

An Overview on Replica Consistency Methods in Distributed Systems and Future Works

Mahsa Beigrezaei^{*1}

¹ Department of Computer Engineering, Yadegar-e-Imam Khomeini (RAH) Shahre Rey Branch, Islamic Azad University, Tehran, Iran.

Received: 07 May 2023, Revised: 27 November 2023, Accepted: 01 December 2023

Paper type: Research

Abstract

Nowadays, applications generate huge amounts of data, in the range of several terabytes or petabytes. This data is shared among many users around the world. Distributed systems such as grid and cloud provide a suitable platform for these applications, enabling the use of these diverse mass data applications in a distributed manner. In these systems, they use data replication to face performance problems, guarantee service quality, and increase data accessibility. Replication, despite its many advantages, also brings administrative costs. The balance between the consistency cost of replication and the benefits of replication is a hotly debated topic among researchers in this field. Therefore, paying attention to the consistency of replication plays an effective role in the efficiency of these systems. Many strategies have been proposed by researchers in the field of data replication consistency. Each of these strategies try to reduce consistency costs and provide effective solutions in this field by considering various parameters such as read rate, write rate, old data tolerance rate, number of replicas and communication bandwidth in determining the consistency levels of replicas. In this article, we will examine the concepts related to replication and replica consistency and categorize its types and review previous works in this field. The done works have been compared from the perspective of system type, decision parameters, compatibility model and improved parameters. At the end, the open issues in this field are raised.

Keywords: Data Replication, Consistency, Models, Cloud, IOT, Distributed System.

* Corresponding Author's email: m.beigrezaei@gmail.com

مروری بر سازگاری تکثیر داده در سیستم‌های توزیع شده

مهسا بیگ رضایی^۱*

^۱ گروه مهندسی کامپیوتر، واحد یادگار امام خمینی (ره) شهری، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۲/۱۷ تاریخ بازبینی: ۱۴۰۲/۰۹/۰۶ تاریخ پذیرش: ۱۴۰۲/۰۹/۱۰

نوع مقاله: پژوهشی

چکیده

سیستم‌های توزیع شده مانند گرید و ابر به منظور مواجهه با مشکلات کارایی، تضمین کیفیت خدمت و افزایش دسترسی پذیری به داده‌ها از تکثیر داده استفاده می‌کنند. تکثیر با وجود مزایای بسیار هزینه‌های مدیریتی نیز به همراه دارد. سازگار نگه داشتن تکثیرها از جمله مهم‌ترین هزینه‌های ناشی از تکثیر است. تعادل بین هزینه سازگاری تکثیر و مزایای تکثیر یک موضوع مورد بحث و داغ در بین محققان این حیطه است. لذا توجه به سازگاری تکثیر نقش موثری در کارایی این سیستم‌ها بازی می‌کند. استراتژی‌های بسیاری توسط محققان در حیطه سازگاری تکثیر داده ارائه شده است. هر کدام از این استراتژی‌ها با در نظر گرفتن پارامترهای مختلفی مانند نرخ خواندن، نرخ نوشتن، نرخ حمل داده‌های قدیمی، تعداد تکثیرها و پهنای باند ارتباطی در تعیین سطوح سازگاری تکثیرها سعی در کاهش هزینه‌های سازگاری و ارائه راهکارهای مؤثر در این حوزه دارند. در این مقاله به مفاهیم تکثیر و سازگاری تکثیر پرداخته می‌شود. دسته‌بندی‌ها و روش‌های سازگاری موجود در این حوزه بررسی می‌شود. کارهای انجام شده در حیطه سازگاری تکثیر داده از دیدگاه‌های مختلفی مانند نوع سیستم، پارامترهای تصمیم‌گیری، ابزار شبیه‌سازی، مدل سازگاری و پارامترهای بهبود داده شده مقایسه می‌شوند. همچنین در پایان، موضوعات باز در این حوزه مطرح می‌شود.

کلیدواژگان: تکثیر داده، سازگاری، مدل‌های سازگاری، سیستم توزیعی، ابر، گرید داده.

* رایانامه نویسنده مسؤول: m.beigrezai@gmail.com

۱- مقدمه

امروزه همراه با پیشرفت و رشد اینترنت، برنامه‌های کاربردی با تعداد زیادی کاربر توزیع شده در سطح جهان تمایل به ارتباط برخط دارند. در این راستا سیستم‌های توزیع شده مانند خوشه‌ها^۱، گرید محاسباتی، ابر محاسباتی و سیستم‌های نظیر به نظیر^۲ پا به عرصه وجود گذاشتند. افزایش کاربران و ازدحام درخواست‌ها و دوری کاربران از منابع، کاهش کیفیت خدمت و تأخیر و نارضایتی کاربران را به همراه دارد. این کاستی‌ها زمانی که درخواست کاربران نیازمند حجم زیادی از داده‌ها باشند، به طور چشمگیری افزایش می‌یابد.

برنامه‌های کاربردی امروزی در بسیاری از زمینه‌ها مثل تجاری، علمی، آموزشی سرگرمی در حال تولید حجم انبوهی اطلاعات است [۱-۵]. به عنوان مثال، برنامه‌ها با کاربرد فیزیک با انرژی بالا^۳ [۶] و بیوانفورماتیک و وب سایت‌های محبوب مانند Facebook, Twitter, Google, Amazon در شبکه‌های گرید، شبکه‌های ابری [۷]، اینترنت اشیا (IOT) [۸] تولیدکننده داده با حجم بالا، در حد ترابایت و پتابایت است. با رشد روز افزون این برنامه‌ها در سیستم‌های توزیع شده، تضمین فاکتورهای کارایی مانند مقیاس پذیری، قابلیت اطمینان، دسترسی پذیری، قابلیت گسترش، کاهش تأخیر و هزینه‌ها، یکی از مشکلات اصلی این محیط‌های توزیع شده است. تکثیر داده یک روش کلیدی و کارا برای مقابله با مشکل‌های مطرح شده و رسیدن به تحمل پذیری خطا و حل مشکل گلوگاه، است. تکثیر داده تکنیکی است که چندین کپی از داده مورد نیاز کاربر را در سطح سیستم توزیعی در نزدیکی کاربر به منظور افزایش کارایی قرار می‌دهد. امروزه استراتژی تکثیر به صورت گسترده و برای افزایش کارایی در شبکه جهانی وب، شبکه‌های نظیر به نظیر [۹]، سنسورها، ابر محاسباتی، گرید [۱۰] و سایر سیستم‌های توزیعی استفاده می‌شود [۱۱-۱۵].

امروز با ظهور سیستم‌های توزیع شده نوین مانند ابر، تکثیر داده دوباره به عنوان موضوعی داغ مورد توجه محققان قرار گرفته است. در این سیستم‌ها، برای رسیدن به کارایی مد نظر داده‌های تعداد زیادی از برنامه‌های کاربردی تکثیر می‌شود. این سیستم‌ها استراتژی‌های مختلف تکثیر و مدیریت سازگاری تکثیر را به کار می‌برند. در استراتژی‌های تکثیر داده، چندین کپی از داده در سرورهای مختلف در مناطق مختلف توزیع می‌شود. این کار به منظور کاهش زمان دسترسی و مصرف پهنای باند، توزیع بار درخواست‌ها و زنده نگه داشتن خدمات داده در زمان خرابی سرورها

است [۱۶، ۱۷] و با وجود این مزایا، تکثیر سربراهایی را به سیستم توزیعی تحمیل می‌کند که در صورت عدم کنترل مناسب این سربراه هزینه‌های ناشی شده از آن بیشتر از سودمندی آن خواهد بود. سازگار نگه داشتن تکثیرها یکی از مهم‌ترین و پرچالش‌ترین هزینه‌های مدیریت تکثیر داده است. مدیریت سازگاری تکثیرها اگر به درستی انجام نشود می‌تواند منجر به کندی سیستم شود. به طوری که تکثیر را تبدیل به معطلی بزرگ در سیستم‌های توزیعی کند. از این رو مدیریت سازگاری تکثیر در سیستم‌های توزیعی با قابلیت خواندن و نوشتن بالا بسیار اهمیت دارد.

با وجود اهمیت مبحث مدیریت سازگاری تکثیر داده، تعداد کمی از مقالات مروری پیشین به طور ویژه به بررسی این مبحث پرداخته‌اند. بر اساس مطالعات ما، مرجع‌های [۱۸-۲۱] مقاله‌های مروری موجودی هستند که به مبحث سازگاری تکثیر داده پرداخته‌اند. در این مرجع‌ها به طور جامع به مروری بر این مبحث پرداخته نشده است. به عنوان نمونه در [۱۸]، تنها به مدیریت سازگاری در شرایط خاص تکثیر خوشبینانه داده پرداخته شده است. در مقاله [۱۹]، تنها در سیستم توزیعی ابر مدیریت سازگاری مورد بررسی قرار گرفته است و تنها به تعدادی محدودی از روش‌های پایه و قدیمی پرداخته شده است. در [۲۱]، تنها به تعدادی کارهای ارائه شده در مبحث الگوریتم‌های سازگاری در سیستم توزیعی گرید پرداخته شده است. تقسیم‌بندی‌های ارائه شده در این مقاله از نقطه نظر جنبه توپولوژی‌های گرید و پارامترهای ارزیابی و نرم‌افزار شبیه‌سازی است. در [۲۰]، پروتکل‌های پایه تکثیر و پروتکل‌های به‌روزرسانی تکثیر در سیستم‌های پایگاه داده‌ای بررسی می‌شود. در این مقاله کارهای جدید کمی مطرح شده و بیشتر تمرکز بر روی کارهای پایه است.

از بررسی کارهای مروری انجام شده می‌توان متوجه شد در هیچ‌کدام از کارهای پیشین تقسیم‌بندی‌ها در تمام سیستم‌های توزیعی بررسی نشده، و کارهای آینده و چالش‌ها و موضوعات باز به طور خاص بررسی نشده‌اند. در این مقاله برای رفع کاستی‌های موجود الگوریتم‌های سازگاری تکثیر را در سیستم‌های توزیعی قدیمی و جدید بررسی می‌کنیم. در این بررسی در ابتدا مفهوم تکثیر و تقسیم‌بندی‌های تکثیر و مفهوم سازگاری تکثیر می‌پردازیم. در ادامه به بررسی روش‌های پایه و مشهور و تقسیم‌بندی‌های موجود و جدید از منظرهای مختلف پرداخته می‌شود. در بخش‌هایی نیز به طور جدا کارهای موجود و جدید در سیستم توزیعی‌های محبوب بررسی می‌شود. در این مقاله سعی می‌شود تقسیم‌بندی‌ها به طور

³ High Energy Physics (HEP)

¹ Clusters

² Peet to Peer

شده است و روش‌های تکثیر به طور کلی به الگوریتم‌های پویا و ایستا طبقه‌بندی می‌شوند [۲۳-۲۲]. در الگوریتم‌های ایستا، تعداد تکثیرها و مکان آنها ثابت است. در مقابل الگوریتم‌های تکثیر پویا با الگوهای دسترسی کاربر و تغییرات سازگار هستند. الگوریتم‌های پویا تغییر نیازمندی‌های سیستم را در نظر می‌گیرند. ماهیت محیط توزیع شده پویا است. بنابراین روش‌های پویا موفق‌تر از ایستا در سیستم‌های توزیع شده واقعی هستند. در شبکه‌های پراکنده بزرگ به دلیل فاصله جغرافیایی، تأخیر دسترسی به داده‌ها زیاد است. تأخیرهای زیاد ممکن است منجر به یک مشکل حیاتی برای ارائه‌دهندگان خدمات و مشتریان شود [۱۷]. از نظر نوع سازمان تصمیم‌گیری می‌توان آن را به دو دسته توزیعی و متمرکز طبقه‌بندی کرد. در این دو روش، چندین گره توزیع شده و یک گره به ترتیب تصمیمات تکثیری را اتخاذ می‌کنند. از سوی دیگر، فرآیند تصمیم‌گیری می‌تواند به صورت برخط یا برون خط انجام شود. الگوریتم‌ها می‌توانند یک هدف یا چند هدف داشته باشند. علاوه بر این، بر اساس درصد تکثیر فایل، تکثیر داده‌ها را می‌توان به تکثیر کامل و تکثیر جزئی تقسیم کرد. در تکثیر کامل ۱۰۰٪ حجم فایل کپی شده و به صورت جزئی، فقط یک قسمت ضروری از فایل کپی می‌شود. در تکثیر جزئی می‌توان از دانه‌بندی‌های مختلف تکثیر مانند بلوک، جدول، ردیف، شی و پایگاه داده استفاده کرد. اغلب تکثیر جزئی مؤثرتر از تکثیر کامل است؛ اما پیچیده‌تر است. اصلی‌ترین زیر ساخت‌ها در سطح بزرگ مقیاس که از تکثیر داده استفاده می‌کنند شبکه‌های نظیر به نظیر، گرید، ابر، لبه و مه هستند. محیط‌هایی که از تکثیر داده بهره می‌برند می‌توانند همگن یا ناهمگن باشند. تاکنون الگوریتم‌های تکثیر زیادی ارائه شده است. ارزیابی کارایی این الگوریتم‌ها توسط روش‌های مختلفی مانند شبیه‌سازی، اثبات ریاضی، پیاده‌سازی در محیط واقعی انجام شده است [۲۴].

۳- سازگاری تکثیرهای داده

بحث سازگاری تکثیر داده در جایگاه‌های مختلفی مثل معماری کامپیوتر، پایگاه داده و سیستم‌های توزیع شده مطرح می‌شود که در هر کدام می‌تواند معانی متفاوتی داشته باشد. در تمام آنها باید به این سؤال پاسخ داده شود که چه رخدادی باید بیفتد، زمانی که تکثیری که احتمالاً به زودی مورد دسترس واقع می‌شود، تغییر کند. بسته به محتوای داده تکثیر شده و نوع برنامه کاربردی که تکثیر به آن وابسته است پاسخ‌های متفاوتی به این سؤال داده می‌شود که تحت عنوان سازگاری به این مقوله پرداخته می‌شود.

جامع‌تری مورد توجه قرار گیرد. در پایان نیز چالش‌ها و کارهای آتی بیان می‌شود تا دید جامع از آینده این حوزه به خواننده داده شود. این مقاله شامل ۷۵ مرجع است که در یک دوره زمانی منتشر شده است. این مقالات شامل کارهای انجام شده توسط محققان از دهه ۱۹۸۰ تا نیمه سال ۲۰۲۳ می‌شود. با توجه به ماهیت آثار تحلیل شده، این مقاله هم مقالات کنفرانس‌ها و هم مجلات، را پوشش می‌دهد به طور خاص، این تحلیل ۱۳ کنفرانس، ۵۰ مجله، هشت یادداشت سخنرانی، یک مرجع آنلاین، و سه کتاب را در نظر پوشش می‌دهد. در بین مجلات، بیشتر آنها با موارد نمایه شده در گزارش استنادی تامسون مجله مطابقت دارد که به عنوان شاخصی مناسب برای اندازه‌گیری تأثیر مجلات در نظر گرفته می‌شود.

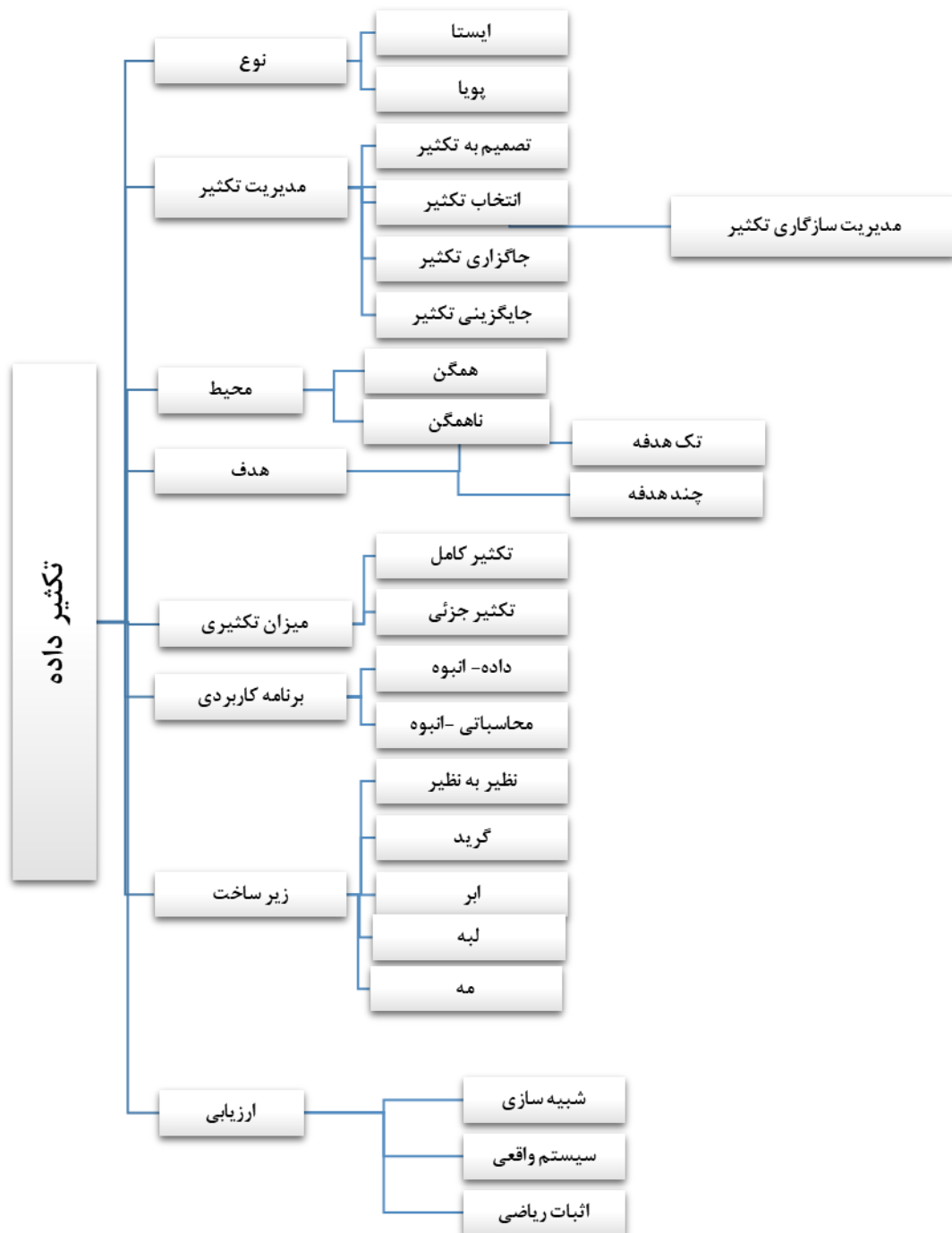
در ادامه این مقاله در بخش ۲ به بررسی تکثیر داده و تقسیم‌بندی‌ها در تکثیر داده می‌پردازیم. در بخش ۳ مفهوم سازگاری، روش‌های مطرح در سازگاری، تقسیم‌بندی‌ها و مدل‌های سازگاری موجود مطرح می‌شود. در بخش ۴ سازگاری در سیستم‌های توزیعی ابر و مه^۱ و اینترنت اشیا مورد بررسی قرار خواهد گرفت. در بخش ۵ چالش‌های موجود و کارهای آتی مطرح شده و در بخش پایانی نتیجه‌گیری بیان می‌شود.

۲- تکثیر داده و تقسیم‌بندی الگوریتم‌های تکثیر

با وجود مزایای بسیار ذکر شده، تکثیر سربراهایی را به سیستم توزیعی تحمیل می‌کند. در صورت عدم کنترل مناسب این سربرار هزینه‌ها بیشتر از سودمندی آن خواهد بود. چرا که عمل تکثیر ذاتاً عملی هزینه‌بردار است و در جهت استفاده از آن منابع ذخیره‌سازی و پهنای‌بند شبکه باید مصرف شود. لذا مدیریت تکثیر، امری خطیر و چالشی مهم در سیستم‌های توزیع شده است. در مدیریت تکثیرهای داده باید به سوال‌های مهمی در این بخش پاسخ داده شود. این سؤال‌ها عبارت‌اند از: ۱. چه فایلی باید تکثیر شود؟ ۲. چه زمانی عمل تکثیر انجام شود؟ ۳. بهترین مکان تکثیر کجاست؟ ۴. چه تکثیرهای باید حذف شوند؟ ۵. چگونه داده‌ها سازگار نگه داشته شوند؟ چهار سؤال اول توسط الگوریتم‌های تکثیر و سؤال آخر توسط مدل‌ها و استراتژی‌های سازگاری پاسخ داده می‌شود [۲۲]. در این مقاله تمرکز اصلی بر روی سوال پنجم است. در واقع این مقاله به بررسی روش‌ها و کارهای انجام شده در پاسخ به سوال پنجم می‌پردازد. سایر سوال‌ها تمرکز این مقاله مروری نمی‌باشد.

در شکل ۱ تکثیر داده‌ها را به طور خلاصه از دیدگاه‌های مختلف بررسی و طبقه‌بندی می‌کنیم. همانطور که در شکل ۱ نشان داده

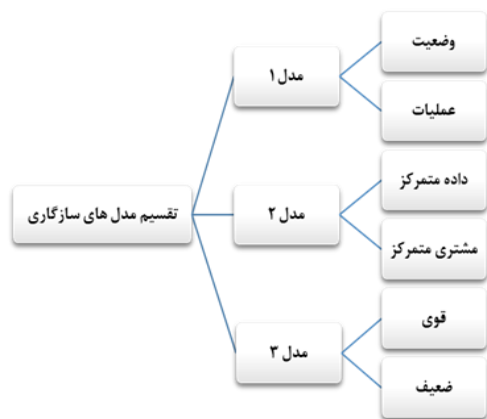
^۱ Fog



شکل ۱. تقسیم‌بندی الگوریتم‌های تکثیر داده از دیدگاه‌های مختلف

مدیریت تکثیر داده، سازگاری تکثیرها است. مشکل سازگاری و ضمانت سازگاری تکثیرها از طریق اعمال سازگاری قوی بر طرف می‌شود. اما به دلیل ماهیت ذاتی تأخیر بالا در سیستم‌های ذخیره‌سازی موجود در سیستم‌های توزیعی استفاده از تکثیر همزمان و سازگاری قوی باعث افزایش ترافیک شبکه و هزینه‌ها و نهایتاً تأخیر در شبکه می‌شود. این تأخیرها هزینه‌های قابل توجهی را به بار می‌آورند. به عنوان مثال هزینه تنها یک ساعت از کار افتادگی و تأخیر برای سیستم که مجوزهای فروش کارت اعتباری

این پاسخ در سیستم‌های توزیع شده اهمیت بالایی دارد [۲۴]. چرا که باید توسعه‌دهنده‌ها از این سؤال پایه‌ای آگاه باشند. به خصوص زمانی که مشتری‌ها انسان نباشند، که این مسئله در سیستم‌های توزیعی امروزی مانند ابر، گرید و نظیر به نظیر یک اتفاق رایج است. در تعریفی دیگر، سازگاری تکثیر داده یعنی معتبر بودن، دقت، قابل استفاده بودن و صحت داده‌های تکثیر شده از نظر برنامه‌هایی که از این داده‌ها استفاده می‌کنند [۱۲، ۲۱]. روش‌های مختلفی برای رسیدن به این تعاریف مطرح شده است. یکی از بخش‌های مهم در



شکل ۲. تقسیم‌بندی‌های مرسوم مدل‌های سازگاری

مدل‌های سازگاری هر یک تعداد محدودی از نیازهای سازگاری را مطرح می‌کنند. بنابراین تا کنون مدل‌های بسیاری مطرح شده است. تقسیم‌بندی دوم توسط آقای تنبام ارائه شده است. در یک تقسیم‌بندی، تنبام مدل‌های سازگاری را در دو دسته داده-متمرکز^۹ و مشتری-متمرکز^{۱۰} قرار می‌دهد. مدل داده-متمرکز بر روی سازگاری یک سیستم، از دیدگاه وسیع سیستم ذخیره‌سازی متمرکز می‌شود.

این تقسیم‌بندی در شکل ۳ به طور کامل‌تری همراه با مثال‌هایی از هر دسته به تصویر کشیده شده است. مثال‌هایی که در این مدل جا می‌گیرند عبارت‌اند از: سازگاری سخت، سازگاری علیت، سازگاری ترتیبی، مدل مشتری-متمرکز بر روی سازگاری از دیدگاه یک مشتری تمرکز دارد و تنها به داده‌های ذخیره شده به وسیله مشتری متمرکز می‌شود. مثال‌ها شامل: سازگاری خواندن یکنواخت^{۱۱}، سازگاری نوشتن یکنواخت^{۱۲}، سازگاری خواندن خود نوشته‌ها^{۱۳} و سازگاری نوشتن به دنبال خواندن^{۱۴} [۲۶]. در یک تقسیم‌بندی خیلی کلی‌تر مدل‌های سازگاری به دو دسته مدل‌های سازگاری قوی و مدل‌های سازگاری ضعیف تقسیم می‌شوند و مدل‌هایی مانند سازگاری شدید^{۱۵}، سازگاری ترتیبی^{۱۶}، سازگاری خطی‌سازی^{۱۷} در دسته قوی و مدل‌هایی مانند سازگاری احتمالی^{۱۸} و خواندن بعد از نوشتن^{۱۹} و خواندن بعد از نوشتن^{۲۰} و سازگاری علیت در دسته ضعیف قرار می‌گیرند [۲۶].

را صادر می‌کند، حدود ۳٫۱ میلیون دلار است. لذا مدیریت سازگاری با حداقل تأخیر یک امر مهم و یک چالش در سیستم‌های توزیعی است [۲۵].

۳-۱- تئوری CAP^۱ و چالش جدید سازگاری و کارایی

تئوری CAP به عنوان یک چالش در طراحی سیستم‌های ذخیره‌سازی مطرح شده است. در این تئوری اثبات شده است که تنها دو خصوصیت از سه خصوصیت (دسترسی‌پذیری، قابلیت حمل) را می‌توان به طور عملی و همزمان فراهم کرد. در توزیعی تضمین سازگاری و دسترسی‌پذیری با اهمیت‌تر است. در این میان به دلیل وجود تأخیر بالا در ماهیت سیستم‌های توزیعی ابر، محققان مصالحه‌ای جدید و مهمتری به نام سازگاری و کارایی را مطرح کردند که مدل‌های سازگاری متفاوتی توسط محققان برای غلبه بر این چالش‌ها مطرح شده است [۲۵].

۳-۲- تقسیم‌بندی موجود در سازگاری تکثیر

در منابع مختلف روش‌های سازگاری از نقطه نظرهای مختلفی تقسیم‌بندی شده‌اند. در شکل ۲ این تقسیم‌بندی‌ها به تصویر کشیده شده‌اند. در ادامه هر کدام را به تفصیل توضیح می‌دهیم. در تقسیم‌بندی اول مدل‌های سازگاری از نظر خصیصه‌های سازگاری به دو دسته کلی سازگاری وضعیت^۲ و سازگاری عملیاتی^۳ تقسیم می‌شود. در سازگاری وضعیت سازگاری بر روی وضعیت داده تمرکز دارد (مثلاً ناسازگاری دو تکثیر^۷، با انحراف عددی بیشتر از x ، تعریف می‌شود). روش‌های موجود در دسته سازگاری وضعیت به دسته‌های نامتغیر^۴، محدوده‌ی خطا^۵ تقسیم‌بندی شده است. روش‌هایی که در دسته سازگاری عملیاتی قرار می‌گیرند تمرکز بر روی عملیات‌ها و نتایجی که بر می‌گردانند است. این دسته خود به تعادل ترتیبی^۶، تعادل مراجع^۷، خواندن - نوشتن متمرکز^۸، تقسیم‌بندی می‌شود [۲۵-۲۶]. از طرفی آقای اندرو تنبام مدل‌های سازگاری را قراردادی بین مراکز داده توزیع شده و پردازش‌ها دانسته که با رعایت آنها نتایج عملیات خواندن نوشتن درست و معتبر خواهد بود [۲۶].

¹¹ Monotonic Reads

¹² Monotonic write

¹³ Read Your Writes

¹⁴ Writes Follow Reads

¹⁵ Strict Consistency

¹⁶ Sequential Consistency

¹⁷ Linearizability

¹⁸ Eventual Consistency

¹⁹ Read-After-Read

²⁰ Read After Write

¹ Consistency, Availability, and Partition Tolerance (CAP)

² State Consistency

³ Operation Consistency

⁴ Invariants

⁵ Error Bounds

⁶ Sequential Equivalence

⁷ Reference Equivalence

⁸ Read-Write Centric

⁹ Data- Centric

¹⁰ Client- Centric



شکل ۳. انواع مدل‌های سازگاری از دیدگاه تنبام

شکل ۴. انواع مدل‌های سازگاری از منظر مدل‌های به‌روزرسانی

معایب: ۱. زمان پاسخ تراکنش‌ها را بالا می‌برد. ۲. برای گره‌های متحرک مناسب نمی‌باشند. ۳. باید کل سیستم اصلی تا زمان به‌روزرسانی سایر تکثیرها منتظر بماند. ۴. در زمانی که تأخیر بالا و لینک‌های ارتباطی دارای احتمال خرابی هستند، این روش کارایی ندارد. ۵. در زمانی که تعداد تکثیرها و تعداد عملیات به‌روزرسانی بالا است سیستم به‌شدت با مشکل کارایی مواجه می‌شود.

• روش‌های غیرهمزمان

در این روش‌ها تأکیدی بر دریافت تاییدیه از نسخه‌های اصلی و ثانویه تکثیرها نداریم. همزمان‌سازی تکثیرها در بازه‌های زمانی انجام می‌شود. میزان این بازه بستگی به تأخیر لینک‌های ارتباطی و بار کاری سرورها دارد. در این مدل به‌روز رسانی باعث بلاک شدن دسترسی به سایر تکثیرها نمی‌شود. همزمان‌سازی تبیل و آرام شده^۱ نام‌های دیگری برای به‌روزرسانی غیرهمزمان است. در سیستم‌های توزیع شده اغلب برنامه‌های توزیعی از این مدل به‌روزرسانی استفاده می‌کنند. این مدل به دو دسته فراگیر و مبتنی بر تقاضا تقسیم می‌شود. در دسته فراگیر، از یک فایل چندین نسخه از فایل همزمان می‌تواند وجود داشته باشد. وقتی یک فایل تغییر می‌کند به‌روزرسانی‌ها به شکل تبیل برای سایر تکثیرها به‌روزرسانی منتشر می‌شوند. در این وضعیت اغلب سطح ضعفی از سازگاری را خواهیم داشت. در این روش هر تکثیر در هر زمانی می‌تواند به‌روزرسانی شود. لذا تعداد برخوردها زیاد خواهد بود. از این روش با نام ذخیره و ارسال^۲ نیز یاد می‌شود. در دسته مبتنی بر تقاضا تا زمانی که کلاینتی در شبکه درخواست دسترسی به فایل را ندهد داده به روز نمی‌شود.

برخی از این مدل‌ها در جدول ۱ به اختصار توضیح داده شده‌اند. مدل‌های سازگاری مانند Adaptive View Fork [۳۶]، VFC3 [۳۷] و Red-Blue که نیازهای ضروری جدید سیستم‌های محبوب جدید مانند ابر، مه و اینترنت اشیاء صنعتی را برطرف می‌کنند، به عنوان مدل‌های جدید در نظر گرفته می‌شوند. برخی از نیازهای حیاتی جدید که این مدل‌ها بر روی آنها تمرکز دارند عبارتند از: امنیت، قابلیت اطمینان، تضمین SLA، همگرایی بیشتر در عملیات و مصرف انرژی. همچنین مدل‌های سازگاری مانند Time Casual، Fork و Casual+ که از سایر مدل‌های پایه توسعه یافته‌اند، به عنوان مدل‌های توسعه یافته در نظر گرفته می‌شوند.

۳-۲-۱- مدل‌های سازگاری از منظر تکنیک به‌روزرسانی و شروع کننده انتشار

مدل‌های سازگاری از دیدگاه تکنیک که در به‌روزرسانی استفاده می‌کنند و از دیدگاه موجودیت شروع کننده به‌روزرسانی قابل بررسی و طبقه‌بندی است. در شکل ۴، از نقطه نظر روش‌های به‌روزرسانی، سازگاری تکثیرها به دو روش همزمان و غیرهمزمان تقسیم می‌شوند [۲۳].

• روش‌های همزمان

در روش همزمان در هر بار عملیات نوشتن تمام تکثیرهای اصلی و ثانویه به‌روزرسانی می‌شوند. در این روش تا زمانی که تمام تکثیرها به‌روزرسانی نشود عملیات تکمیل شده در نظر گرفته نمی‌شود. در واقع این روش تراکنش‌ها به صورت واحد و اتمیک در نظر گرفته می‌شود. این روش با نام‌های به‌روز رسانی مشتاق و بدبینانه نیز شناخته می‌شوند. مزایا و معایب این دسته از روش‌ها عبارتند از:

مزایا: ۱. داده‌ها در هر زمان به‌روز هستند. ۲. هیچ برخوردی بین داده‌ها وجود ندارد.

² Store And Forward

¹ Relaxed Synchronous

جدول ۱. شرح مختصری از برخی مدل‌های سازگاری رایج تقسیم‌بندی شده در پنج دسته داده متمرکز، مشتری متمرکز، ترکیبی، جدید و

توسعه یافته

داده متمرکز (data - centeric)	
Strict [۲۷]	قوی ترین مدل سازگاری است. تکثیرها در سطح جهانی با زمان مطلق جهانی همگام می‌شوند.
Sequential [۲۶]	این مدل شبیه سازگاری strict است. با این حال، ضعیف‌تر است. فرآیندها عملیات نوشتن را مشاهده می‌کنند که دارای رابطه علیت با همان ترتیب هستند، اما عملیات خواندن اجرا شده توسط سایر فرآیندها قابل مشاهده نیست.
Causal [۱۷]	این مدل روابط علی بالقوه بین عملیات را تعیین می‌کند. این تضمین می‌کند که عملیات مرتبط با علت به همان ترتیب توسط همه فرآیندها بازدید می‌شود. بنابراین اگر عملیات خواندن نتیجه برخی از عملیات نوشتن باشد، تا زمانی که تمام عملیات نوشتن تمام نشده باشد، انجام نخواهد شد.
Weak [۲۶]	این مدل یک متغیر همگام‌سازی را به عنوان نشانه مشخص می‌کند. عملیات نوشتن و خواندن را می‌توان توسط فرآیندهایی که دارای این نشانه هستند اجرا کرد. در این مدل، داده‌های مشترک با سازگاری متوالی در متغیر همگام‌سازی قابل دسترسی است.
Releases [۲۸]	از متغیر همگام‌سازی یا قفل استفاده می‌کند. این فرآیند، در مرحله اول، از یک قفل برای دسترسی به یک منبع مشترک (مثنی) درخواست می‌کند. سپس در مرحله دوم، قفل آزاد می‌شود و مقدار تغییر یافته ماکت به ماکت‌های دیگر ارسال می‌شود.
Releas -lazy [۲۸]	این مدل یک بسط ضعیف قوام انتشار است. یک فرآیند می‌تواند نمایش نتایج عملیات نوشتن خود را تا عملیات همگام‌سازی بعدی به تاخیر بیندازد. از مهر زمانی استفاده می‌کند تا تشخیص دهد که نسخه جدیدترین به‌روزرسانی است یا قدیمی.
Entry [۲۹]	در این مدل قبل از دسترسی به هر داده، قفل آن دریافت می‌شود و سپس می‌توان هر عملیاتی را انجام داد.
مشتری متمرکز (client centric)	
Monotonic read [۲۶]	این مدل تضمین می‌کند که فرآیند هرگز ارزش قبلی مورد X را در زمانی که فرآیند مشاهده می‌کند مشاهده نکند. بانک اطلاعات ایمیل توزیع شده نمونه ای از این مدل است که در آن هر کاربر یک صندوق پستی در چندین نسخه دارد.
monotonic write [۲۶]	در این مدل، عملیات نوشتن بر روی یک داده توسط یک فرآیند زمانی قابل قبول است که عملیات نوشتن قبلی را به طور کامل روی آن انجام داده باشد.
Read Your Write [۲۶]	در این مدل، یک مشتری از هر مقدار نوشته شده جدید به ترتیب انجام توسط مشتری بازدید می‌کند. در واقع، پس از اتمام عملیات نوشتن قبلی، مشتری مقدار تازه نوشته شده را می‌خواند.
Write follow Read [۲۶]	این مدل به عنوان جلسه علیت نیز شناخته می‌شود. عملیات نوشتن روی یک آیتم داده زمانی توسط مشتری قابل قبول است که مشتری آخرین مقدار مورد-داده (item-data) را خوانده باشد. تویپتر نمونه ای از استفاده از این مدل است.
ترکیبی (Combination/ Hybrid)	
Fork Sequential [۳۰]	در این مدل، مشتریان متعددی که می‌توانند یک تراکنش را به طور مستقیم یا غیرمستقیم مشاهده کنند، تمامی عملیات‌های قبلی را بر اساس تاریخچه رویداد خود مشاهده می‌کنند. این مدل اتمی بودن عملیات مشتری را تضمین نمی‌کند. این نقص در مدلی به نام Fork Linearize Consistency (FLC) برطرف شده است.
Fork Linearize [۳۱]	
RedBlue Novel [۳۲]	این مدل عملیات را بر اساس نوع اجرا به عملیات قرمز و آبی تقسیم می‌کند. ترتیب اجرای عملیات قرمز باید در همه سایت‌ها یکسان باشد اما ترتیب رنگ آبی مهم نیست
جدید	
RedBlue [۳۲]	این مدل عملیات را بر اساس نوع اجرا به عملیات قرمز و آبی تقسیم می‌کند. ترتیب اجرای عملیات قرمز باید در همه سایت‌ها یکسان باشد اما ترتیب عملیات آبی مهم نیست
Timed [۳۳]	در این مدل، اگر عملیات نوشتن در زمان t انجام نشود، پس از زمان Δ ، سایر گره‌ها قابل مشاهده هستند. این ترکیبی از ناسازگاری و ترتیب قدیمی را ارائه می‌دهد و به آن قوام دلتا می‌گویند.
Fork [۳۴]	برای سیستم‌های ذخیره‌سازی با کلاینت‌های غیر قابل اعتماد استفاده می‌شود و همزمانی و امنیت را افزایش می‌دهد. این سه ویژگی اصلی را اجرا می‌کند، از جمله ۱. کلاینت سالم لیستی از عملیات صحیح را به سرور ارسال می‌کند. ۲. یک کلاینت سالم تمام عملیات قبلی خود را مشاهده می‌کند. ۳. همه مراجعان سالم جنبه یکسانی از عملیات آن مشتری سالم دارند.
توسعه یافته	
Fork* [۳۵]	Fork* (FC) مشابه سازگاری fork است با این تفاوت که در خاصیت سوم، از ویژگی join-at-once استفاده می‌کند و دقت و بهبود بازرسی سیستم را بهبود می‌بخشد.

مزایا و معایب روش‌های غیرهمزمان عبارتند از:

۱. برنامه‌ها نیاز به داشتن اطلاع از به‌روز رسانی‌ها و وجود تکثیرها در محل‌های دور ندارند، لذا کارایی کلی سیستم افزایش می‌یابد. ۲. در زمان به‌روزرسانی یک تکثیر سایر تکثیرهای همان قابل دسترس هستند. ۳. در زمان خراب شدن تکثیرها اصلی سایر تکثیرها قائل دسترس هستند. ۴. پایداری و دسترسی پذیری را تضمین می‌نماید. ۵. به ارمغان آوردن کاهش زمان دسترسی و افزایش کارایی به دلیل در دسترس بودن تکثیرها.

مزایا: ۱. برنامه‌ها نیاز به داشتن اطلاع از به‌روز رسانی‌ها و وجود تکثیرها در محل‌های دور ندارند، لذا کارایی کلی سیستم افزایش می‌یابد. ۲. در زمان به‌روزرسانی یک تکثیر سایر تکثیرهای همان

- استراتژی‌های ترکیبی

در این استراتژی‌ها دو روش مبتنی بر فشار و مبتنی بر کشیدن را به صورت ترکیبی استفاده می‌کنند.

۴- روش‌های سازگاری تکثیر داده از منظر نوع سیستم‌های توزیعی

در این بخش به بررسی مدل‌های سازگاری در هر دسته از خدمات توزیعی می‌پردازیم.

۴-۱- سازگاری در ابر محاسباتی

همانطور که محاسبات ابری به طور فزاینده‌ای محبوب می‌شود، همزمان خدمات ذخیره‌سازی ابر نیز به ایجاد امنیت، کارایی بالا و در دسترس بودن با هزینه کم برای کاربران خود بیشتر توجه دارند. انتظار می‌رود ذخیره‌سازی ابر به نیروی اصلی بازارهای ذخیره‌سازی آینده، تبدیل شود. همان‌طور که گفته شده تکنیک تکثیر داده یکی از تکنیک‌های محبوب در افزایش کارایی در سیستم‌های ابر است. تکثیر داده به‌عنوان یک تکنیک کلیدی در محاسبات ابری، با چالش سازگاری تکثیرها مواجه می‌شود. در ابر مانند سایر سیستم‌های توزیعی، تحقق سازگاری تکثیرها به حل کردن همگام‌سازی بین دو تکثیر یا بیشتر می‌پردازد. در راستای برآورده کردن این هدف، سیستم با هزینه‌ها و چالش‌هایی مواجه خواهد شد. محققان مدل‌های سازگاری تکثیر داده مختلفی را در تلاش برای برآورده کردن این اهداف در ابر ارائه کرده‌اند.

ازجمله سازگاری‌های پایه‌ای موجود در ابر، می‌توان به سازگاری قوی، ضعیف و سازگاری احتمالی اشاره کرد [۳۸]. در شکل مرسوم، سازگاری قوی (مشهور به سازگاری مشتاق^۴)، با تغییر یک داده، تضمین می‌شود که همه تکثیرهای آن فوراً به‌روز رسانی شوند سپس به عملیات بعدی اجازه اجرا شدن داده شود. در نسخه دیگر از سازگاری قوی به نام سازگاری قوی-تنبیل^۵ در اولین درخواست خواندن، آخرین نسخه داده تغییر داده شده و به همه تکثیرهای موجود ارسال می‌شود. در سازگاری قوی هیچ تفاوتی بین تکثیرها وجود ندارد؛ بنابراین، در هر دسترسی تازه‌ترین داده به دست خواهد آمد؛ اما در این روش هزینه‌های نگهداری به طور قابل توجهی افزایش می‌یابد، طوری که در دسترس بودن تکثیرها و عملکرد و کارایی سیستم را کاهش می‌دهد.

معایب: ۱. تعداد برخوردها در شرایطی که چند مستر داریم بالا می‌باشد ۲. افزایش برنامه‌های محاسباتی نیازمند به داده باعث وابستگی تنگاتنگ روش با محل تکثیر و وابستگی انتخاب تکثیر با کارایی به‌روزرسانی غیرهمزمان خواهد شد. ۳. مدلی ضعیف در تضمین سازگاری است. ۴. تعیین بازه زیاد یا کم منجر به افزایش سربار و یا کاهش درجه سازگاری خواهد شد.

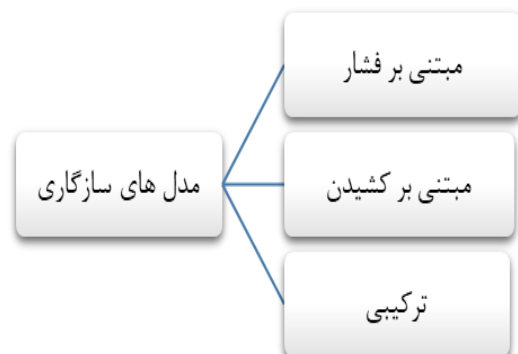
شکل ۵ تقسیم‌بندی مدل‌های سازگاری از نقطه نظر اینکه در انتشار به‌روز رسانی تکثیرها چه کسی شروع‌کننده است (سرور یا مشتری) روش‌های سازگاری موجود به سه دسته مبتنی بر کشیدن^۱ و مبتنی بر فشار دادن^۲، ترکیبی^۳ تقسیم‌بندی می‌شوند.

- روش‌های سازگاری مبتنی بر کشیدن (Pull-based)

سناریو این دسته مبتنی بر مشتری است به این معنی که مشتری با بیرون کشیدن داده‌ی به روز از سرور به سمت خود تغییرات را مقداردهی می‌کند (درخواست به سرور و انتقال از سرور). زمانی که تعداد به‌روزرسانی‌های کمتری در مقایسه با خواندن فایل توسط مشتری وجود دارد، از این روش می‌توان استفاده کرد. یکی از معایب این روش این است که هیچ به‌روزرسانی انجام نمی‌شود اگر سرور از کار افتاده شود.

- روش‌های سازگاری مبتنی بر فشار (push-based)

در این روش سرور، داده‌های به روز را به مشتری‌ها منتشر می‌کند (بدون توجه به درخواست از سمت کلاینت). سرور در این مکانیسم‌ها نقش پررنگی در یکسان‌سازی تکثیرها دارد. زمانی که تعداد به‌روزرسانی و تعداد خواندن‌ها بالا می‌باشد این روش مفید و کاربردی است. زمانی که همه تکثیرها از سرور ارسال داده شوند زمان پاسخ تکثیرها ناچیز خواهد بود.



شکل ۵. انواع مدل‌های سازگاری از منظر شروع‌کننده بروز رسانی

⁴ Egger

⁵ Lazy

¹ Pull Based

² Push Based

³ Hybrid

معمولی) مؤثر بوده است. این بهبود در حالی است که نیازها را در مورد کیفیت خدمات و اطلاعات ارائه شده به کاربران نهایی حفظ می‌کند.

در [۴۳]، نویسندگان استدلال می‌کنند که هزینه‌های مالی باید هنگام ارزیابی و انتخاب یک سطح سازگاری در ابر در نظر گرفته شوند. بر این اساس یک معیار جدید به نام بهره‌وری هزینه-سازگاری تعریف می‌کنند. بر اساس این معیار یک مدل سازگاری غیرهمزمان اقتصادی مؤثر ساده را به نام Bismar ارائه می‌دهند، که به صورت انطباقی سطح سازگاری را در زمان اجرا به منظور کاهش هزینه‌های مالی تنظیم می‌کند در حالی که به طور همزمان یک کسر کمی از خواندن‌های قدیمی را حفظ می‌کند. Bismar روی یک مدل احتمالی سازگاری تکیه می‌کند که خواندن‌های قدیمی و هزینه‌های مرتبط برنامه کاربردی را با توجه به نرخ خواندن/نوشتن فعلی و تأخیر شبکه تخمین می‌زند. Bismar در سیستم ذخیره‌سازی Cassandra روی Grid'5000 پیاده‌سازی شد. نتایج نشان داده روش مذکور هزینه‌ها را پایین نگه می‌دارد در شرایطی که تعداد خواندن‌های داده قدیم از تعداد قابل تحمل خواندن‌های قدیمی در برنامه‌های کاربردی تجاوز نمی‌کند.

در [۴۴]، یک سازگاری جدید در ابر به عنوان یک مدل خدمت به نام Caas ارائه شده است، که متشکل از یک ابر داده بزرگ و چندین ابر حسابرس کوچک است. در مدل Caas، یک ابر داده توسط یک ارائه‌دهنده خدمات ابر نگهداری می‌شود و یک گروه از کاربران که یک ابر حسابرس را تشکیل می‌دهند، می‌توانند تشخیص دهند آیا داده‌های ابر سطح وعده داده شده از سازگاری را ارائه می‌دهند یا خیر. در ادامه الگوریتم‌هایی برای تعیین کمیت شدت موارد نقض با دو معیار تشابه موارد نقض و کهنگی ارزش خواندن طراحی شده است.

در [۴۵]، الگوریتم‌های مؤثری را پیشنهاد کرده‌اند که به برخی سؤالات پاسخ می‌دهد. این سؤالات عبارتند از: چگونه یک کاربر می‌تواند بررسی کند که آیا یک سیستم ذخیره‌سازی وعده‌اش را از سازگاری برآورده می‌کند؟ تضمین تنها سازگاری احتمالی توسط یک سیستم چه مشکلاتی می‌تواند داشته باشد؟ الگوریتم‌ها با تجزیه و تحلیل ردی از تعاملات بین ماشین مشتری و یک انبار ارزش کلیدی، می‌توانند گزارش دهند که آیا رد، امن، منظم یا اتمیک است و اگر نیست چه تعداد موارد نقض وجود دارد. الگوریتم‌ها بر اساس رده‌هایی از کلیدهایی که بر اساس سازگاری شرطی ذخیره شده (Pahoehoe) ارزیابی شده‌اند. نتایج اجراء، تعدادی کم و در مواردی صفر مورد نقض را نشان داده است. به طور کلی این مقاله به چگونگی

در [۳۹]، کاری دیگر با موضوع سازگاری داده، مکانیسم تطبیقی و غیرهمزمان بر مبنای برنامه کاربردی برای سازگاری تکثیر در ابر پیشنهاد شد. مدیریت کپی‌ها در این مدل یک ساختار متمرکز دارد. دامنه اجرایی آن شامل یک گره اصلی، سه گره جانشین و گره‌های فرزند است. گره اصلی مهم‌ترین بخش از دامنه اجرایی است که اطلاعات همه گره‌ها را نگه می‌دارد و به طور همزمان، مسئول برای تأیید تمام درخواست‌های به‌روزرسانی است و سپس یک عملیات مربوطه را با توجه به استراتژی سازگاری تکثیر اجرا می‌کند. هدف آن‌ها حل کردن مشکلاتی مانند: تراکم شبکه و دسترسی‌پذیری پایین است. نتایج نشان می‌دهد که مکانیسم تطبیقی و ترکیبی پیشنهاد شده در این مقاله مقدار عملیات به‌روزرسانی را به طور چشمگیری کاهش می‌دهد. در شرایطی که به طور عمده نیاز برنامه‌های کاربردی از نقطه نظر سازگاری تضمین می‌شود. این روش غیرهمزمان و ترکیبی (مبتنی بر فشار و کشیدن) است. در این مقاله برای ارزیابی از شبیه‌ساز optorsim بهره برده شده است.

در [۴۰]، یک مدل تحلیلی سازگاری غیرهمزمان مبتنی بر صف‌بندی M/M/1 برای سیستم ذخیره‌سازی ابر خصوصی پیشنهاد شده است. در روش پیشنهادی برای رسیدگی کردن به تضادهای به‌روزرسانی، از روش احتمال حذف بر اساس درخواست‌های به روز شده استفاده شده است. علاوه بر این، یک روش دیگر نیز به منظور بهبود خوانایی پس از به‌روزرسانی و سازگاری ذخیره‌سازی در محیط ابر ارائه شده است.

در [۴۱]، یک روش جدید غیرهمزمان و ترکیبی در ابر به نام هارمونی پیشنهاد شده است که به صورت انطباقی سطح سازگاری را در زمان اجرا بر طبق نیازمندی‌های برنامه تعیین می‌کند. ایده اصلی پشت هارمونی ساخت یک مدل ارزیابی هوشمند از خواندن‌های قدیمی است. در نتیجه، سازگاری مورد نظر برنامه‌های کاربردی را در حال دستیابی به عملکرد خوب برآورده می‌کند. نتایج نشان داده که هارمونی می‌تواند به عملکردی خوب بدون تجاوز از تعداد قابل تحمل از خواندن‌های قدیمی دست یابد.

در [۴۲]، با استفاده از VFC³ (Versatile Framework for Consistency)، یک مدل سازگاری غیرهمزمان جدید برای داده‌های تکثیر شده در سراسر مراکز داده ابر با چارچوب و پشتیبانی کتابخانه در پیشنهاد شده تا درجه زیادی از سازگاری را برای معانی مختلف داده به اجرا درآورد. این مقاله یک نمونه معماری از چارچوب VFC³ را پیاده‌سازی و ارزیابی کرده و نتایج امیدوارکننده‌ای را آشکار کرده است. این روش در بهبود کیفیت خدمت، کاهش زمان تأخیر، پهنای باند و افزایش توان برای یک مجموعه‌ای از بارکاری کلیدی (و

پشتیبانی می‌کند. در سمت سرور، از سازگاری علیت به موقع (TCC) نیز پشتیبانی می‌کند. علاوه بر این، از رویکرد مشتری متمرکز قوی‌تر است و نسبت به رویکردهای داده‌متمركز انعطاف پذیرتر است. با وجود تحمل پارتیشن، روش پیشنهادی ثبات و در دسترس بودن داده‌ها را برآورده می‌کند. این مقاله به منظور ارزیابی بارهای کاری متفاوت، روی خوشه‌های کاساندررا اجرا شده است. نتایج تجربی مبتنی بر مقایسه بین روش پیشنهادی با ALL، ONE، QUORUM، و همچنین سازگاری‌های CC، در یک خوشه کاساندررا با ۲۴ گره، گواهی می‌دهد که رویکرد پیشنهادی به طور متوسط نرخ خواندن کهنه را تا ۲۴ درصد در حجم کار کاهش داده است.

توسعه سیستم‌های چند ابری با سطح قابل قبولی از تاخیرهای خدماتی که به طور قابل توجهی بر کیفیت تجربه کاربران نهایی تأثیر می‌گذارد، یک مشکل تحقیقاتی باز است. در [۴۹]، یک مدل شبیه‌سازی سازگاری داده برای یک سیستم چند ابری با یک میان‌افزار توزیع‌شده جغرافیایی را در Matlab ارائه می‌شود. این روش به منظور تخمین زمان برای نوشتن به‌روزرسانی‌های داده‌ای که توسط کاربران نهایی در پایگاه‌های داده همه ارائه‌دهندگان خدمات ابری آغاز شده است پیشنهاد می‌شود. نتایج شبیه‌سازی نشان می‌دهد که مدل پیشنهادی می‌تواند در انتخاب یک پروتکل سازگاری داده‌ای که بر معماری یک سیستم چند ابری تأثیر می‌گذارد مفید باشد.

در [۵۰]، یک رویکرد سازگاری پویا، به نام سازگاری دینامیک مه ابری (CFDC^۱) پیشنهاد می‌شود. این روش مزایای رویکردهای خوش‌بینانه و بدبینانه را برای دستیابی و اطمینان از خدمات سازگاری محلی و جهانی در سیستم‌های ابر و مه ترکیب می‌کند. رویکرد CFDC پیشنهادی یک راه حل مناسب برای افزایش مقیاس محیط ابر و مه فراهم می‌کند تا عملکرد و کیفیت خدمات آن سیستم‌ها را به حداکثر برساند.

در مقاله [۵۱]، یک روش جایگذاری تکثیر پویا مبتنی بر ریسک اعتباری گره به نام RPCC در محیط ابر طراحی و اجرا شده است. هدف این روش استفاده از قرار دادن تکثیر برای متعادل کردن رابطه بین عملکرد سیستم و سازگاری تکثیر است. برای رسیدن به هدف، RPCC از روش تطبیقی غیرمتمركز برای ایجاد و قرار دادن تکثیر به صورت پویا استفاده می‌کند. این روش فایل‌هایی با محبوبیت بالا و نرخ به‌روزرسانی پایین را برای ایجاد تکثیر در هنگام ایجاد کپی‌های جدید انتخاب می‌کند. در همین حال، نقشه ریسک اعتباری را با توجه به بار گره، درجه گره و تاخیر دسترسی ترسیم

ارزیابی و اندازه‌گیری سازگاری واقعی مشاهده شده به وسیله مشتری، هنگام استفاده از انبارهای ارزش کلیدی می‌پردازد.

در [۴۶]، سازگاری علیت برای سیستم‌های توزیعی بزرگ مقیاس ابر مورد بررسی قرار گرفته است. اغلب کارهای پیشین سازگاری برای تکثیر کامل فایل مطرح شده است و تعداد کمی به سازگاری در حالت تکثیر جزئی پرداخته‌اند. در این مقاله دو الگوریتم ضعیف و تنبل برای حل چالش‌های موجود در این حیطه پیشنهاد شده است. این مقاله دو الگوریتم پیشنهادی خود را با نام‌های Full-Track و Opt-Track برای تکثیر داده به صورت جزئی ارائه داده‌اند. الگوریتم اول بهینه است. بهینگی از این جنبه است که در این الگوریتم، هر بار که علیت داده از بین می‌رود، داده‌ها در نزدیک‌ترین فاصله زمانی به‌روزرسانی می‌شوند. الگوریتم دوم نیز بهینه است به این معنی که اندازه فراداده حمل شده در پیام‌ها و ذخیره شده در سیاه‌های مربوطه به صورت محلی را کاهش می‌دهد. علاوه بر این، الگوریتم سومی مشتق شده از الگوریتم دوم ارائه می‌شود که سایز پیام، سربار، زمان پردازش و هزینه ذخیره‌سازی محلی در هر سایت را در حالت تکثیر کامل فایل را نیز کاهش می‌دهد. نتایج شبیه‌سازی حکایت از کاهش سربار در پروتکل‌های سازگاری تکثیر جزئی (Opt-Track و Opt-Track) نسبت به پروتکل‌های سازگاری تکثیر کامل (Opt-Track-CRP) در هزینه‌های انتقال و ذخیره‌سازی دارد. زمانی که سایر فایل‌ها بالا و میزان بار کاری نیز افزایش می‌یابد روش‌های بهینه (Opt-Track و Opt-Track) پیشنهادی در تکثیر جزئی موفقیت بیشتری از خود نشان داده‌اند.

در [۴۷]، یک الگوی تراکنش غیرهمزمان جدید برای سیستم توزیعی ابر ارائه شده است، که به طراحان این امکان را می‌دهد به تضمین سازگاری روی داده‌ها به جای سطح تراکنش بپردازند و همچنین روش پیشنهادی آنان این امکان را می‌دهد که تضمین سازگاری در زمان اجرا به صورت خودکار تغییر کند. آزمایش‌ها نشان داده که استراتژی تطبیقی ارائه شده در این مقاله به کاهش قابل توجه زمان پاسخ و هزینه‌ها از جمله هزینه جریمه ناسازگاری منجر می‌شود. این روش به طور قابل توجهی هزینه‌های کلی را کاهش می‌دهد.

در [۴۸]، روشی به نام سازگاری علیت زمان‌بندی دقیق (STCC) را به عنوان یک مدل سازگاری ترکیبی در محیط ابر ارائه می‌کند. در این مدل سازگاری در دو سمت مشتری و سمت سرور در نظر گرفته شده است. در سمت مشتری، این مدل از روش‌های سازگاری خواندن یکنواخت، نوشتن یکنواخت، و نوشتن به دنبال خواندن

^۱ Cloud Fog Dynamic Consistency

بهترین انتخاب برای این محیط پویا خواهد بود. مدل‌های سازگاری‌های تطبیقی با توجه به شرایط مدل مناسب سازگاری تعیین می‌کنند.

۴-۲- سازگاری تکثیر داده در اینترنت اشیا و محاسبات مه

اینترنت اشیا شبکه‌ای از دستگاه‌های هوشمند متصل است. سنسورها، گوشی‌های هوشمند، دوربین‌ها و وسایل نقلیه نمونه‌هایی از این دستگاه‌های هوشمند هستند [۵۳] و [۵۸]. امروزه تعداد دستگاه‌های هوشمند و حجم داده‌های تولید شده در حال افزایش است. تخمین زده می‌شود که تعداد دستگاه‌های اینترنت اشیا (IoT) تا سال ۲۰۲۵ از ۷۵,۴۴ میلیون نفر فراتر رود [۵۹]. ابر طیف گسترده‌ای از نیازهای اینترنت اشیا را فراهم می‌کند. اما، بسیاری از برنامه‌های کاربردی در اینترنت اشیا نیاز به ذخیره و مدیریت داده‌های جریان، پردازش وظایف بلادرنگ، کاهش تأخیر، تضمین امنیت و مقیاس‌پذیری، و مدیریت تعداد زیادی از درخواست‌های مشتری حتی بدون استفاده از اینترنت دارند. رایانش ابری به دلیل متمرکز بودن، فاصله جغرافیایی بین منابع و درخواست‌کننده و حجم بالای ترافیک قادر به پاسخگویی به همه نیازهای برنامه‌های کاربردی در اینترنت اشیا نیست. به اینترنت هم نیاز دارد. امروزه، محاسبات مه به عنوان یک الگوی محاسباتی توزیع شده برای غلبه بر کاستی‌های ابر در حال ظهور است. مه خدمات را در لبه شبکه می‌آورد. بنابراین دسترسی به خدماتی با تأخیر کم را فراهم می‌کند. باید توجه داشته باشیم که مه جایگزینی برای ابر نیست. آن با چالش‌هایی مانند مدیریت منابع متنوع، امنیت، تأخیر، مقیاس‌پذیری، پویایی، پیچیدگی بالا، و ناهمگونی مواجهه است. از این رو، سازگاری مشابه در محاسبات مه یکی از مسائل مهم در عملکرد برنامه‌های کاربردی اینترنت اشیا است. از سوی دیگر، برنامه‌های کاربردی جدید در اینترنت اشیا الزامات سازگاری جدیدی دارند. به عنوان مثال، اکثر راه‌حل‌های سازگاری فعلی در ابر، زمینه داده‌ها و مشتریان را در نظر نمی‌گیرند و فقط یک تنظیم سازگاری جهانی را ارائه می‌دهند. آنها تمام درخواست‌های مشتری را به عنوان یک جعبه سیاه در نظر می‌گیرند [۶۰]. در ابر، همه مشتریان دارای ضمانت‌های یکسان خواندن و نوشتن هستند. در مقابل در مه، این فرض ممکن است درست نباشد و زمینه مشتری سازگاری مورد نیاز را تعیین می‌کند. در برنامه مبتنی بر مه، امکان استفاده از زمینه، به ویژه مکان وجود دارد. باید بتواند سازگاری زمینه را در سطح عملیات (خواندن و نوشتن) به طور جداگانه فراهم کند. موقعیت مکانی یکی از حساس‌ترین کلاس‌های زمینه در مه

می‌کند. RPCC گره‌های بهینه را برای قرار دادن تکثیرها بر اساس نقشه اعتباری انتخاب می‌کند. آزمایش‌های گسترده نشان می‌دهد که RPCC می‌تواند عملکردهای برتری را در افزایش احتمال به‌روزرسانی سازگاری به طور متوسط حدود ۵ درصد، کاهش تأخیر دسترسی حدود ۱۰ درصد و تعادل بار بالاتر نسبت به سایر روش‌های مشابه به دست آورد.

۴-۱-۱- محبوب‌ترین روش‌های سازگاری‌ها در ابر

در ابر روش‌های سازگاری متنوعی توسط محققان ارائه شده است. با این وجود روش‌های سازگاری مبتنی بر روش سازگاری نهایی یا احتمالی از محبوبیت بیشتری در ابر برخوردار هستند. در سازگاری احتمالی همه تکثیرها فوراً به‌روزرسانی نمی‌شود، بنابراین سیستم اختلاف تکثیرها را تحمل می‌کند؛ اما قول می‌دهد همه تکثیرها در یک زمان خاص سازگار باشند. در واقع سازگاری با تأخیر انداختن به‌روزرسانی‌ها، باعث افزایش سرعت دسترسی کاهش مصرف منابع شبکه می‌شود. در واقع نسبت به سازگاری‌های قوی هزینه‌ها به‌روزرسانی را کاهش می‌دهد.

به همین دلیل ارائه‌دهندگان ابر به منظور کاهش هزینه تمایل بیشتری به سازگاری نهایی دارند. سازگاری احتمالی اجازه می‌دهد سیستم برخی داده‌های قدیمی را برگرداند. اما نهایتاً تضمین می‌کند که داده‌ها در نهایت سازگار شوند. بسیاری از سیستم‌های امروزی مثل [۵۲] Cassandra، [۲۹] Amazon Dynamo، [۵۳] Google Big Table، [۵۴] Volodemort، [۵۵] Hbase، [۵۶] Riak مبتنی بر این راه‌حل‌های سازگاری نهایی هستند. البته این سیستم‌ها در شرایط خواندن و نوشتن‌های زیاد، عملکردشان به طور چشم‌گیری افت می‌کند که این میزان کاهش عملکرد در برخی از سیستم‌ها تا ۶۶,۶ درصد نیز گزارش شده است که بسیار نگران‌کننده است [۵۷]. سازگاری قوی و سازگاری احتمالی هر دو فقط برای شرایطی خاص مناسب هستند. در محیط محاسبات ابری، مشتریان ذخیره‌سازی در ابر، متنوع هستند. از طرفی برنامه‌های کاربردی به یک نوع ثابت و مشخص از سازگاری قوی یا سازگاری احتمالی، نیاز ندارند. نیاز سازگاری یک برنامه کاربردی در دوره‌های زمانی مختلف متغیر است. به عنوان مثال نرخ خواندن و نوشتن در یک فروشگاه اینترنتی، در بازه زمانی تعطیلات نسبت به زمان‌های دیگر بیشتر خواهد بود. لذا در این بازه به سطح سازگاری قوی‌تری نسبت به زمان‌های دیگر احساس می‌شود. استراتژی سازگاری تکثیر باید برای برنامه‌های کاربردی در شرایط مختلف مناسب باشد؛ و با تغییر الگوی دسترسی برنامه‌های کاربردی وفق‌پذیر باشد؛ بنابراین، سازگاری تطبیقی تکثیر

همگام‌سازی داده‌ها تنظیم شده‌اند. در این روش کلاینت مستقیماً با گره‌های لبه تعامل دارد که تأخیر پاسخ تعامل با ابر DC را کاهش می‌دهد. ارزیابی PGCE در مقایسه با مدل‌های موجود نشان می‌دهد که عملکرد بهتری در تأخیر پاسخ و توان عملیاتی دارد.

در [۶۵]، مدلی مشابه روش ارائه شده در [۶۴] ارائه شده است. این روش یک مدل سازگاری علیت از همکاری لبه-ابر مبتنی بر پروتکل گروه‌بندی است. مراکز داده در این روش نیز با جدول هش پارتیشن‌بندی می‌شوند و تکثیر داده به صورت جزئی در گره‌های لبه انجام می‌شود. این روش در محیط همکاری edge-cloud در همان زمان، یک الگوریتم همگام‌سازی گروهی به نام Imp_Paxos اجرا می‌کند، در این روش به‌روزرسانی فقط نیاز به همگام‌سازی با گروه اصلی دارد که باعث کاهش تأخیر مشاهده به‌روزرسانی و کاهش سربار همگام‌سازی داده‌ها می‌شود.

در [۶۶]، یک روش مدیریت منابع و یک الگوریتم مدیریت تکثیر، که شامل تخصیص تکثیر و استراتژی حفظ سازگاری پیشنهاد شده است. در این روش‌ها با تکیه بر هزینه مالی به‌روزرسانی، سعی دارند هزینه‌های زمانی و هزینه از جمله هزینه‌های مالی سازگاری تکثیرها را برای مشتریان کاهش دهند. آزمایش‌ها نشان می‌دهد علاوه بر این الگوریتم مدیریت تکثیر به طور موثر زمان انتقال داده و سربار ذخیره‌سازی را کاهش دهد.

در [۵۰]، به منظور کاهش اتلاف منابع ذخیره‌سازی، یک استراتژی ایجاد تکثیر پویا بر اساس محبوبیت بلوک و بار گره (DRC-BN) پیشنهاد شده است. همچنین به منظور حل مشکل ناهماهنگی داده‌ها ناشی از همزمانی چند کاربر، یک استراتژی نگهداری سازگاری تکثیر بر اساس الگوریتم Fast Paxos (RC-FP) پیشنهاد شده است. استراتژی RC-FP عملکرد جامع گره‌ها را در نظر می‌گیرد و گره با عملکرد بهتر را با توجه به عملکرد جامع گره‌ها به عنوان رهبر انتخاب می‌کند و زمان نگهداری سازگاری را کاهش می‌دهد. نتایج تجربی نشان می‌دهد که استراتژی‌های پیشنهادی از نظر زمان نگهداری سازگاری و توان عملیاتی گره عملکرد خوبی دارند. در عین حال، استراتژی‌های پیشنهادی هزینه‌های کلی را نیز کاهش می‌دهند.

مقاله [۶۷]، به بررسی ایجاد تکثیر پویا، بازیابی سریع تکثیر و حفظ سازگاری تطبیقی در سیستم edg cloud می‌پردازد. قبل از حل مشکل ناسازگاری داده‌ها ناشی از به‌روز رسانی مکرر تکثیرها، استراتژی بازیابی سریع تکثیر را در نظر می‌گیرد. روش سازگاری پیشنهادی با نام ACP-IMP برای حل مشکل سازگاری شبیه‌سازی در محیط‌هایی با نرخ شکست بالای شبکه و گره پیشنهاد

است [۶۱]. سازگاری منطقه‌ای نمونه‌ای از سازگاری مبتنی بر زمینه مکان است که اطلاعات مربوط به سطوح مورد نیاز سازگاری را جمع می‌کند. بر اساس این اطلاعات، سطح مناسبی از سازگاری ارائه می‌شود. برنامه‌های کاربردی ترافیک تمایل دارند از این ثبات استفاده کنند [۶۲].

در [۶۳]، یک طرح مدیریت سازگاری حافظه نهان مبتنی بر محبوبیت کارآمد را پیشنهاد شده است، که هدف آن تضمین تازگی داده‌های اینترنت اشیا بازگردانده شده توسط روترهای مسیریاب و جلوگیری از هزینه‌های سیگنالینگ سنگین معرفی شده در همان زمان است. شبیه‌سازی‌های گسترده‌ای تحت توپولوژی‌های بدون ترس و باینری درختی در دنیای واقعی انجام شد، و نتایج مربوطه کارایی طرح پیشنهادی را در حذف به‌موقع داده‌های اینترنت اشیا منسوخ ذخیره‌شده توسط حافظه پنهان درون شبکه CCN ثابت کرده‌اند.

در [۸]، دو استراتژی برای مدیریت تکثیر و سازگار نگهداشتن داده‌های اینترنت اشیا و ثبات در زیرساخت‌های مه ارائه شده است. استراتژی‌ها برای هر داده، شماره تکثیر مناسب و مکان آن‌ها را انتخاب می‌کنند تا تأخیر دسترسی به داده‌ها و هزینه همگام‌سازی ماکت‌ها را کاهش دهند. این کار با رعایت سطح سازگاری لازم انجام می‌شود. همچنین، یک پلتفرم ارزیابی مبتنی بر شبیه‌ساز iFogSim را پیشنهاد می‌شود تا کاربران بتوانند استراتژی‌های خود را برای تکثیر داده‌های اینترنت اشیا و مدیریت سازگاری پیاده‌سازی و آزمایش کنند. آزمایش‌های نشان می‌دهد که هنگام استفاده از استراتژی‌های پیشنهادی، تأخیر خدمت را می‌توان تا ۳۰ درصد در مورد زیرساخت‌های Fog کوچک و ۱۳ درصد در مورد زیرساخت‌های مه در مقیاس بزرگ در مقایسه با iFogStor، کاهش داد. یک استراتژی پیشرفته و محبوب جاگذاری داده در زیرساخت مه است.

در [۶۴]، یک مدل مبتنی بر سازگاری علیت برای ذخیره‌سازی توزیع شده بر اساس تکثیر جغرافیایی جزئی و ساختار همکاری به نام Cloud-Edge(PGCE) پیشنهاد شده است. این مدل مبتنی بر معماری شبکه توزیع شده همکاری Cloud-Edge است، و مجموعه داده‌های ابری توسط یک تابع هش به زیر مجموعه‌های متعدد تقسیم می‌شود تا این زیرمجموعه‌ها در گره‌های لبه‌ای که نزدیک به شبکه کاربر هستند برای تحقق بخشیدن به تکرار جغرافیایی جزئی، ذخیره شود. در همان زمان، مکانیسم تثبیت مهر زمانی و خدمت پردازش فراداده برای اجرای سازگاری داده‌ها بین گره‌ها با فرض اطمینان از علت، کاهش هزینه‌های سربار پردازش ابر داده و

نسخه‌ها می‌باشد.

در [۳۴]، الگوریتم مدیریت سازگار موثری در گرید داده ارائه می‌شود که در آن داده‌های تکثیر را ثابت نگه می‌دارد. در این روش، گره‌ای که حاوی تکثیر است، وظیفه مدیریت سازگاری را بر عهده دارد. در هر گره، نرخ به‌روزرسانی تکثیر بر اساس میزان تغییر در فایل داده اصلی و نرخ درخواست دسترسی به نسخه مشابه موجود محاسبه می‌شود. مدل پیشنهادی تأخیرهای دسترسی را کاهش می‌دهد. نتایج تجربی نشان می‌دهد که مدل پیشنهادی در مقایسه با رویکردهای خوش بینانه و بدبینانه از نظر زمان اجرای کارو تعداد به‌روزرسانی‌ها عملکرد بهتری دارد.

در جدول ۲ مقایسه‌ای از برخی از مدل‌های سازگاری ارائه شده توسط محققان از منظر نوع سیستم توزیعی، دسته‌بندی مدل سازگاری ارائه شده، محیط پیاده‌سازی یا شبیه‌سازی، پارامترهای مهم دخالت داده شده در مدل سازگاری و پارامترهایی بهبود داده شده آورده شده است.

جدول ۲. مقایسه مدل‌های سازگاری پیشین.

شماره مرجع	نوع سیستم توزیعی	دسته‌ی مدل سازگاری	محیط پیاده‌سازی یا شبیه‌ساز	پارامترهای لحاظ شده در تصمیم‌گیری	پارامترهای بهبود داده شده
[۶۸]	نظیر به نظیر	ترکیبی، تنبل، غیرهمزمان، توزیع شده	شبیه‌ساز جدید مبتنی بر رویداد (Event Based)	تعداد به‌روزرسانی‌ها وضعیت قطعی و وصلی گره‌ها	نرخ پرسش معتبر اما غلط ^۲ ، نرخ دانلود معتبر اما غلط ^۳
[۷۰]	گرید داده	تنبل، نیمه‌متمرکز، غیرهمزمان، ترکیبی، (ترکیبی از سازگاری ضعیف و قوی)	شبیه‌ساز Optorsim	فرکانس دسترسی‌ها	زمان پاسخ، تعداد برخوردها
[۷۲]	گرید داده	غیرهمزمان، ترکیبی، (ترکیبی از سازگاری قوی و ضعیف)	شبیه‌ساز Optorsim	فرکانس دستیابی به فایل و ظرفیت رسانه ذخیره‌سازی	مدت زمان تکثیر
[۵۴]	گرید داده	مبتنی بر فشار، نیمه متمرکز، همزمان، سازگاری قوی	شبیه‌ساز Optorsim	درجه همسایگی گره‌های در ساخت درخت انتشار	زمان پاسخ
[۷۳]	گرید داده	غیرهمزمان، (ترکیبی از قوی و ضعیف)	شبیه‌ساز به زبان جاوا	بارکاری و کیفیت خدمت	تعداد واگرای‌ها و زمان پاسخ
[۴۷]	ابر محاسباتی	غیرهمزمان، ترکیبی از سازگاری قوی و ضعیف	پیاده‌سازی بر روی سیستم ذخیره‌سازی Amazon's Simple Storage Service (S3)	احتمال برخورد به‌روزرسانی‌ها در هر رکورد، تعداد به‌روزرسانی هزینه به‌روزرسانی	هزینه در هر تراکنش ^۴ ، هزینه در زمان اجرا در هر تراکنش، هزینه سراسری در هر تراکنش
[۳۹]	ابر محاسباتی	غیرهمزمان، ترکیبی، (ترکیبی از قوی و ضعیف)، تنبل	شبیه‌ساز Optorsim	فرکانس به‌روز رسانی، فرکانس دسترسی، نوع تکثیر (اصلی یا ثانویه)	درصد دسترسی به داده به‌روز، میزان حجم تراکنش‌ها،
[۴۰]	ابر محاسباتی	تنبل، ضعیف، غیرهمزمان	مدل ارزیابی مبتنی بر صفت‌بندی MMI	متوسط نرخ خدمت، نرخ به‌روزرسانی، تعداد تکثیرها	احتمال خواندن‌های درست

³ Downloading Query False Valid Rate

⁴ Cost Per Transaction

¹ Computing Element

² Query False Valid Rate

[۴۲]	ابر محاسباتی	غیرهمزمان، (ترکیبی از قوی و ضعیف)، تنبل	یک نمونه اولیه پیاده‌سازی شده به زبان جاوا	حداکثر تفاوت نسبی در محتویات دو داده شی داده، مدت زمانی که تکثیر می‌تواند داده قدیمی را تحمل کند، تعداد به‌روزرسانی‌های ممکن برای یکشی بدون به روز کردن تکثیرها	کارایی حافظه نهان، کارایی در میزان تأخیر و گذردهی، توان عملیاتی، میزان مصرف پهنای باند
[۴۳]	ابر محاسباتی	غیرهمزمان، (ترکیبی از قوی و ضعیف)، تنبل	پیاده‌سازی بر روی سیستم ذخیره‌سازی Cassandra	تأخیر شبکه، نرخ خواندن و نوشتن	بهره‌وری هزینه سازگاری ^۱ ، هزینه مالی و نرخ خواندن‌های تازه، هزینه، نرخ خواندن داده‌های قدیمی
[۵۷]	ابر محاسباتی	غیرهمزمان، ضعیف، ترکیبی، تنبل	ذکر نشده	کلاک سیستم، آخرین زمان نوشتن	سربرار تبادل پیام، زمان انتقال، مدت زمان میانجی‌گری، پیچیدگی زمانی و فضایی
[۴۱]	ابر محاسباتی	غیرهمزمان، (ترکیبی از قوی و ضعیف)، تنبل	در محیط Grid'5000 و Amazon EC2	نرخ تحمل خواندن داده‌های قدیمی توجه به داده‌های قدیمی و ساخت مدل مبنی بر این خواندن‌های قدیمی و مبتنی بر نیاز برنامه ارائه یک مدل ارزیابی هوشمند	توان عملیاتی، تأخیر خواندن، تعداد خواندن‌های قدیمی،
[۷۱]	نظیر به نظیر	غیرهمزمان، (ترکیبی از قوی و ضعیف)، تنبل	Peersim	نرخ خواندن و نوشتن، فرم‌های به‌روزرسانی	نرخ پیام‌های به‌روز، زمان به‌روزرسانی، نرخ موفقیت به‌روزرسانی
[۳۴]	گريد داده	غیرهمزمان، (ترکیبی از قوی و ضعیف)، تنبل	Optorsim	نرخ درخواست، زمان درخواست‌های قبلی، آخرین زمان درخواست داده	میانگین زمان اجرای درخواست‌ها،
[۶۷]	لبه-ابر	غیرهمزمان	متلب	تعداد پیام‌ها، هزینه ارتباط ورزولوشن‌ها، زمان خدمت	میانگین دسترسی-پذیری، زمان پاسخ، تعادل بار، زمان اتمام کار
[۴۸]	ابر	غیر همزمان، ضعیف	خوشه‌های کاساندر	برچسب زمانی خواند و برچسب زمانی نوشتن	تأخیر عملیات خواندن، توان عملیاتی، بار کاری، تعداد خواندن‌های داده‌های قدیمی
[۶۴]	ابر-لبه	غیر همزمان، ضعیف	شبیه ساز کلودسیم	مهر زمانی، فاصله، علیت	هزینه‌های سربرار پردازش ابر داده و همگام‌سازی داده‌ها، تأخیر زمان پاسخ
[۶۵]	ابر-لبه	ترکیبی، ضعیف	شبیه ساز کلودسیم	علیت، تثبیت مهر زمانی،	توان عملیاتی، تأخیر عملیات و تأخیر دید به‌روزرسانی
[۶۶]	ابر-لبه	غیر همزمان، ضعیف	پیاده‌سازی در مراکز داده Hangzhou	منشا داده- زمان عملیات نوشتن، اطلاعات به‌روزرسانی	میانگین زمان بررسی ثبات و بازیابی، میانگین زمان حفظ سازگاری، تعداد دسترسیها، تعداد تکثیرها، زمان انتقال داده
[۷۴]	گريد	قوی-ضعیف-همزمان غیر همزمان	شبیه‌سازی به زبان جاوا در GlouboS	آخرین زمان به‌روزرسانی	Load balancing and QoS
[۶۷]	ابر-لبه	غیر همزمان-قوی	OptorSim	متوسط تأخیر درخواست، درجه گره، همسایگی	نسبت بار متوسط، احتمال آپدیت موفقیت آمیز،
[۵۰]	ابر-مه	غیر همزمان، ضعیف	شبیه ساز CloudSim	مهر زمانی(آخرین زمان به‌روزرسانی)	تعداد خواندن‌های ناسازگار، میانگین تعداد موارد کیفیت خدمت برآورده شده

^۱ Consistency-Cost Efficiency

۵- موضوعات باز و کارهای آتی

اکثر کارهای ارائه شده با موضوع مدل‌های سازگاری، اغلب در شبیه‌سازها مورد تست و ارزیابی قرار گرفته‌اند. ارزیابی روش‌های پیشنهادی در حیطه سازگاری در محیط‌های واقعی سیستم توزیعی یکی از کارهای آتی است که مورد توجه محققان در این حیطه است. اکثر کارهای ارائه شده به اندازه کافی سازگار و وفق‌پذیر با محیط نیستند. ارائه مدل‌های سازگارتر منعطف‌تر و وفق‌پذیرتر با محیط پویای سیستم‌های توزیعی یکی دیگر از مباحث نیمه کاره و مطلوب محققان این حیطه است. اغلب کارهای ارائه شده هزینه مالی را برای درخواست‌کنندگان سازگاری مد نظر قرار نداده‌اند. در کارهای پیشین اغلب مدل ارائه شده تک سروری بوده‌اند و تعداد کمی از کارها به مدل‌های چند سروری پرداخته‌اند. وجود چند سرور تکثیر باعث برخورد داده خواهد شد. پرداختن به این مسئله یکی دیگر از موضوعات باز در این حیطه است. ترکیب الگوریتم‌های تکثیر پویا با مدل‌های سازگاری که هر دو با هم به بهترین نحوه عمل کنند یکی دیگر از موضوعات در کارهای آتی می‌تواند باشد. چراکه اغلب الگوریتم‌های تکثیر فرض کرده‌اند داده‌ها فقط خواندنی هستند و همچنین اغلب مدل‌های سازگاری پیشین مستقل از الگوریتم تکثیر عمل می‌کنند. برنامه‌های کاربردی نیازمند سطوح مختلفی از سازگاری هستند. لذا توجه به پارامترهای مختلف در برآورده کردن نیازی‌های سازگاری برنامه‌های کاربردی در تعیین سطوح سازگاری برای داده‌های برنامه کاربردی یکی دیگر شکاف‌ها در کارهای پیشین است. یکی از مسائلی که از دید محققان در کارهای پیشین پنهان مانده است؛ اهمیت توجه به محبوبیت یک فایل و تعیین سطوح سازگاری بر اساس تخمین محبوبیت یک فایل در زمان آتی است. در واقع فایل‌های محبوب به دلیل نیاز بیشتر در دسترسی به نسخه به روز آنها، نیازمند سازگاری قوی‌تری هستند. در واقع با پیشگویی دسترسی آتی به یک فایل می‌توان احتمال نیاز به سازگاری قوی برای آن فایل را پیشگویی کرد و با به روز کردن نسخه محبوب زمان پاسخ و کارایی سیستم را افزایش داد. در کارهای پیشین در تعیین سطوح سازگاری برای یک فایل میزان محبوبیت فایل‌ها در سطح یک فایل مطرح شده است. زمانی که یک تکثیر در یک سایت نسبت به سایر تکثیرهای آن فایل بیشتر مورد دسترس قرار می‌گیرد، دادن سطوح سازگاری قوی‌تر می‌تواند موجب افزایش کارایی شود. توجه به محبوبیت تکثیر فایل در یک سایت خاص به صورت جدا در عوض محبوبیت فایل در کل سیستم می‌تواند باعث کاهش دسترسی به داده قدیمی شود. لذا توجه بیشتر به محبوبیت تکثیرها به صورت جدا در مدل‌های پیشنهادی می‌تواند از جمله کارهای آتی

در نظر گرفته شود. در تعداد کمی از کارها مدل سازگاری مبتنی بر یک مدل ریاضی بیان شده است. با توجه به اعتبار قطعی مدل ریاضی، ارائه روش‌های سازگاری مبتنی بر مدل‌های ریاضی می‌تواند به عنوان کارهای آتی در نظر گرفته شود.

امروزه به کاربران به ازای پرداخت هزینه، خدمت ارائه می‌شود. لذا ارایه مدل‌های مبتنی بر پرداخت هزینه را نیز می‌توان از چالش‌های پیشرو در نظر گرفته و در پژوهش‌های آتی مورد بررسی قرار داد. در سیستم‌های توزیعی مدرن مانند ابر ارائه خدمات مبتنی بر پرداخت هزینه مالی است. در واقع کاربران خدمات خود را از فراهم‌کنندگان اجاره می‌کنند. از این رو خدمت باید طوری ارائه شوند که هم برای اجاره‌کننده کم هزینه و هم برای فراهم‌کننده خدمت سودآور باشد. از این رو ارائه مدل‌های سازگاری که به این دو هزینه به طور همزمان توجه کنند از کارهای آتی در این حوزه است.

در تعداد کمی از کارهای قبلی، کیفیت خدمات (QoS) در نظر گرفته شده است. اکثر کارهای ارائه شده فقط زمان پاسخ را به عنوان معیار QoS در نظر می‌گیرند. ارائه روش‌های مدیریت سازگاری که به طور همزمان فهرستی جامع از پارامترهای کیفیت خدمات مانند توان عملیاتی، در دسترس بودن، مقیاس پذیری، امنیت، مصرف انرژی و قابلیت اطمینان را در نظر می‌گیرند در آینده می‌تواند بسیار جالب توجه باشد.

۶- نتیجه‌گیری

محبوبیت سیستم‌های توزیعی روز به روز در حال افزایش است. از طرفی حجم داده‌های تولیدی توسط برنامه‌های کاربردی در سیستم‌های توزیعی و میزان نیاز به دسترسی مؤثر و کارا به داده‌ها نیز در حال افزایش است. تکثیر داده تکنیکی کلیدی برای رسیدن به کارایی و دسترسی‌پذیری است. داده‌ها در سیستم‌های توزیعی توسط کاربران و برنامه‌ها دائماً در حال تغییراند. از این رو حفظ و برقراری سازگاری تکثیرهای داده در این سیستم‌ها یک نیاز حیاتی است. از گذشته دور تا به اکنون، تلاش‌های قابل توجه‌ای برای توسعه استراتژی‌های سازگاری مختلف اختصاص یافته است. نتیجه این تلاش‌ها ظهور استراتژی‌های مختلف بوده است.

در این مقاله یک بررسی کلی در مورد استراتژی‌های سازگاری تکثیر داده در سیستم‌های توزیعی محبوب از جمله گرید، ابر، مه، لبه، نظیر به نظیر انجام شده است. تقسیم‌بندی‌های الگوریتم تکثیر از دیدگاه نوع تکثیر، بخش‌های مدیریت تکثیر، محیط تکثیر، هدف الگوریتم تکثیر و روش ارزیابی الگوریتم تکثیر بیان شد. با توجه به تمرکز بر روی سازگاری تکثیر، مفهوم سازگاری تکثیر بررسی شد و

مراجع

- انواع تقسیم‌بندی‌های موجود از دیدگاه‌های مختلف از جمله خصوصیت، دیدگاه آقای تنبام، میزان همسان بودن نسخه بیان شد. از طرفی در یک طبقه‌بندی دقیق‌تر، ما مدل‌های سازگاری‌های پایه را به پنج دسته داده-متمرکز، مشتری-متمرکز، ترکیبی، جدید و بسط یافته تقسیم کردیم. تقسیم‌بندی مدل‌های سازگاری از نقطه نظر تکنیک به‌روزرسانی ذکر شد. در این بررسی مدل‌های سازگاری به دو دسته همزمان و غیرهمزمان تقسیم شدند. همچنین از نقطه نظر اینکه در انتشار به‌روز رسانی تکثیرها چه کسی شروع‌کننده بوده است (سرور یا مشتری)، روش‌های سازگاری موجود به سه دسته مبتنی بر کشیدن و مبتنی بر فشاردادن، ترکیبی تقسیم‌بندی شدند. علاوه بر این در هر نوع سیستم توزیعی مدل‌های سازگاری ارائه شده توسط محققان مطرح شدند و در جدولی برخی از مدل‌های سازگاری ارائه شده از منظر نوع سیستم توزیعی، دسته‌بندی مدل سازگاری ارائه شده، محیط پیاده‌سازی و شبیه‌سازی، پارامترهای مهم دخالت داده شده در مدل سازگاری و پارامترهایی بهبود داده شده مقایسه شدند.
- بررسی‌ها نشان داد مدل‌های پیشنهادی محققان در تمام سیستم‌های توزیعی اغلب به یک یا چند مورد از موارد زیر می‌پرداختند: قابلیت اطمینان، مقیاس‌پذیری، تحمل خطا، متعادل‌سازی بار و غیره. این استراتژی‌ها سعی می‌کردند پارامترهای خاصی مانند تازگی (قدیمی نبودن)، گستره ساخت، هزینه ارتباط، هزینه به‌روزرسانی، زمان پاسخ‌دهی، مصرف پهنای باند، تأخیر دسترسی، متعادل‌سازی بار، هزینه تعمیر و نگهداری، زمان اجرای کار، تحمل خطا، و قرار دادن تکثیر استراتژیک را بهبود بخشند. هدف اکثریت استراتژی‌های همگام‌سازی یا سازگاری تکثیر داده دستیابی به صحت و کیفیت خدمات تکثیرهای داده است. همچنین بررسی‌ها نشان می‌دهد به دلیل هزینه‌های بالا سازگاری‌های قوی، تمایل به سمت استفاده از سازگاری‌های ضعیف در سیستم‌های توزیعی مانند ابر بسیار بالا است. البته میزان کاهش مطلوب سطح سازگاری تا حد بسیار زیادی بستگی به الگوهای دسترسی برنامه کاربردی و نرخ به‌روزرسانی داده‌های تکثیر شده و همچنین هدف به کارگیری آن داده‌ها دارد. لذا خدمت‌گیرندگان نیازمند سطوح سازگاری متفاوتی هستند. در این مقاله در بخشی جدا به کارهای آتی در این حوزه پرداخته شد. مدیریت سازگاری تکثیر مبتنی بر کیفیت خدمت، مدیریت سازگاری تکثیر مبتنی بر هزینه‌های مالی، مدیریت پویای سازگاری تکثیر مبتنی بر نیاز پویای برنامه‌های کاربردی، برخی از این کارهای آتی مطرح شده است.
- [1] R. Moore, C. Baru, A. Rajasekar, R. Marciano, and M. Wan, "Data Intensive Computing, In "The Grid: Blueprint for a New Computing Infrastructure", eds. I. Foster and C. Kesselman." Morgan Kaufmann, San Francisco, 1999.
 - [2] M. Beigrezaei, A. Toroghi Haghghat, and S. Leili Mirtaheri, "Minimizing data access latency in data grids by neighborhood-based data replication and job scheduling," *Int. J. Commun. Syst.*, p
 - [3] A. M. Rahmani, Z. Fadaie, and A. T. Chronopoulos, "Data placement using Dewey Encoding in a hierarchical data grid," *J. Netw. Comput. Appl.*, vol. 49, pp. 88–98, 2015.
 - [4] U. Tos, R. Mokadem, A. Hameurlain, and T. Ayav, "Achieving query performance in the cloud via a cost-effective data replication strategy," *Soft Comput.*, vol. 25, no. 7, pp. 5437–5454, 2021.
 - [5] X. Dong, J. Li, Z. Wu, D. Zhang, and J. Xu, "On dynamic replication strategies in data service grids," in *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, 2008, pp. 155–161.
 - [6] M.-C. Lee, F.-Y. Leu, and Y. Chen, "PFRF: An adaptive data replication algorithm based on star-topology data grids," *Futur. Gener. Comput. Syst.*, vol. 28, no. 7, pp. 1045–1057, 2012.
 - [7] C. Li, M. Song, M. Zhang, and Y. Luo, "Effective replica management for improving reliability and availability in edge-cloud computing environment," *J. Parallel Distrib. Comput.*, 2020.
 - [8] M. I. Naas, L. Lemarchand, P. Raipin, and J. Boukhobza, "IoT data replication and consistency management in fog computing," *J. Grid Comput.*, vol. 19, no. 3, pp. 1–25, 2021.
 - [9] S. Sun, W. Yao, B. Qiao, M. Zong, X. He, and X. Li, "RRSD: A file replication method for ensuring data reliability and reducing storage consumption in a dynamic Cloud-P2P environment," *Futur. Gener. Comput. Syst.*, vol. 100, pp. 844–858, 2019.
 - [10] M. Beigrezaei, A. T. Haghghat, and H. R. Kanan, "A new fuzzy based dynamic data replication algorithm in data grids," in *2013 13th Iranian Conference on Fuzzy Systems (IFSC)*, 2013, pp. 1–5, doi: 10.1109/IFSC.2013.6675676.
 - [11] K. Rajaretnam, M. Rajkumar, and R. Venkatesan, "Rplb: A replica placement algorithm in data grid with load balancing," *Int. Arab J. Inf. Technol.*, vol. 13, no. 6, 2016.
 - [12] H. K. H. So and R. Brodersen, "A unified hardware/software runtime environment for FPGA-based reconfigurable computers using BORPH," *Trans. Embed. Comput. Syst.*, vol. 7, no. 2, pp. 1–28, 2008, doi: 10.1145/1331331.1331338.
 - [13] A. M. Rahmani, L. Azari, and H. A. Daniel, "A file group data replication algorithm for data grids," *J. Grid Comput.*, vol. 15, no. 3, pp. 379–393, 2017.
 - [14] G. Hager and G. Wellein, *Introduction to high performance computing for scientists and engineers*. CRC Press, 2010.
 - [15] M. Beigrezaei, A. T. Haghghat, M. R. Meybodi, and M. Runiassy, "PPRA: A new pre-fetching and prediction based replication algorithm in data grid," in *2016 6th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2016, pp. 257–262.
 - [16] B. L. Chamberlain, D. Callahan, and H. P. Zima, "Parallel programmability and the chapel language," *Int. J. High Perform. Comput. Appl.*, vol. 21, no. 3, pp. 291–312, 2007.
 - [17] L. Azari, A. M. Rahmani, H. A. Daniel, and N. N. Qader, "A data replication algorithm for groups of files in data grids," *J. Parallel Distrib. Comput.*, vol. 113, pp. 115–126, 2018.
 - [18] Y. Saito, "Consistency management in optimistic replication algorithms," *INTERNET, AOnlineU*, vol. 15, pp. 1–18, 2001.

- [39] X. Wang, S. Yang, S. Wang, X. Niu, and J. Xu, "An application-based adaptive replica consistency for cloud storage," in *2010 Ninth International Conference on Grid and Cloud Computing*, 2010, pp. 13–17.
- [40] Y. N. Aye, "Data Consistency on Private Cloud Storage System," *Int. J. Emerg. Trends Technol. Comput. Sci. Vol.*, vol. 1, 2012.
- [41] H.-E. Chihoub, S. Ibrahim, G. Antoniu, and M. S. Perez, "Harmony: Towards automated self-adaptive consistency in cloud storage," in *2012 IEEE International Conference on Cluster Computing*, 2012, pp. 293–301.
- [42] S. Esteves, J. Silva, and L. Veiga, "Quality-of-service for consistency of data geo-replication in cloud computing," in *European Conference on Parallel Processing*, 2012, pp. 285–297.
- [43] H.-E. Chihoub, S. Ibrahim, G. Antoniu, and M. S. Perez, "Consistency in the cloud: When money does matter!," in *2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*, 2013, pp. 352–359.
- [44] Q. Liu, G. Wang, and J. Wu, "Consistency as a service: Auditing cloud consistency," *IEEE Trans. Netw. Serv. Manag.*, vol. 11, no. 1, pp. 25–35, 2014.
- [45] X. Ren, W. Ru-chuan, and K. Qiang, "Efficient model for replica consistency maintenance in data grids," in *International Symposium on Computer Science and its Applications*, 2008, pp. 159–162.
- [46] A. Stiemer, I. Fetai, and H. Schuldt, "Analyzing the performance of data replication and data partitioning in the cloud: the BEOWULF approach," in *2016 IEEE international conference on big data (Big Data)*, 2016, pp. 2837–2846.
- [47] T. Kraska, M. Hentschel, G. Alonso, and D. Kossmann, "Consistency rationing in the cloud: Pay only when it matters," *Proc. VLDB Endow.*, vol. 2, no. 1, pp. 253–264, 2009, doi: 10.14778/1687627.1687657.
- [48] H. N. S. Aldin, H. Deldari, M. H. Moattar, and M. R. Ghods, "Strict timed causal consistency as a hybrid consistency model in the cloud environment," *Futur. Gener. Comput. Syst.*, vol. 105, pp. 259–274, 2020.
- [49] O. Kozina and M. Kozin, "Simulation Model of Data Consistency Protocol for Multicloud Systems," in *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek)*, 2022, pp. 1–4.
- [50] [N. Mostafa, "A dynamic approach for consistency service in cloud and fog environment," in *2020 fifth international conference on fog and mobile edge computing (FMEC)*, 2020, pp. 28–33.
- [51] S. Sun, X. Wang, and F. Zuo, "RPCC: A Replica Placement Method to Alleviate the Replica Consistency under Dynamic Cloud," in *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*, 2020, pp. 729–734.
- [52] A. Lakshman and P. Malik, "Cassandra: a decentralized structured storage system," *ACM SIGOPS Oper. Syst. Rev.*, vol. 44, no. 2, pp. 35–40, 2010.
- [53] [OpenACC Working Group, "The OpenACC Application Programming Interface 3.0," *Openacc.Org Doc.*, pp. 1–118, 2019.
- [54] C.-T. Yang, W.-C. Tsai, T.-T. Chen, and C.-H. Hsu, "A one-way file replica consistency model in data Grids," in *The 2nd IEEE Asia-Pacific Service Computing Conference (APSCC 2007)*, 2007, pp. 364–373.
- [55] M. N. Vora, "Hadoop-HBase for large-scale data," in *Proceedings of 2011 International Conference on Computer Science and Network Technology*, 2011, vol. 1, pp. 601–605.
- [56] [R. Klophaus, "Riak core: Building distributed applications without shared state," in *ACM SIGPLAN Commercial Users of Functional Programming*, 2010, p. 1.
- [19] R. A. Campêlo, M. A. Casanova, D. O. Guedes, and A. H. F. Laender, "A brief survey on replica consistency in cloud environments," *J. Internet Serv. Appl.*, vol. 11, no. 1, pp. 1–13, 2020.
- [20] J. Curtis, "Consistency of data replication protocols in database systems: a review," 2014.
- [21] P. Vashisht, A. Sharma, and R. Kumar, "Strategies for replica consistency in data grid—a comprehensive survey," *Concurr. Comput. Pract. Exp.*, vol. 29, no. 4, p. e3907, 2017.
- [22] K. S. Maabreh, "An Enhanced University Registration Model Using Distributed Database Schema," *KSII Trans. Internet Inf. Syst.*, vol. 13, no. 7, 2019.
- [23] I. Foster, C. Kesselman, and S. Tuecke, "The anatomy of the grid: Enabling scalable virtual organizations," *Int. J. High Perform. Comput. Appl.*, vol. 15, no. 3, pp. 200–222, 2001.
- [24] H. Yu and A. Vahdat, "Design and evaluation of a conit-based continuous consistency model for replicated services," *ACM Trans. Comput. Syst.*, vol. 20, no. 3, pp. 239–282, 2002.
- [25] T. Kraska, M. Hentschel, G. Alonso, and D. Kossmann, "Consistency rationing in the cloud: pay only when it matters," *Proc. VLDB Endow.*, vol. 2, no. 1, pp. 253–264, 2009.
- [26] A. S. Tanenbaum and M. Van Steen, *Distributed systems: principles and paradigms*. Prentice-Hall, 2007.
- [27] J. Du, S. Elnikety, and W. Zwaenepoel, "Clock-SI: Snapshot isolation for partitioned data stores using loosely synchronized clocks," in *2013 IEEE 32nd International Symposium on Reliable Distributed Systems*, 2013, pp. 173–184.
- [28] P. Keleher, A. L. Cox, and W. Zwaenepoel, "Lazy release consistency for software distributed shared memory," *ACM SIGARCH Comput. Archit. News*, vol. 20, no. 2, pp. 13–21, 1992.
- [29] L. Iftode, J. P. Singh, and K. Li, "Scope consistency: A bridge between release consistency and entry consistency," in *Proceedings of the eighth annual ACM symposium on Parallel algorithms and architectures*, 1996, pp. 277–287.
- [30] C. Cachin, I. Keidar, and A. Shraer, "Fork sequential consistency is blocking," *Inf. Process. Lett.*, vol. 109, no. 7, pp. 360–364, 2009.
- [31] C. Cachin, A. Shelat, and A. Shraer, "Efficient fork-linearizable access to untrusted shared memory," in *Proceedings of the twenty-sixth annual ACM symposium on Principles of distributed computing*, 2007, pp. 129–138.
- [32] C. Li, D. Porto, A. Clement, J. Gehrke, N. Preguiça, and R. Rodrigues, "Making geo-replicated systems fast as possible, consistent when necessary," in *Presented as part of the 10th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 12)*, 2012, pp. 265–278.
- [33] [F. J. Torres-Rojas and E. Meneses, "Convergence through a weak consistency model: Timed causal consistency," *CLEI Electron. J.*, vol. 8, no. 2, pp. 1–2, 2005.
- [34] J. Li, M. N. Krohn, D. Mazieres, and D. E. Shasha, "Secure Untrusted Data Repository (SUNDR)," in *Osdi*, 2004, vol. 4, p. 9.
- [35] A. J. Feldman, W. P. Zeller, M. J. Freedman, and E. W. Felten, "SPORC: Group collaboration using untrusted cloud resources," 2010.
- [36] Y. Saito and M. Shapiro, "Optimistic replication," *ACM Comput. Surv.*, vol. 37, no. 1, pp. 42–81, 2005.
- [37] M. Senftleben and K. Schneider, "Operational characterization of weak memory consistency models," in *International Conference on Architecture of Computing Systems*, 2018, pp. 195–208.
- [38] [D. Terry, V. Prabhakaran, R. Kotla, M. Balakrishnan, and M. K. Aguilera, "Transactions with Consistency Choices on Geo-Replicated Cloud Storage."

- experience in edge cloud computing system,” *Inf. Sci. (Ny)*, vol. 516, pp. 33–55, 2020.
- [67] J. Guo, C. Li, and Y. Luo, “Fast replica recovery and adaptive consistency preservation for edge cloud system,” *Soft Comput.*, vol. 24, pp. 14943–14964, 2020.
- [68] J. Lan, X. Liu, P. Shenoy, and K. Ramamritham, “Consistency maintenance in peer-to-peer file sharing networks,” in *Proceedings the Third IEEE Workshop on Internet Applications. WIAPP 2003*, 2003, pp. 90–94.
- [69] H. Shen, “IRM: Integrated file replication and consistency maintenance in P2P systems,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 1, pp. 100–113, 2009.
- [70] R.-S. Chang and J.-S. Chang, “Adaptable replica consistency service for data grids,” in *Third International Conference on Information Technology: New Generations (ITNG’06)*, 2006, pp. 646–651.
- [71] [X. Meng and C. Zhang, “An ant colony model based replica consistency maintenance strategy in unstructured P2P networks,” *Comput. Networks*, vol. 62, pp. 1–11, 2014.
- [72] [S. C. Choi and H. Y. Youn, “Dynamic hybrid replication effectively combining tree and grid topology,” *J. Supercomput.*, vol. 59, pp. 1289–1311, 2012.
- [73] [E. Anderson, X. Li, M. A. Shah, J. Tucek, and J. J. Wylie, “What Consistency Does Your {Key-Value} Store Actually Provide?,” 2010.
- [74] G. Belalem and B. Yagoubi, *Collaborative Negotiation to Resolve Conflicts among Replicas in Data Grids*. INTECH Open Access Publisher, 2010.
- [75] G. Belalem, C. Haddad, and Y. Slimani, “An effective approach for consistency management of replicas in Data Grid,” in *2008 IEEE International Symposium on Parallel and Distributed Processing with Applications*, 2008, pp. 11–18.
- [57] L. Xiong, L. Yang, Y. Tao, J. Xu, and L. Zhao, “Replication strategy for spatiotemporal data based on distributed caching system,” *Sensors*, vol. 18, no. 1, p. 222, 2018.
- [58] M. I. Naas, J. Boukhobza, P. R. Parvedy, and L. Lemarchand, “An extension to ifogsim to enable the design of data placement strategies,” in *2018 IEEE 2nd International Conference on Fog and Edge Computing (ICFEC)*, 2018, pp. 1–8.
- [59] R. Oma, S. Nakamura, D. Duolikun, T. Enokido, and M. Takizawa, “Fault-tolerant fog computing models in the IoT,” in *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 2018, pp. 14–25.
- [60] A. V. Dastjerdi, H. Gupta, R. N. Calheiros, S. K. Ghosh, and R. Buyya, “Fog computing: Principles, architectures, and applications,” in *Internet of things*, Elsevier, 2016, pp. 61–75.
- [61] M. Verma, N. Bhardwaj, and A. K. Yadav, “Real time efficient scheduling algorithm for load balancing in fog computing environment,” *Int. J. Inf. Technol. Comput. Sci*, vol. 8, no. 4, pp. 1–10, 2016.
- [62] H. F. Atlam, R. J. Walters, and G. B. Wills, “Fog computing and the internet of things: a review,” *big data Cogn. Comput.*, vol. 2, no. 2, p. 10, 2018.
- [63] B. Feng, A. Tian, S. Yu, J. Li, H. Zhou, and H. Zhang, “Efficient Cache Consistency Management for Transient IoT Data in Content-Centric Networking,” *IEEE Internet Things J.*, 2022.
- [64] [T. Junfeng, B. Wenqing, and J. Haoyi, “PGCE: A distributed storage causal consistency model based on partial geo-replication and cloud-edge collaboration architecture,” *Comput. Networks*, vol. 212, p. 109065, 2022.
- [65] J. Tian, H. Jia, and W. Bai, “CCECGP: causal consistency model of edge-cloud collaborative based on grouping protocol,” *J. Supercomput.*, pp. 1–24, 2022.
- [66] C. Li, J. Bai, Y. Chen, and Y. Luo, “Resource and replica management strategy for optimizing financial cost and user

The Belief of Persian Text Mining Based on Deep Learning with Emotion-Word Separation

Hossein Alikarami¹, Amir Massoud Bidgoli^{2*}, Hamid Haj Seyed Javadi³

¹ Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran

² Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran

³ Department of Mathematics and Computer Science, Shahed University, Tehran, Iran

Received: 29 June 2023, Revised: 17 November 2023, Accepted: 12 December 2023

Paper type: Research

Abstract

Belief analysis or the classification of texts based on the feelings and opinions of users on websites and social media helps people, companies and organizations to make important decisions. Belief mining includes a system for analyzing people's opinions and feelings about an entity such as products, people, and organizations, according to the opinions, messages and tweets of users in social media. In this article, the belief analysis of Persian texts based on the messages, comments and tweets of users in social media and websites of four datasets using two deep learning methods, CNN, LSTM, taking into account the sense of the word, in two poles, positive and negative with intervals -2 and +2 are classified. In the proposed method, first the process of data pre-processing based on character to number conversion, removing the list of extra words and multi-word analysis is done, then for belief analysis and classification of Persian texts CNN, LSTM machine learning algorithm with word sense separation (WSD) is used to recognize the intensity of emotions according to the words. We call the proposed model CNN_WSD and LSTM_WSD. In the proposed method, the Persian Twitter dataset is used for evaluation and then it is compared with other machine learning and deep learning methods, DNN, CNN, LSTM, in the implementation of this method, python software is used. The accuracy rate of the proposed method for LSTM-WSD and CNN-WSD is 95.8 and 94.3%, respectively.

Keywords: Belief mining, Natural Language Processing (NLP), Deep Learning, Text Mining.

* Corresponding Author's email: DrBidgoli@gmail.com; am_bidgoli@iau-tnb.ac.ir

اعتقادیابی متون فارسی بر اساس یادگیری عمیق با تفکیک احساس-کلمه

حسین علی کرمی^۱، امیرمسعود بیدگلی^۲، حمید حاج سیدجوادی^۳
^۱ گروه مهندسی کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران
^۲ گروه مهندسی کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران
^۳ گروه ریاضی و علوم کامپیوتر، دانشگاه شاهد، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۴/۰۸ تاریخ بازبینی: ۱۴۰۲/۰۸/۲۶ تاریخ پذیرش: ۱۴۰۲/۰۹/۲۱
نوع مقاله: پژوهشی

چکیده

اعتقادکاوی یا طبقه‌بندی متون بر اساس احساس و عقیده کاربران در وبسایت‌ها و رسانه‌های اجتماعی به مردم، شرکت‌ها و سازمان‌ها کمک می‌کند تا بتوانند تصمیم‌گیری‌های مهم را انجام دهند. اعتقادکاوی شامل یک سیستم برای تحلیل عقاید و احساسات مردم درباره یک موجودیت مانند محصولات، افراد، سازمان‌ها با توجه به نظرات، پیام‌ها و توییت‌های کاربران در رسانه‌های اجتماعی می‌باشد. در این مقاله اعتقادکاوی متون فارسی بر اساس پیام‌ها، نظرات و توییت‌های کاربران در رسانه اجتماعی با استفاده از دو روش یادگیری عمیق LSTM, CNN با در نظر گرفتن شدت احساس کلمات و دیدگاه کاربران، در دو قطب مثبت و منفی با بازه ۲- و ۲+ طبقه‌بندی شده‌اند. در روش پیشنهادی ابتدا فرآیند پیش‌پردازش داده‌ها بر اساس تبدیل کاراکتر به عدد، حذف لیست واژه‌های اضافی و تحلیل چند واژه‌ای انجام می‌شود، در مرحله دوم جهت اعتقادکاوی و طبقه‌بندی متون فارسی از الگوریتم‌های یادگیری عمیق CNN و LSTM با تفکیک احساس بر روی کلمات (WSD) استفاده می‌شود تا شدت احساسات را با توجه به کلمات تشخیص دهد، در مرحله سوم دیدگاه کاربران با طبقه‌بندی متون در چهار دسته سیاسی، اجتماعی، اقتصادی و فرهنگی تشخیص داده می‌شود. ما مدل پیشنهادی را CNN_WSD و LSTM_WSD می‌نامیم. در این مقاله از مجموعه داده فارسی توییت برای ارزیابی و مقایسه روش پیشنهادی با سایر روش‌های یادگیری ماشین و یادگیری عمیق از جمله LSTM, CNN, DNN استفاده شده، پیاده‌سازی این روش با نرم‌افزار پایتون انجام شده است. میزان دقت روش پیشنهادی برای LSTM-WSD و CNN-WSD به ترتیب 95.8 و 94.3 درصد است.

کلید واژگان: اعتقادکاوی، پردازش زبان طبیعی (NLP)، یادگیری عمیق، متن کاوی.

۱- مقدمه

بر اساس دیدگاه آنها در دسته‌های مثبت و منفی بین بازه $+2$ و -2 تحلیل می‌کند.

در اعتقادکاوی و تحلیل احساسات بر اساس نظرات کاربران یادگیری عمیق چالش‌ها و مشکلاتی مانند هزینه بالای آموزش بر اساس زمان یا حافظه استفاده شده، عدم وجود واژگان غنی و کامل، ابعاد بالای فضای ویژگی و ابهام در تشخیص مثبت یا منفی برخی از جملات وجود دارد. از سوی دیگر، علم یادگیری عمیق توانسته با پیشرفت خود به بسیاری از مسائل حوزه پردازش زبان طبیعی پاسخ دهد و جایگزینی مناسب برای روش‌های سنتی باشد، یادگیری عمیق تاکنون از خود عملکرد بسیاری خوبی در بسیاری از شاخه‌های پردازش زبان طبیعی، خصوصاً تحلیل احساسات نشان داده است. مهمترین مزیت این روش، بی‌نیازی از استخراج دستی ویژگی‌ها است که به جای تخصص در حوزه زبان‌شناسی بر دسترسی به حجم بالای داده‌ها تکیه دارد.

برای مقابله با این مشکلات، در این مقاله از روش جدید یادگیری عمیق مبتنی بر شدت احساس کلمات بر اساس دیدگاه کاربران، به تحلیل احساسات و اعتقادکاوی نظرات پرداخته شده است. در این پژوهش روش پیشنهادی برای اعتقادکاوی نظرات فارسی کاربران در شبکه اجتماعی توییتر اهداف زیر را دنبال می‌کند:

- در پیش پردازش کلمات اضافی با توجه به لیست حذف می‌شود و داده‌ها به عدد و بصورت برداری تبدیل می‌شوند.
- در روش پیشنهادی بدلیل استفاده از روش یادگیری عمیق (CNN, LSTM) با تفکیک شدت احساس بر روی کلمه بر اساس دیدگاه کاربران در دقت رده‌بندی و تحلیل احساسات پیشرفت خوبی بدست آمده است.
- در روش پیشنهادی یک روش جدید با یادگیری عمیق با تفکیک و تحلیل احساسات در متون فارسی ارائه شده که متون در قسمت رمزنگاری به عدد تبدیل می‌شود تا متون به بردار تبدیل شود، این روش باعث افزایش دقت در اعتقادکاوی شود.
- ابتدا کلمات برگرفته از بدنه اصلی اسناد آموزشی با روش اسکپ‌گرام^۱ مشخص می‌شود و با مقایسه کلمات با واژگان، شدت احساسات آن مشخص می‌شود و در دسته‌های مثبت و منفی در بازه $+2$ و -2 مشخص می‌شود.
- تحلیل احساسات کاربران بر اساس شدت احساس کلمات می‌تواند دید دقیق و صحیح‌تری نسبت به نظرات کاربران در زمینه‌های مختلف را برای تصمیم‌گیری به ما بدهد.

تحلیل احساسات که همچنین نظرکاوی نامیده می‌شود، بخشی از مطالعات است که به تحلیل عقاید، احساسات و نگرش‌های مردم درباره موجودیت‌هایی همچون محصولات، سرویسها، سازمان‌ها، افراد، رخدادها و موضوعات خاص می‌پردازد [۱].

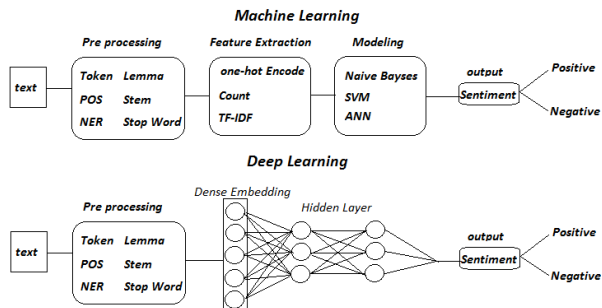
در فعالیت‌هایی که تاکنون بر روی متون فارسی انجام شده اغلب به طبقه‌بندی متون با توجه به زمینه و حوزه مورد مطالعه پرداخته شده است، در زمینه اعتقادکاوی و تحلیل احساسات بر روی نظرات و متون فارسی پیشرفت چشمگیری انجام نشده، در فعالیت‌هایی که تاکنون انجام شده اغلب بدون در نظر گرفتن معنا، شدت حساسیت کلمات و دیدگاه کاربران و کلمات اضافی فقط نظرات را در دو گروه مثبت و منفی طبقه‌بندی می‌کنند. وقتی در تحلیل احساسات کاربران نظرات در گروه‌های مثبت و منفی طبقه‌بندی می‌شود، اهمیت و ضرورت این موضوع احساس می‌شود که شدت احساسات نیز مورد ارزیابی قرار گیرد، در این مقاله با استفاده از روش پیشنهادی اعتقادکاوی با یادگیری عمیق، ابتدا نظرات کاربران در دو قطب مثبت و منفی دسته‌بندی شده و برای هر قطب شدت احساسات بین بازه -2 و $+2$ نسبت داده شده است، سپس دیدگاه کاربران از نظر سیاسی، اجتماعی، اقتصادی و فرهنگی تعیین می‌شود، اعتقادکاوی با تعیین شدت احساسات بر اساس دیدگاه کاربران باعث می‌شود که سازمان‌ها، شرکت‌ها و محققان به راحتی نظرات کاربران را در زمینه‌های مختلف تحلیل کنند و تصمیم‌گیری مناسب و تعیین اهداف را انجام دهند.

بیشتر پژوهش‌های انجام شده در حوزه پردازش اطلاعات متنی و تحلیل احساسات، بر روی داده کاوی و طبقه‌بندی اطلاعات، مانند تحلیل نظرات در دو قطب مثبت یا منفی تمرکز دارند. این در حالی است که در فرایند تصمیم‌گیری توسط مدیران سازمان‌ها و شرکت‌ها نیاز به اطلاع دقیق از دیدگاه و احساس کاربران می‌باشد، به عنوان مثال اگر درصد نظرات کاربران در دو قطب مثبت و منفی برابر باشد نمی‌توان تصمیم‌گیری انجام داد، اما در اعتقاد کاوی تعیین می‌کند که نیمی از کاربران با یک دیدگاه تعیین شده نظر مثبت و نیمی از آنها با دیدگاه دیگر نظر منفی دارند و یا حتی کاربرانی که از یک دیدگاه مشترک نظر داده‌اند چند درصد از آنها دیدگاه مثبت یا منفی دارند.

در اعتقادکاوی با تحلیل احساسات کاربران بر اساس شدت احساسات کلمات، زمینه‌ها و دیدگاه کاربران در دسته‌های سیاسی، اجتماعی، اقتصادی و فرهنگی مشخص می‌شود و نظرات کاربران را

^۱ Skip-gram

احساسات بین دو رویکرد یادگیری ماشین و یادگیری عمیق را نشان می‌دهد. شبکه‌های عصبی مصنوعی و یادگیری عمیق در حال حاضر بهترین راه حل‌ها را برای بسیاری از مشکلات در زمینه‌های اعتقادکاوی، تحلیل احساسات، تشخیص گفتار و پردازش زبان طبیعی ارائه می‌دهند.



شکل ۱. تفاوت بین دو رویکرد تحلیل احساسات با روش‌های یادگیری ماشین (بالا) و یادگیری عمیق (پایین).

یادگیری عمیق شامل شبکه عصبی عمیق (DNN)، شبکه عصبی حلقوی (CNN)، شبکه عصبی مکرر (RNN) و LSTM است [۲].

شبکه عصبی عمیق: یک شبکه عصبی با بیش از دو لایه است که برخی از آن‌ها لایه‌های پنهان هستند. شبکه‌های عصبی عمیق از مدل‌سازی ریاضی پیچیده برای پردازش داده‌ها به روش‌های مختلف استفاده می‌کنند.

شبکه عصبی حلقوی: یک نوع خاص از شبکه عصبی رو به جلو است که در اصل در مناطقی مانند بینایی ماشین، سیستم‌های توصیه‌کننده و پردازش زبان طبیعی به کار می‌رود. این یک معماری شبکه عصبی عمیق است، که به طور معمول از لایه‌های محکم و جمع با مخزن تشکیل شده است تا ورودی‌های یک لایه طبقه‌بندی کاملاً متصل را فراهم کند.

حافظه کوتاه مدت (LSTM): یک نوع خاص از شبکه عصبی مکرر (RNN) است که قادر به استفاده از حافظه طولانی به عنوان ورودی توابع فعال‌سازی در لایه پنهان است. سایر روش‌های یادگیری عمیق در [۳] بیان شده است.

۳- کارهای گذشته

در اوایل دهه ۱۹۶۰ مردم شروع به مطالعه طبقه‌بندی متن کردند. در آن زمان قوانین طبقه‌بندی با توجه به پدیده‌ها و قوانین زبان نوشته می‌شد. تا دهه ۱۹۹۰، مردم شروع به مطالعه تکنولوژی طبقه‌بندی خودکار مبتنی بر کامپیوتر کردند. در این روش‌ها ابتدا به پیش برچسب‌گذاری داده‌ها، یادگیری قوانین و سپس به طبقه‌بندی و آموزش نمونه‌های جدید از دسته‌های ناشناخته به

در سال‌های اخیر، محققین مطالعات متعددی در زمینه تحلیل احساسات با یادگیری عمیق برای زبان انگلیسی ارائه داده‌اند اما در زمینه اعتقادکاوی و تحلیل دیدگاه‌های کاربران تحقیقات زیادی انجام نشده است. همچنین تحلیل احساسات در متون فارسی نیز پیشرفت چشمگیری نداشته است. به همین دلیل در این کار به اعتقادکاوی با تحلیل شدت احساس کلمات بر اساس دیدگاه کاربران پرداخته‌ایم. در این کار جدیدترین مدل‌های یادگیری عمیق، مانند شبکه‌های عصبی عمیق (DNN)، شبکه‌های عصبی مکرر (RNN) و شبکه‌های عصبی کانولوشن (CNN) و مدل‌های یادگیری ماشین و روش پیشنهادی جدید CNN_WSD، LSTM_WSD بررسی و ارزیابی شده است، تا مشکلات مختلف در رابطه با تجزیه و تحلیل احساسات را حل کند. در روش پیشنهادی برای اعتقادکاوی و رده‌بندی نظرات از یادگیری عمیق با تفکیک شدت احساسات کلمات بر روی مجموعه داده‌های فارسی استفاده شده، این روش سبب استخراج ویژگی‌های مناسب و افزایش دقت و تشخیص شدت احساسات در بازه +۲ و -۲ می‌شود، بنابراین روش پیشنهادی برای بهبود تصمیم‌گیری و نظرکاوی در متون فارسی اهمیت بالایی دارد.

در این مقاله در بخش اول مقدمه‌ای در زمینه اعتقادکاوی و رده‌بندی متون فارسی، در بخش دوم ادبیات تحقیق و پیش زمینه‌ای در مورد روش‌های یادگیری عمیق، در بخش سوم به کارهای گذشته در زمینه اعتقادکاوی و تحلیل احساسات پرداخته شده، در بخش چهارم پیش پردازش متون فارسی و اعتقادکاوی نظرات با روش پیشنهادی CNN_WSD و LSTM_WSD ارائه شده، در بخش پنجم پردازش داده‌ها، در بخش ششم نتایج و آزمایشات روش پیشنهادی و مقایسه با سایر روش‌ها و سپس به نتیجه‌گیری پرداخته شده است.

۲- ادبیات تحقیق

در روش پیشنهادی اعتقادکاوی مبتنی بر الگوریتم یادگیری عمیق می‌باشد، قبل از بررسی و تحلیل مدل بهینه شبکه عصبی عمیق در روش پیشنهادی مقدماتی از یادگیری عمیق بیان می‌شود.

یادگیری عمیق یک رویکرد چند لایه را با لایه‌های پنهان شبکه عصبی تطبیق می‌دهد. در روش‌های یادگیری سنتی ماشین، ویژگی‌ها به صورت دستی یا با استفاده از روش‌های انتخاب ویژگی مشخص و استخراج می‌شوند. اما در مدل‌های یادگیری عمیق، ویژگی‌ها به طور خودکار استخراج می‌شوند و به دقت و عملکرد بهتری دست می‌یابند. شکل ۱ اختلاف‌های طبقه‌بندی و تحلیل

گرفته شده است [۸]. در مقایسه روش تعبیه کلمات با مدل کیف واژه^۱، روش تعبیه کلمات در مدل سازی متن کلمه و معنی کلمه نیز مؤثر است. بعد از تولید اسناد مناسب، می توان از مدل های مختلف شبکه عصبی و سایر روش های یادگیری ماشین، مانند ماشین های بردار پشتیبان، برای تولید کلمات تعبیه شده و طبقه بندی عقاید استفاده کرد [۹].

یکی از اولین پژوهش های صورت گرفته در زمینه نظر کاوی برای زبان فارسی مربوط به گردآوری مجموعه داده ای با نام PersianClues است. این پژوهش با استفاده از یک روش ابتکاری بدون ناظر به تحلیل احساسات می پردازد. در واقع تغییری که در این روش صورت گرفته اضافه کردن مجموعه کلمات حاوی بار معنایی به عنوان بردار ویژگی ها در مرحله یادگیری [۱۰]، پژوهش دیگری نیز تحت عنوان ایجاد یک سیستم نظر کاوی با استفاده از الگوریتم های ناظر انجام گرفته است. در گام نخست آن، یک لغتنامه احساس برای زبان فارسی به کمک شبکه واژگانی فارسی موجود، فارسی نت، گسترش داده شده است. این پژوهش با استفاده از سه الگوریتم یادگیری ماشین، شامل: ماشین بردار پشتیبانی، بیز ساده و رگرسیون منطقی به ارزیابی روش پیشنهادی خود پرداخته است [۱۱].

یکی از روش های اعتقاد کاوی و تحلیل احساسات بر روی متون فارسی، تحلیل احساسات کاربران بر اساس دیدگاه آنها است که با روش های یادگیری عمیق CNN، DNN و LSTM و تکنیک تعبیه کلمات^۲ و TF-IDF به آن پرداخته شده بر اساس نتایج این مقاله روش تعبیه کلمات^۳ با LSTM برای تحلیل احساسات مبتنی بر دیدگاه کاربران نتایج بهتری را داشته است [۳].

دو و هانگ^۴ مقاله ای تحت عنوان تحقیقات طبقه بندی متن با شبکه های عصبی مکرر مبتنی بر توجه را ارائه دادند که روش پیشنهاد شده در این مقاله مزایای هر دو روش را در نظر می گیرد. با استفاده از یک سازوکار توجه بر روی یادگیری وزن برای هر کلمه استفاده می شود. در این روش، کلمات کلیدی وزن بیشتری خواهند داشت و کلمات رایج وزن کمتری خواهند داشت. بنابراین، در نظر کاوی متون و اعتقادیابی نه تنها همه کلمات را در نظر می گیرد بلکه توجه بیشتری به کلمات کلیدی نیز می کند [۱۲].

یکی دیگر از پژوهش های صورت گرفته در زبان فارسی تحت عنوان

طور خودکار پرداخته شد. نتایج نشان می دهد که در زمینه حجم داده های بزرگ، دقت طبقه بندی آن بسیار بهتر از تعریف قوانین است. بنابراین تحقیقات کنونی بر طبقه بندی خودکار متون با الگوریتم های هوش مصنوعی متمرکز است [۴].

طبقه بندی متون یکی از وظایف اصلی یادگیری ماشین است. هدف آن طراحی الگوریتم های مناسب برای استخراج ویژگی ها و طبقه بندی متون به صورت خودکار است. در گذشته، اساساً از طبقه بندی کلمات کلیدی و طبقه بندی سنتز معنایی با شبکه عصبی استفاده می شد.

محققان در طول سال های ۲۰۰۰-۲۰۱۵ به ارزیابی و مقایسه روش های تجزیه و تحلیل احساسات و متن کاوی پرداخته اند [۵]. در زمینه تحلیل های احساسی مانند استخراج احساسات، نظر کاوی، استخراج نظرات و اعتقاد کاوی، روش هایی برای تجزیه و تحلیل احساسات و اندیشه کاوی در کلمه، جمله و سطح سند، برای نظرات مصرف کنندگان در عبارات بیان شده است. با توجه به فازی بودن شخصیت، تکنیک های یادگیری ماشین سنتی نمی توانند نظرات را به خوبی نشان دهند. برای رفع این مشکل روش تحلیل احساسات با فاز معنایی برای حل مسئله پیشنهاد شده است [۶].

مطالعه افکار عمومی می تواند اطلاعات ارزشمندی را در اختیار ما قرار دهد. تجزیه و تحلیل احساسات در شبکه های اجتماعی، مانند توییتر یا فیس بوک، به ابزاری قدرتمند برای تحلیل نظرات کاربران تبدیل شده و کاربردهای گسترده ای دارد. با این حال، چالش هایی در صحت تحلیل احساسات در پردازش زبان طبیعی (NLP) پیش آمده. در سال های اخیر نشان داده شده است که مدل های یادگیری عمیق یک راه حل امیدوارکننده برای چالش های تحلیل احساسات و اعتقاد کاوی است. در مطالعاتی مانند قطبیت احساسات که از یادگیری عمیق برای حل مشکلات تحلیل احساسات استفاده شده است، مدل هایی با استفاده از فرکانس سند، فرکانس معکوس (TF-IDF) و تعبیه کلمه بر روی یک سری مجموعه داده ها اعمال شده اند [۲].

با توجه به ویژگی های مورد استفاده برای بهره برداری از نظرات مصرف کننده، مدل کیف واژه ها بیانگر سند سنتی است که در آن فرکانس های کلمه برای هر کلمه (عبارت) در واژگان محاسبه می شود [۷]. با این حال، این رویکرد منجر به بازنمایی اسناد پراکنده در ابعاد بالا می شود، علاوه بر این، این روش معنی کلمه را نادیده می گیرد. برای غلبه بر این مشکلات، روش تعبیه کلمات به جای تک کلمات معرفی شده اند، که یک زمینه کوتاه در نظر

¹ Bag-of-Words

² Word Embedding

نیستند و منجر به افزایش خطاها در طبقه‌بندی می‌شوند، ویژگی‌های بهینه انتخاب می‌شوند و سپس این ویژگی‌ها به ماشین یادگیری داده می‌شود [۱۷].

در زمینه تحلیل احساسات، عنان روشی را برای اعتقادیابی و تفکر کاوی در سطح کلمه، جمله و سند برای نظرات مصرف کنندگان به زبان چینی ارائه کرد [۱۸]، او یک رویکرد مجموعه‌ای برای انتخاب ویژگی ارائه کرد، که چندین لیست ویژگی‌های فردی را که با روش‌های مختلف انتخاب ویژگی به دست آمده‌اند جمع می‌کند تا زیر مجموعه ویژگی‌های قوی‌تر و کارآمدتری به دست آید. تا تجزیه و تحلیل احساسات در مورد بررسی محصول بر اساس جاسازی کلمات وزن دار و شبکه‌های عصبی عمیق انجام دهد [۱۹].

یک معماری شبکه عصبی کانولوشن دو طرفه توسط وانگ پیشنهاد شد که از دو لایه دو طرفه LSTM و GRU استفاده می‌کند تا با اتصال دو لایه پنهان از جهت‌های مخالف به یک زمینه، هم زمینه‌های گذشته و هم آینده را استخراج کند [۲۰].

وانگ و همکاران یک روش افزایش بازنمایی خلاف واقع (CRA) را برای آموزش و طبقه‌بندی احساسات در اعتقادیابی نظرات کاربران ارائه کردند تا نتیجه عملکرد تعمیم دامنه هدف را بهبود بخشند [۲۱].

سیجی مای و همکاران یک چارچوب جدید با نام HyCon برای یادگیری متضاد ترکیبی نمایش سه وجهی برای احساسات چندوجهی ارائه کردند [۲۲].

ترکی و روی یک روش تشخیص سخنان منفی را با استفاده از تجسم ابری کلمه و یادگیری گروهی با بردار شمارنده ارائه کردند، آنها از یک چارچوب محاسباتی استفاده کردند که تکنیک‌های تقویت داده‌ها را برای بازنمایی و عملکرد بهتر استفاده کردند [۲۳].

۴- روش پیشنهادی یادگیری عمیق با تفکیک شدت احساس کلمات

در روش پیشنهادی مجموعه داده توپیتر فارسی برای تحلیل احساسات و اعتقادکاوی نظرات کاربران استفاده شده است. تحلیل احساسات براساس قطبیت انجام شده است، قطبیت جملات در این پیکره به صورت عددی بین $+2$ و -2 نمایش داده شده‌اند که عدد کوچکتر نشانگر قطبیت کمتر (بار منفی بیشتر) است.

هدف اصلی در متن‌کاوی، تحلیل احساسات کاربران بر اساس

بهربرداری از یادگیری عمیق در تحلیل احساسات است. در این پژوهش از مدل یادگیری عمیق شامل شبکه عصبی کانولوشن استفاده شده است. و در نهایت مدل یادگیری عمیق معرفی شده خود را با روش‌های کم عمق یادگیری ماشین همچون پرسپترون چندلایه مقایسه نموده‌اند [۱۳].

بصیری و کبیری مقاله‌ای تحت عنوان بهبود تجزیه و تحلیل احساسات و اعتقادیابی در زبان فارسی با استفاده از اصلاح واژگان را ارائه دادند، که هدف از تجزیه و تحلیل احساسات مبتنی بر اصلاح واژگان این است که مشکل استخراج افکار مردم از نظرات آنها در وب را با استفاده از واژگان کلمات از پیش تعریف شده رفع کنند. با این حال، اعتقادکاوی و نظرکاوی برای زبان فارسی در مقایسه با انگلیسی متفاوت است. استفاده از روش مبتنی بر واژگان در فارسی، یک رشته جدید است. منابع محدودی برای تحلیل احساسات و نظرکاوی در زبان فارسی وجود دارد که دقت روش‌های موجود مبتنی بر واژگان پایین‌تر از زبان‌های دیگر است [۱۴].

اورولاگین^۲ مقاله‌ای تحت عنوان یک رویکرد جدید برای تجزیه و تحلیل احساسات و تجسم فکری و طبقه‌بندی خلاصه اخبار ارائه داد که در این کار روش‌های موثر برای استخراج داده‌ها را پیشنهاد می‌کند. در این روش بصورت یک مرور کلی و خلاصه‌ای از متن و تحلیل احساسات می‌تواند احساسات بیان شده در متن را به صورت محاسبات به دست آورد. خلاصه‌سازی متن و تحلیل احساسات بر روی اخبار بی بی سی انجام شده است. روش خلاصه‌سازی با متن جایگزینی برای تجزیه و تحلیل احساسات مورد استفاده قرار می‌گیرد و طرح‌های تجسم سه بعدی برای نشان دادن اطلاعات احساسات ارائه شده است [۱۵].

کیویو و لی^۳ مقاله‌ای تحت عنوان تجزیه و تحلیل احساسات متن کوتاه در میکرو بلاگ براساس تجزیه وابستگی را ارائه دادند، این روش برای حل مشکلات ارتباط بین کلمات عاطفی و اصلاح‌کننده و احساسات متن کوتاه از طریق ساختار احساسات و قوانین محاسبه احساسات پیشنهاد شده، که به تحلیل احساسات متن کوتاه کمک می‌کند. احساس متن کوتاه با توجه به تأثیرات مختلف روابط میان جملات و سهم هر جمله به محاسبه اندیشه متن کوتاه پرداخته می‌شود [۱۶]. همچنین با در نظر گرفتن معنای کلمات در نظرکاوی، با توجه به اینکه برخی از ویژگی‌های انتخابی مناسب

¹ Sentiment Analysis

² Siddhaling Urologin

³ Lirong Qiu, Jie Li

c نمایانگر شعاع پنجره متن است (چند کلمه اطراف را در نظر می‌گیریم و $p(\omega + 1|\omega_t)$ احتمال کلمه خروجی با توجه به کلمات ورودی محاسبه شده با استفاده از الگوریتم سلسله مراتبی (softmax) است.

$$\rho(\omega_0|\omega I) = \prod_{j=1}^{L(\omega)-1} \sigma(\|n(\omega, j + 1) = ch(n(\omega, j))\|v'_{n(\omega, j)}v_{\omega I}) \quad (2)$$

جایی که ω_0 و ωI به ترتیب کلمات ورودی و خروجی هستند. v_{ω} و v'_{ω} به ترتیب نشانگرهای بردار کلمات ورودی و خروجی را نشان می‌دهند. $n(\omega, j)$ گره زام در درخت باینری است. $L(\omega)$ طول مسیر در درخت است. $ch(n)$ یک گره فرزند را نشان می‌دهد. $\sigma(x)$ یک تابع عملکرد سیگموئیدی را مشخص می‌کند.

در صورتی که اگر x صحیح باشد $[X] = 1$ ؛ در غیر این صورت $[X] = -1$.

برای به دست آوردن بازنمایی سند برای لایه بعدی در معماری CNN_WSD, LSTM_WSD، میانگین مقادیر بردارها از ماتریس وزن تعبیه شده محاسبه شد.

برای تکمیل بازنمایی احساس-کلمه با قطبیت و شدت احساسات، از چندین واژگان احساساتی از پیش تعریف شده استفاده کردیم. برای به دست آوردن یک ارزیابی احساساتی معتبر، پیشنهاد می‌شود به یک واژگان واحد اعتماد نکنید. علاوه بر این، ترکیبی از شاخص‌های احساسات مبتنی بر واژگان، مشکل حساسیت به عقاید غیرمستقیم را که معمولاً در مدل‌های یادگیری ماشین وجود دارد، غلبه می‌کند. برای محاسبه قطبیت احساسات، از دو واژگان دست‌ساز کلمات مثبت و منفی استفاده کردیم [۱۴].

یکی از کمبودهای این واژگان این است که وزن مساوی بدون توجه به شدت احساس آنها به همه کلمات اختصاص می‌یابد. برای پرداختن به این مسئله، ما شاخص‌های شدت احساسات به دست آمده از واژگان زیر را با نقاط قوت احساساتی از پیش آموزش دیده درج کردیم: بنابراین، نمرات کلی مثبت و منفی را می‌توان برای هر واژگان محاسبه کرد. علاوه بر این، ترکیب چندین واژگان، پوشش واژگان بالاتری را تضمین می‌کند [۲۵].

برای به دست آوردن جایگاه n-gram، وزن هر n-gram به شرح زیر محاسبه می‌شود:

$$\omega_{ij} = (1 + \log(tf_{ij})) \times \log(N/df_i) \quad (3)$$

جایی که ω_{ij} وزن i -ام n-gram را در سند زام (بررسی) بیان می‌کند. $j = 1, 2, \dots, N$ ؛ df_i و tf_{ij} به ترتیب بیانگر فرکانس

قطبیت، تشخیص شدت احساسات با تفکیک احساس-کلمه است. در این مقاله تحلیل احساسات با روش پیشنهادی یادگیری عمیق مبتنی بر احساس-کلمه انجام می‌شود، که برای آنالیز دقیق و صحیح‌تر نظرات در دو مرحله زیر انجام می‌شود:

در مرحله اول احساسات کاربران در دسته‌های مثبت و منفی در باز ۲+ و ۲- با در نظر گرفتن کلمه-احساس مشخص می‌شود. به صورتی که کلمات با استفاده از مدل اسکپ گرام آموزش داده می‌شود.

در مرحله دوم واژگان بدست آمده از بدنه اصلی اسناد با چند واژگان مقایسه می‌شود تا قطبیت و شدت احساسات مبتنی بر واژگان را اضافه کند.

برای آنالیز داده‌ها پیش‌پردازش بر روی اسناد آموزشی انجام شده و داده‌های گسیخته حذف شده، و بردارهای مبتنی بر $tf-idf$ از متن ورودی به دست آمده و بردارهای حاصله، نرمال‌سازی می‌شوند. در مرحله پیش‌پردازش کاهش ابعاد بر روی متون انجام می‌گردد [۲۴]. در نهایت بعد از پیش‌پردازش که در زیر جزئیات آن شرح داده شده، داده‌های پیش‌پردازش شده به عنوان داده‌های آموزشی برای اعتقادکاوی و رده‌بندی نظرات روش پیشنهادی اعمال می‌گردد.

در تحلیل احساسات و دسته‌بندی خودکار نظرات از یادگیری عمیق مبتنی بر احساس-کلمه CNN_WSD, LSTM_WSD بر روی اسناد انجام می‌شود. CNN (Convolutional Neural Networks), LSTM (Long-Short-Term Memory) پنهن متراکم برای پردازش در ویژگی‌های ورودی استفاده می‌شود. این دو نمایندگی شامل احساس-کلمه و n-gram است.

بازنمایی کلمه-احساس در دو مرحله تولید می‌شود: در مرحله اول، تعبیه کلمات با استفاده از مدل اسکپ گرام آموزش داده می‌شود. مرحله دوم، واژگان به دست آمده از بدنه اصلی اسناد ورودی با چند واژگان مقایسه شده است تا قطبیت و شدت احساسات مبتنی بر واژگان را اضافه کند [۳].

برای محاسبه ماتریس وزن تعبیه شده، عملکرد تعبیه شده برای هر کلمه ω_t در واژگان اعمال می‌شود. تابع تعبیه برای دنباله $\omega = \{\omega_1, \omega_2, \dots, \omega_t, \dots, \omega_T\}$ کلمات آموزش داده شده تا عملکرد تابع هدف زیر به حداکثر برسد:

$$E = \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c} \text{Logp}(\omega_t + j|\omega_t) \quad (1)$$

¹ Outlier

² Term Frequency-Inverse Document Frequency

الگوریتم ۱. الگوریتم روش پیشنهادی تحلیل احساسات

Parameters: $n(w, j)$ is the j th node in the binary tree. $L(w)$ is the path length in the tree. $ch(n)$ is a child node. $\sigma(x)$ specifies a sigmoid function.

Input: Set of words in a sentiment lexicon $w = \{w_1, w_2, \dots, w_n\}$.

Output: Well-trained sentiment- word embeddings Et
Initialization

$w \text{ embed}(w) // \text{Equation (1)}$

$i = 0, n = \text{total number of documents}$

while $i < n$ do

if $P(W_o|W_i)$

v_w is the vector marker of the input words and v'_w represent the vector marker of the output words.

$x = \text{Softmax}(Wt + j) // \text{Equation (2)}$

If x is integer, $[[x]] = 1$; Otherwise $[[x]] = -1$.

If x is correct, $[[x]] = 1$; Otherwise $[[x]] = -1$.

end if where $(W_{ij}$ weight of each n -gram in the document) // Equation (3)

Repeating number or using the frequency product of each word

$tf_idf // \text{Equation (4)}$

$i = i + 1$

Obtain well-trained sentiment- word embeddings(WSD): using synonym Word and the words of the context around it.

end while

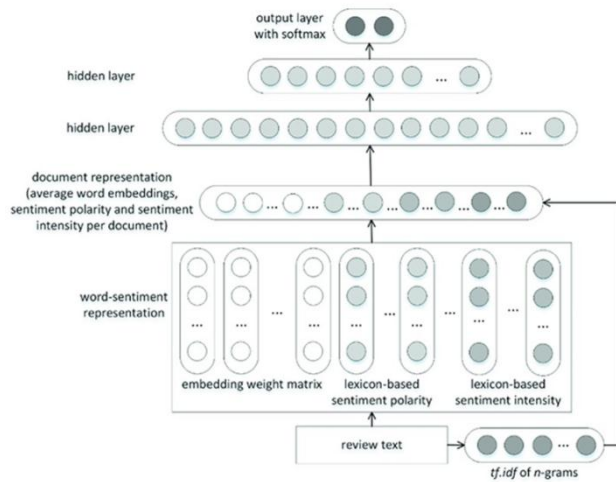
return.

مراحل طراحی روش پیشنهادی در ادامه بیان شده است.

۴-۱- پیش‌پردازش و خلاصه‌سازی اولیه

در حقیقت، پیش‌پردازش وظیفه نگاشت متن داده شده به یک نمای منطقی را بر عهده دارد. به عبارت دیگر استخراج ویژگی و وزندهی و کاهش ابعاد در این قسمت انجام می‌گیرد. بسته به کاربرد استخراج ویژگی می‌تواند بسیار ساده و یا بسیار مفصل باشد. تحلیل واژگانی شامل عملیات مربوط به یکسان سازی متن، قواعد مربوط به نشانه‌گذاری‌ها و مرزبندی بین کلمات می‌باشد. بعد از این مرحله عموماً دسته‌ای از کلمات بی‌ارزش که متناوباً تکرار می‌شوند و بار معنایی خاصی ندارند: مانند حرف ربط («و»، «که»، «تا»، «وقتیکه»، «اگر»، «اما»، «اینکه»، حرف اضافه («به»، «با»، «از»، «در»، فعل ربطی («است»، «بود»، «شد») و حرف تعریف («یک» در «یک دانشجوی نمونه کسی است که...») از متن داده حذف می‌شوند. سپس با استفاده از الگوریتم‌های ریشه‌یابی، به منظور بهینه‌سازی ویژگی‌های استخراج شده، کلمات ریشه‌یابی می‌شوند. در نهایت با استفاده از گروه‌های اسمی کلمات دسته‌بندی می‌گردند.

اصطلاحات و اسناد^۱ هستند. بنابراین، طول بررسی در نظر گرفته می‌شود، و وزن نسبتاً بالاتری به n -gram نادر اختصاص می‌یابد. برای پردازش بیشتر، n -gramها با توجه به وزن آنها رتبه‌بندی می‌شوند، و n -gramهای برتر برای ورود به لایه بازنویسی اسناد در معماری CNN_WSD, LSTM_WSD انتخاب می‌شوند.



شکل ۲. معماری پیشنهادی برای استخراج نظرات و اعتقادکاوی

دو لایه پنهان بعدی برای پردازش رابطه پیچیده بین نمایش اسناد و کلاسهای مثبت / منفی از خروجی استفاده می‌شود. برای جلوگیری از اضافه کردن اتصالات و اثربخشی آموزش، به ترتیب از یکپارچه‌سازی ترک خوردگی ۰٫۲ و ۰٫۵ برای ورودی و دو لایه پنهان استفاده کردیم. الگوریتم نزول شیب مینی دسته‌ای با $b = 100$ مینی دسته^۲، میزان یادگیری ۰٫۱ و ۱۰۰۰ تکرار، رفتار همگرایی خوب و پایدار را برای ما فراهم کرده است. اعداد مختلف nh_1 و nh_2 در دو لایه پنهان برای به دست آوردن معماری مطلوب مورد آزمایش قرار گرفت. همانطور که در زیر ارائه می‌شود، بهترین نتایج برای $nh_1 = 25$ و $nh_2 = 24$ نوروں بدست آمد. توجه داشته باشید که ما همچنین با یک لایه پنهان آزمایش کرده‌ایم اما بدون پیشرفت. عملکرد هدف با از دست دادن آنتروپی متقاطع نشان داده شد. پیچیدگی کلی مدل ارائه شده می‌تواند به صورت:

$$O(b \times I \times (m \times nh_1 + nh_1 \times nh_2 + nh_2 \times n_0))$$

بیان شود، که I تعداد تکرارها را نشان می‌دهد. m تعداد ویژگی‌های موجود در لایه ارائه اسناد را نشان می‌دهد. nh_1 و nh_2 به ترتیب تعداد نوروں‌ها در لایه‌های پنهان اول و دوم و لایه خروجی را نشان می‌دهند.

¹ Term and Document Frequency

² Mini-batches

الگوریتم ۲. نحوه تخصیص شدت احساس کلمات

کمتر به وجود می‌آیند که بعد از انجام اعمالی نظیر رده‌بندی، جواب‌هایی مشابه داده‌های اولیه تولید می‌کنند، در روش پیشنهادی برای حل این مسئله از پارامترهای معکوس فرکانس (tf-idf) و مطابقت کلمات در کلاس (uni) و یکنواختی (ICF) مربوط به هر کلمه استفاده می‌کند، سپس کلمه ای را انتخاب می‌کند که حد آستانه کوچکتر از ICF و uni مشخص و بزرگتر از حد آستانه تعیین شده داشته باشد.

در حقیقت این پارامترها سعی می‌کنند با امتیاز دادن به تکرار کلمه در کلاس‌های کمتر و تعداد رخداد بیشتر آن در یک کلاس، کلماتی را به عنوان ویژگی نهایی انتخاب کنند که نماینده یک کلاس خاص باشند.

۴-۱-۲- بازنمایی متون با روش تعبیه واژه‌های عصبی

برای وزندهی به ویژگی‌ها می‌توان از رویکردهای متفاوتی بهره برد. در ساده‌ترین حالت این وزندهی می‌تواند به صورت باینری انجام شود. انتخاب دیگر وزندهی به هر کلمه با توجه به تعداد تکرار هر کلمه و یا استفاده از حاصلضرب فرکانس هر کلمه در معکوس فرکانس سند که معمولاً به صورت زیر تعریف می‌شود:

$$TF - IDF(ti, dj) = tf(ti, dj) \times \log(N/(N(ti))) \quad (5)$$

ti تعداد اسنادی از مجموعه $N(ti)$ و N نماینده تعداد کل اسناد می‌باشد.

در نظر کاوی داده‌ها با یادگیری عمیق ابتدا داده‌های متنی باید با روش‌های رمزنگاری به عدد تبدیل شود، زیرا هر مدل پایه ریاضی دارد. با وجود روش‌های برداری فوق که اغلب در الگوریتم‌های یادگیری ماشین سنتی استفاده می‌شود، ما از تعبیه واژه‌های عصبی در مدل‌های یادگیری عمیق بهره مند می‌شویم که هر کلمه به یک بردار کم ابعاد به نام ویژگی کلمه تبدیل می‌شود [۲۶]. تعریف کلمات در این نمایش باعث می‌شود که کلمات مشابه مستقیماً پیدا شوند. بنابراین می‌توانیم از این مزیت برای کاوش نظر نیز استفاده کنیم. به طور خلاصه، هر جمله رمزگذاری می‌شود، کلمات بعدی برداری می‌شوند. در نتیجه، هرچه بردارهای مشابه بیشتری در میان جملات قرار بگیرند، شباهت بیشتری خواهند داشت. شایان ذکر است که تعبیه کلمه عصبی نه تنها یک ردیاب مترادف است بلکه یک روش برای یافتن کلمات از یک خانواده (به عنوان مثال گربه، سگ) است. این بردارها به دو صورت لایه جاسازی آنلاین و جاسازی کلمات از قبل آموزش داده شده انجام می‌شود، روش لایه جاسازی آنلاین به مجموعه داده موجود متکی است و در فرایند یادگیری عصبی عملی خواهد شد. در

```
1: A set of words in Vocabulary = (W1,W2,W3,...,WV)
2: A dictionary of input words and their corresponding sentiment polarity
3: // Get sentiment polarity of words based on SentiWordNet and
   Sentipers lexicons
4: Sentiment_class = { } // Dictionary of words and their corresponding
   polarity class
5: for each W in Vocabulary:
6: Score = 0
7: Word = Lemmatization(W) // Lemmatize the word
8: Synsets = GetSynsets_SentiWordNet(Word) // Obtain the synsets of a
   word from SentiWordNet
9: if length(Synsets) > 0 do // That means the word exists in
   SentiWordNet lexicon
10: Score = Average( Synsets.positive_scores) - Average(
   Synsets.negative_scores)
11: else
12: Score = getPolarity_SenticNet(Word)
13: if Not Score do // That means the word does not exist in the SenticNet
   lexicon
14: Score = getPolarity_VADER(Word)
15: end if
16: end if
17: // Assign sentiment class based on the score obtained before
18: append (W: "Strong Negative") to Sentiment_class if Score <= -1
19: append (W: "Negative") to Sentiment_class if -1 < Score < 0
20: append (W: "Neutral") to Sentiment_class if Score == 0
21: append (W: "Positive") to Sentiment_class if 0 < Score < 1
22: append (W: "Strong Positive") to Sentiment_class if Score >= 1
23: end for return Sentiment_class
```

همانطور که در الگوریتم ۲ خلاصه شده است، کلمه مورد نظر با ارزش امتیاز احساسات آن از مجموعه داده جایگزین شده است. از طرف دیگر، از آنجایی که کلمه ممکن است در قطبیت‌های مختلف در SentiWordNet و Sentpres ظاهر شود، از تفاوت میانگین امتیازات مثبت و منفی استفاده کردیم. این با فرمول زیر نشان داده می‌شود:

$$Score = \sum_i = 1kSynsetp(i)k - \sum_i = 1kSynsetn(i)k \quad (4)$$

در فرمول ۴ مقدار k به تعداد ظاهر کلمه اشاره دارد، $Synsetp$ نشان دهنده امتیاز مثبت و $Synsetn$ نشان دهنده امتیاز منفی است.

۴-۱-۱- انتخاب ویژگی اولیه

انتخاب ویژگی، تکنیکی است که برای مواجهه با داده‌های با ابعاد بالا^۱ استفاده می‌شود. مسئله اصلی این است که داده‌هایی که دارای ابعاد بالا هستند زمان بیشتری را برای پردازش صرف می‌کنند. یکی از راه‌های کم کردن زمان محاسبات، انتخاب ویژگی‌هایی از فضای مسئله است که در تعیین جوابها موثر هستند و از باقی ویژگی‌ها صرف نظر می‌شود. بدین صورت داده‌هایی با ابعاد

¹ High dimensional

Ranked	Emotion Class
-2	Furious
-1	Angry
0	Neutral
+1	Happy
+2	Delighted

قطبیت هر جمله از یک مجموعه E بین بازه +۲ و -۲ انتخاب شده، $E = \{-2, -1, 0, +1, +2\}$ که در جدول ۱ نمایش داده شده است.

۴-۳- پردازش داده

عملکرد یادگیری ماشین و یادگیری عمیق اغلب به اندازه و کیفیت داده‌های آموزشی بستگی دارد که جمع آوری آنها اغلب خسته‌کننده است [۲۷]، بنابراین، از چندین مجموعه داده متعادل استفاده شده است.

در این مقاله از مجموعه داده توئیتر فارسی استفاده شده، نظرات فارسی در مجموعه داده شبکه اجتماعی توئیتر شامل ۱۰۰۰۰۰۰ سند است، برای هر دسته نظرکاوی با مثبت و منفی بودن نظر کاربران نمایش داده می‌شود.

۵- نتایج و آزمایشات

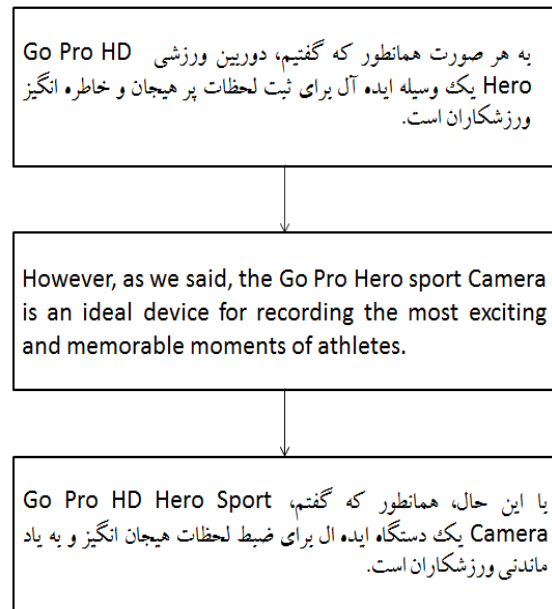
در اعتقادیابی هر کلمه به عنوان یک ویژگی در داده‌های آموزشی برای رده‌بندی در نظر گرفته می‌شود و کلمات در متون فارسی بر اساس نرخ تاثیر در نظرکاوی در رده‌بندی اندازه‌گیری می‌شود و بر اساس میزان دقت Acc و زمان رده‌بندی اندازه‌گیری می‌شود.

این آزمایشات بر روی مجموعه داده توئیتر با ۱۰۰۰۰۰۰ سند انجام شده است. ما از مدل‌های یادگیری عمیق مبتنی بر شدت احساسات کلمات با TF-IDF و تعبیه کلمات بر روی مجموعه داده‌های توئیتر استفاده کرده‌ایم. برای یادگیری تعبیه کلمات (word embeddings)، از مدل Skip-Gram برای آموزش داده‌ها استفاده شده است [۲۸-۳۲].

۵-۱- مجموعه داده توئیتر

نظرات فارسی در مجموعه داده شبکه اجتماعی توئیتر شامل ۱۰۰۰۰۰۰ سند است، برای هر دسته نظرکاوی با مثبت و منفی بودن نظر کاربران نمایش داده می‌شود.

حقیقت، بردارهای خروجی از ورودی با استفاده از هیچگونه عمل ریاضی محاسبه نمی‌شوند. بنابراین، هر کلمه در جملات همانطور که ظاهر می‌شود با یک عدد صحیح رمزگذاری می‌شود. در این حالت، Vs تعداد کلمات مجموعه واژگان را نشان دهد و Ev بعد تعبیه بردارها را نشان می‌دهد. شکل ۳ نحوه آنلاین بودن را نشان می‌دهد.



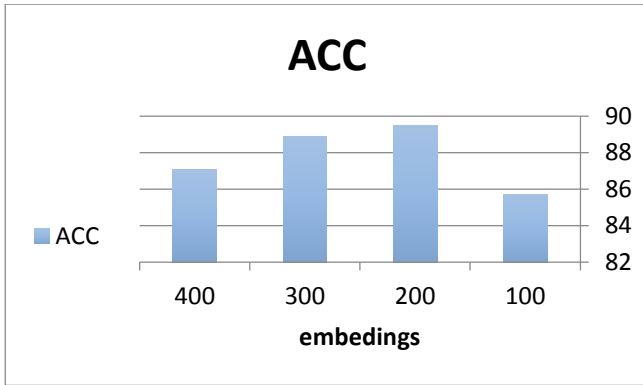
شکل ۳. لایه جاسازی آنلاین

۴-۲- اعتقادکاوی

در اعتقادکاوی، هم بازخوردهای مثبت و هم منفی مجموعه داده استفاده شده که نظرات فارسی کاربران در توئیتر می‌باشد را با استفاده از روش پیشنهادی CNN_WSD, LSTM_WSD رده‌بندی و درصد بازخورد مثبت و منفی نظرات کاربران در مورد هر یک از موضوعات استخراج می‌شود تا مشخص شود که آیا از نظر کاربران درصد بازخورد مثبت هریک از موضوعات بیشتر است یا درصد بازخورد منفی، درصد مثبت و منفی هر یک از موضوعات با توجه به مجموع فراوانی نسبی تکرار هر یک از لغات مثبت (زیبا، عالی، دوست داشتنی و غیره) و منفی (بد، زشت، نامناسب، افتضاح و غیره) مربوط به هر موضوع محاسبه می‌شود. در مجموعه داده آموزشی هر جمله با ویژگی‌هایی مانند قطب، کلمات کلیدی و اهداف شرح داده می‌شود.

جدول ۱. مجموعه‌ای از احساسات رتبه‌بندی شده بر اساس قطبیت

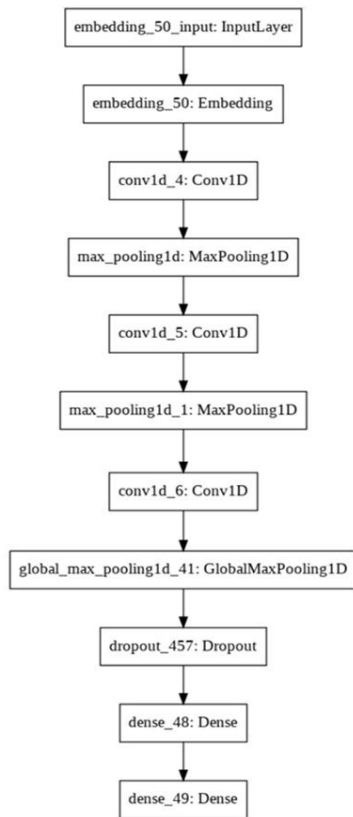
جدول ۲. رتبه‌بندی جملات مجموعه داده توئیتر



شکل ۵. تأثیر تعداد تعبیه کلمات بر عملکرد مدل CNN_WSD, LSTM_WSD با دو لایه پنهان از $nh1 = 25$ و $nh2 = 24$ نرون.

۵-۳- مدل‌ها

در این قسمت از ساختار LSTM مبتنی بر تعبیه کلمات برای لایه‌های شبکه عصبی استفاده شده است. نخستین ساختار حافظه طولانی کوتاه مدت دوطرفه (LSTM) است که بر پایه شبکه‌های عصبی بازگشتی طراحی شده و دوطرفه بودن آن امکان دریافت اطلاعات توسط گذشته و آینده را به لایه خروجی آن اضافه میکند. در شکل ۶ لایه‌های تعبیه شده برای این ساختار قابل مشاهده است.



شکل ۶. ساختار لایه‌های در مدل‌های پیشنهادی LSTM-WSD

Ranked	-2	-1	0	+1	+2
Sentences	5394	93998	425084	294538	180986

مجموعه داده توییت‌ر شامل فیلدهای زیر می‌باشد:

- “target” is the polarity of the tweet;
- “id” is the unique ID of each tweet;
- “date” is the date of the tweet;
- “query_string” indicates whether the tweet has been collected with any particular query keyword (for this column, 100% of the entries labeled are with the value “NO_QUERY”);
- “user” is the Twitter handle name of the user who tweeted;
- “text” is the verbatim text of the tweet.

R	J	I	H	G	F	E	D	C	B	A	1
tweet	place	name	username	user_id	timezone	time	date	created_at	conversation_id		
۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱

شکل ۴. نمونه مجموعه داده توییت‌ر

ما از فیلدهای target, text برای انجام آزمایشات استفاده می‌کنیم. تعداد جمله‌های تمام دسته‌ها ۱۰۰۰۰۰۰ می‌باشد که تعداد داده‌های هر دسته نیز در جدول ۲ مشخص شده است [۳].

۵-۲- پارامترها

از آنجا که هر متن به صورت مجموعه‌ای از کلمات تعبیه شده در آمده، در لایه اول مدل‌های شبکه عصبی خود، تعداد نرون‌ها به اندازه بیشترین طول جملات بر حسب کلمه است. در این مجموعه داده، طولانی‌ترین متن موجود شامل ۲۵۷ کلمه بوده و بنابراین در لایه اول ۲۵۷ نرون خواهیم داشت.

همانطور که در شکل ۵ نشان داده شده است، ما با تنظیمات مختلف مدل را آزمایش کرده‌ایم. در تعبیه کلمات به کمک Keras، در مدل‌های از قبل تعیین شده بهترین عملکرد با تعبیه کلمه $embed_size = 300$ بدست آمد. از آنجا که یادگیری تعبیه در خود شبکه عصبی صورت می‌گیرد این ابعاد $num_words = 2000$ در نظر گرفته شده است. مدل اسکپ گرام در محیط پایتون پایتون با استفاده از سیستم یادگیری عمیق کلب گوگل^۱ آموزش داده شد.

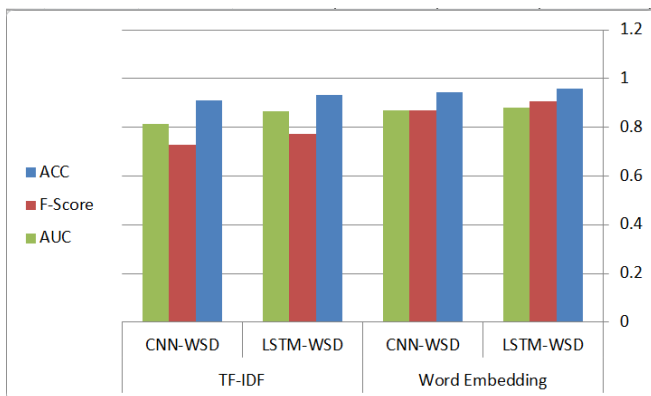
^۱ Colab google

نتایج شکل ۸ نشان می‌دهد، بهترین رفتار با ترکیب LSTM و تعبیه کلمات (Word embedding) نشان داده می‌شود و بالاترین مقادیر ACC، F_Score و AUC با تعبیه کلمات LSTM + داده شد. ما می‌توانیم تأیید کنیم که تعبیه کلمات یک تکنیک مناسب‌تر از TF-IDF برای انجام تجزیه و تحلیل احساسات است، پس از تجزیه و تحلیل نتایج مربوط به کیفیت پیش‌بینی‌ها، لازم است اطلاعاتی در مورد هزینه محاسباتی مرتبط با القاء مدل‌ها به دست آوریم.

جدول ۴ زمان پردازش مورد نیاز برای القاء مدل‌ها از مجموعه داده تویتر را نشان می‌دهد و شامل زمان CPU مورد نیاز در آزمایش‌ها است. این جدول نشان می‌دهد که استفاده از TF-IDF، به زمان محاسبه طولانی‌تری نسبت به استفاده از تعبیه کلمات (Word Embedding) نیاز دارد. این یکی دیگر از دلایلی است که تکنیک تعبیه کلمات را توصیه می‌کنیم. با این حال، LSTM با TF-IDF و هم با تعبیه کلمات وقت‌گیرترین الگوریتم است.

در نهایت، خلاصه کلی از نتایج بدست آمده در آزمایشات را در زیر توضیح می‌دهیم:

دو مدل یادگیری عمیق (CNN و LSTM) با تفکیک شدت احساس کلمات برای انجام آزمایش‌های تجزیه و تحلیل احساسات استفاده شد. مشخص شد که مدل LSTM بهترین نتیجه بین زمان پردازش و دقت نتایج را ارائه می‌دهد.



شکل ۸. نمودار ارزیابی روش‌های یادگیری عمیق در مجموعه داده تویتر

جدول ۴. زمان آزمایشات مدل‌های مختلف روش پیشنهادی با پردازنده گرافیکی GPU

Experiments	TF-IDF		Word Embedding	
	CNN-WSD	LSTM-WSD	CNN-WSD	LSTM-WSD
Tweeter Dataset	2min 48s	23min 52s	8min 01s	10min 39s

در این قسمت ساختار CNN مبتنی بر تعبیه کلمات (WSD برای لایه‌های شبکه عصبی استفاده شده است، در روش پیشنهادی شبکه عصبی کانولوشن (CNN) است که یکی از موفق‌ترین ساختارهای شبکه‌های عصبی است این مدل برای داده‌های متنی خصوصاً در مسائل طبقه‌بندی متن نیز به خوبی عمل میکند تنظیمات CNN در روش استفاده شده در شکل ۷، قابل مشاهده است.

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 40, 300)	4500300
conv1d_1 (Conv1D)	(None, 40, 64)	57664
conv1d_2 (Conv1D)	(None, 40, 32)	6176
max_pooling1d_1 (MaxPooling1D)	(None, 13, 32)	0
conv1d_3 (Conv1D)	(None, 13, 16)	1552
conv1d_4 (Conv1D)	(None, 13, 8)	264
global_average_pooling1d_1 (GlobalAveragePooling1D)	(None, 8)	0
dense_1 (Dense)	(None, 1)	9

Total params: 4,565,965
Trainable params: 65,665
Non-trainable params: 4,500,300

شکل ۷. ساختار لایه‌های در مدل‌های پیشنهادی CNN

۵-۴- نتایج

در آزمایشات ما، سه روش ارزیابی بر روی مجموعه داده تویتر در نظر گرفته می‌شود: دقت (Acc)، مساحت زیر منحنی (AUC) و F-Score. برای ارزیابی عملکرد مدل پیشنهادی، از روش k-fold cross validation استفاده شده که k برابر با ۱۰ می‌باشد.

نتایج تعبیه کلمات برای روش LSTM_WSD: ACC برابر با ۰.۹۵۸، درصد و F-Score برابر با ۰.۹۰۷ و AUC برابر با ۰.۸۶۴ است و روش CNN_WSD با ACC برابر با ۰.۹۴۳، درصد و F-Score برابر با ۰.۸۷۱ و AUC برابر با ۰.۸۸۹ است.

نتایج جدول ۳ نشان می‌دهد استفاده از تعبیه کلمات برای تحلیل احساسات با یادگیری عمیق نتایج بهتری نسبت به TF-IDF برای ارزیابی دقت (Acc)، مساحت زیر منحنی (AUC) و F-Score دارد.

جدول ۳. نتایج آزمایش‌های ACC، F-Score و AUC برای تویتر

مجموعه داده‌ها

Experiments	TF-IDF		Word Embedding	
	CNN-WSD	LSTM-WSD	CNN-WSD	LSTM-WSD
ACC	0.910036	0.934178	0.943248	0.958863
F-Score	0.727000	0.772000	0.871000	0.907000
AUC	0.814408	0.864623	0.868279	0.881397

LSTM نسبت به DNN و CNN در مورد اخیر چندان قابل توجه نیست، استفاده از این دو روش را می‌توان از نظر کاهش زمان و هزینه محاسباتی مناسب‌تر دانست، البته انواع مختلف مجموعه داده‌ها بر نتایج تجزیه و تحلیل احساسات متفاوت تأثیر می‌گذارد.

جدول ۵. مقایسه و ارزیابی نتایج روش پیشنهادی با سایر مدل‌های یادگیری عمیق

Model	AUC	Acc [%]	F-score	Testing time [s]
LSTM_WSD (this study)	88.13 ± 0.91	0.958 ± 0.011	0.907 ± 0.009	10.398 ± 0.226
CNN_WSD (this study)	86.82 ± 0.91	0.943 ± 0.005	0.871 ± 0.006	8.012 ± 0.131
LSTM [3]	84.05 ± 0.28	0.917 ± 0.003	0.841 ± 0.002	2.042 ± 0.128
CNN [33]	84.29 ± 0.17	0.921 ± 0.001	0.844 ± 0.003	8.139 ± 0.286
DNN [34]	84.14 ± 0.71	0.842 ± 0.002	0.828 ± 0.005	8.14 ± 0.31
Tree-LSTM & Discourse-LSTM [35]	86.06 ± 0.28	0.867 ± 0.006	0.851 ± 0.003	4.15 ± 0.11
GRU, and hybrid approaches [9]	85.29 ± 0.16	0.851 ± 0.002	0.854 ± 0.006	11.2 ± 0.31
RNN [36]	84.74 ± 0.91	0.850 ± 0.003	0.838 ± 0.008	9.16 ± 0.21
Coattention-LSTM + Location [37]	87.52 ± 0.19	0.891 ± 0.002	0.871 ± 0.003	5.032 ± 0.18
Bi-LSTM [38]	84.85 ± 0.06	0.871 ± 0.002	0.841 ± 0.003	3.042 ± 0.45

جدول ۶. نتایج آزمایش‌های ACC، F-Score و AUC برای توییت‌ها. مجموعه داده‌ها.

Experiments	TF-IDF		Word Embedding	
	DNN-WSD	LSTM-WSD	DNN-WSD	LSTM-WSD
ACC	0.757757	0.934178	0.790962	0.958863
F-Score	0.763832	0.772000	0.788766	0.907000
AUC	0.764996	0.864623	0.788166	0.881397

اگرچه مدل LSTM در هنگام استفاده از تعبیه کلمات از بالاترین درجه دقت برخوردار بود، اما زمان پردازش آن بیشتر از مدل CNN بود. مدل LSTM هنگام استفاده از تکنیک TF-IDF موثر نیست و زمان پردازش بسیار بیشتر آن منجر به نتایجی می‌شود که بهتر نیستند. برای ارزیابی جامع اثربخشی مدل پیشنهادی، عملکرد آن را در مقابل مدل‌های موجود زیر مقایسه کردیم. در جدول ۵ روش‌های مختلف یادگیری عمیق از جمله LSTM، CNN، DNN، Tree-LSTM، RNN، GRU، Bi-LSTM و Coattention-LSTM که از سال‌ها ۲۰۱۸ تا ۲۰۲۳ توسط محققین برای تحلیل احساسات بر روی مجموعه داده توییت‌ها استفاده شده است، با استفاده از نرم‌افزار پایتون برای مجموعه داده فارسی توییت‌ها پیاده‌سازی و شبیه‌سازی شده‌اند و در حالت برابر ارزیابی شده است، با توجه به اینکه روش‌های یادگیری عمیق عملکرد بهتری نسبت به روش‌های یادگیری ماشین دارد ما فقط روش‌های پرکاربرد یادگیری عمیق را مورد ارزیابی و مقایسه قرار داده‌ایم.

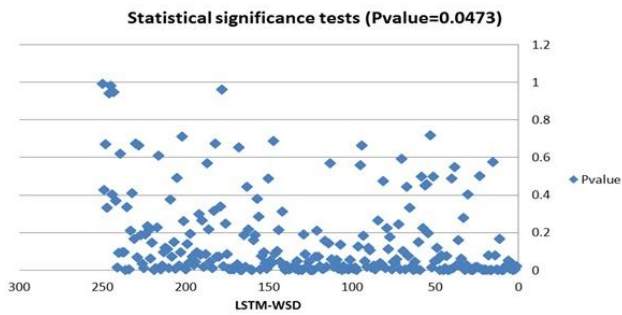
نتایج جدول ۵ نشان می‌دهد، بهترین رفتار با ترکیب LSTM-WSD و تعبیه کلمات نشان داده می‌شود و بالاترین مقادیر ACC، F_Score و AUC با تعبیه کلمات LSTM + داده شد. همچنین روش CNN-WSD با تعبیه کلمات نیز نسبت به اکثر روش‌های مورد مقایسه نتایج بهتری داشته و زمان اجرای مدل CNN نسبت به LSTM بهتر می‌باشد، نتایج بدست آمده نشان می‌دهد روش‌های LSTM، Tree LSTM و Bi LSTM زمان اجرای پایین‌تری نسبت به مدل‌های تعبیه کلمات و TF-IDF برای انجام تجزیه و تحلیل احساسات است، اما میزان دقت طبقه‌بندی این سه مدل به ترتیب ۰.۹۱٪، ۰.۸۶٪ و ۰.۸۴٪ درصد است. پس از تجزیه و تحلیل نتایج مربوط به کیفیت پیش‌بینی روش‌های مورد ارزیابی ما در جدول ۶ نتایج روش LSTM-WSD را برای مدل‌های تعبیه کلمات و TF-IDF را با روش DNN-WSD (حساس به کلمه) مقایسه می‌کنیم.

نتایج جدول ۶ نشان می‌دهد در رفتار هر دو مدل تعبیه کلمات و TF-IDF مدل مورد ارزیابی DNN-WSD برای تحلیل احساسات بر روی مجموعه داده توییت‌ها عملکرد مناسبی ندارد و مدل‌های LSTM و CNN عملکرد بهتری نسبت به DNN عملکرد بهتری دارند. جداول ۴ و ۶ نشان می‌دهد که استفاده از TF-IDF، به زمان محاسبه طولانی‌تری نسبت به استفاده از تعبیه کلمات نیاز دارد. این یکی دیگر از دلایلی است که تکنیک تعبیه کلمات را توصیه می‌کنیم. با این حال، LSTM وقت‌گیرترین الگوریتم است، هم با TF-IDF و هم با Word Embedding. با توجه به اینکه بهبود

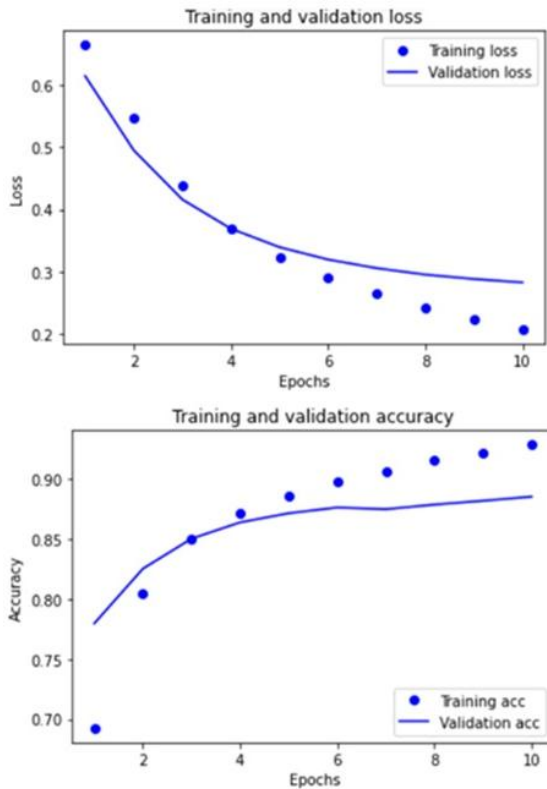
آزمون‌های معنی‌داری آماری که ما انجام دادیم راهی برای ارزیابی نتایج آزمون‌ها در اختیار ما قرار داد. سطح قابل اعتمادی که معمولاً برای نتایج استفاده می‌شود ۰.۰۵ است، همچنین به صورت $p = 0.05$ نوشته می‌شود و به عنوان سطح p شناخته می‌شود. همانطور که در شکل ۱۱ نشان داده شده است، آزمون‌های معناداری آماری را برای آموزش و اعتبار سنجی اضافه کردیم.

شکل ۱۲ عملکرد مدل پیشنهادی LSTM+WSD را بر اساس ۱۰ دوره نشان می‌دهد.

حداکثر میزان دقت برای ۱۰ تکرار ۰.۹۳٪ و میزان خطا ۰.۰۱۸۹ برای LSTM آموزش مدل تعبیه کلمه است، اما در Validation میزان دقت ۰.۸۸٪ و میزان خطا برای ۱۰ تکرار ۰.۲۸۱ است.



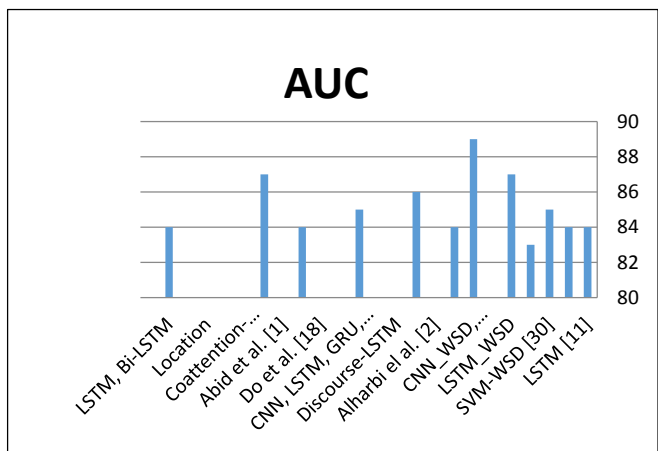
شکل ۱۱. نمودار Pvalue برای آزمون‌های معناداری آماری.



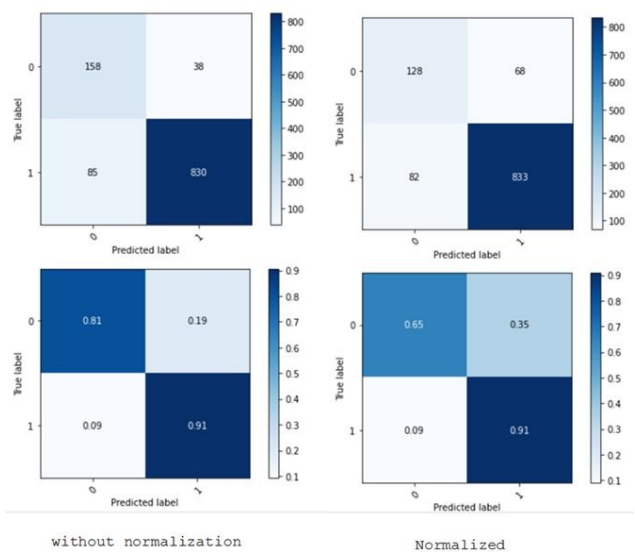
شکل ۱۲. نمودار دقت بر اساس دوره‌ها.

نتایج شکل ۹ نشان می‌دهد که میزان مساحت زیر منحنی AUC روش LSTM_WSD برابر با ۰.۸۶۴۹، CNN_WSD برابر با ۰.۸۸۹۲ است.

نمودار سردرگمی (Confusion) بر اساس سطح دقت مدل پیشنهادی LSTM و جاسازی کلمه برای مجموعه داده SentiPers طراحی شده است. این نمودار به دو صورت نرمال شده و بدون نرمال‌سازی طراحی شده است. به منظور ارزیابی عملکرد مدل LSTM و جاسازی کلمه، به صورت گرافیکی بر اساس ۱۰ دوره از مجموعه داده SentiPers نمایش داده می‌شود (شکل ۱۰).



شکل ۹. مقایسه میزان AUC روش پیشنهادی نسبت به سایر روش‌ها



شکل ۱۰. نمودار سردرگمی برای LSTM+WSD

۵-۵- بحث

با توجه به جداول ۳ تا ۶ می‌توان نتیجه گرفت که Recall، دقت، AUC و F-score روش پیشنهادی LSTM و تعبیه کلمه بهتر از سایر روش‌های ارزیابی شده است.

این نتایج برای ترکیب LSTM و TF-IDF به ترتیب ۰,۷۷۶، ۰,۷۷۴، ۰,۶۴۱ و ۰,۵۶۹ است. نتایج به‌دست‌آمده در مقایسه با تعبیه کلمات و TF-IDF برای مجموعه داده توپیتر و سایر مجموعه‌های داده مبتنی بر روش LSTM نشان می‌دهد که نتایج تعبیه کلمات در LSTM بهتر از TF-IDF است.

همچنین برای ترکیب CNN + جاسازی کلمه و DNN + جاسازی کلمه، Recall ۰,۸۰۲ و ۰,۷۹۱، میزان دقت ۰,۸۰۲، ۰,۷۹۱، FScore 0.788، 0.801 است.

نتایج به‌دست‌آمده نشان می‌دهد که روش تعبیه کلمه و LSTM با دقت ۹۳ درصد و درصد خطای ۰,۱۸۹ نسبت به سایر روش‌های ارزیابی شده عملکرد بهتری دارند. اما زمان اجرای LSTM و تعبیه کلمه بیشتر از CNN و DNN است.

در این مقاله محدودیت‌های پژوهشی وجود دارد. ابتدا، ما کلیدواژه‌های اصلی با فرکانس بالا را برای باور کاوی انتخاب کردیم و ممکن است برخی از کلیدواژه‌های مهم با فرکانس پایین نادیده گرفته شده باشند.

دوم: نتایج نشان می‌دهد که باور کاوی شامل رشته‌های بسیاری مانند علوم کامپیوتر، زبان‌شناسی و مهندسی برق است که نشان‌دهنده روند تحقیقات بین‌رشته‌ای است.

بنابراین، کارهای آینده باید از اسناد متنوع برای کشف ماهیت بین رشته‌ای استفاده کنند.

۵-۶- پیشنهادات و کاربردها

با توجه به اینکه حجم داده در شبکه‌های اجتماعی و وبسایت‌ها بالا است و بر اساس قطبیت کلمات تحلیل احساسات و اعتقادکاوی بر روی جملات انجام می‌شود پیشنهاد می‌شود از روش‌های یادگیری عمیق برای افزایش دقت طبقه‌بندی در تحلیل احساسات و اعتقادکاوی استفاده شود.

نتایج بدست آمده از این پژوهش نشان می‌دهد که عملکرد مدل یادگیری LSTM حساس به کلمه (WSD) نسبت به سایر مدل‌های یادگیری عمیق DNN، RNN و CNN بهتر است. همچنین مدل تعبیه کلمات عملکرد بهتری نسبت به TF-IDF

برای تجزیه و تحلیل نظرات دارد.

به محققان پیشنهاد می‌شود بر روی مدل‌های یادگیری عمیق LSTM و کانولوشن + تعبیه کلمات و مدل‌های ترکیبی بر اساس حساسیت به کلمات و دیدگاه کاربران فعالیت کنند.

تحلیل احساسات و اعتقادکاوی داده‌های شبکه‌های اجتماعی باعث درک نظرات کاربران شده و کاربردهای گسترده‌ای در تصمیم‌گیری و سیاست‌گذاری‌ها دارد و در زمینه‌های مختلف سیاسی، اجتماعی، اقتصادی، فرهنگی و ورزشی کاربرد دارد.

به عنوان مثال در زمینه اقتصادی تحلیل نظرات کاربران بر روی محصولات می‌تواند به فروشگاه‌ها و تولید کنندگان کمک کند تا نظرات مشتریان را تحلیل کند تا سیاست‌گذاری و تصمیم‌های مناسب را جهت افزایش جذب مشتریان انجام دهند.

در زمینه سیاسی به دولت‌ها کمک می‌کند در زمینه‌های مختلف نظرات کاربران را در شبکه‌های اجتماعی تحلیل کنند و سیاست‌گذاری‌های خود را بر اساس دیدگاه و نظرات مردم انجام دهد.

در تحقیقات آینده، با بررسی انجمن‌های احساسات-کلمه (word-sentiment) و همچنین دیدگاه‌های کاربران در حوزه‌های مختلف می‌توان تجزیه و تحلیل کامل‌تری انجام داد. یکی از محدودیت‌های مدل پیشنهادی این است که در آن فقط ویژگی‌های محلی ضبط شده‌اند. استخراج ویژگی n-gram استفاده شده در این تحقیق، شباهت معنایی یا توانایی تمایز کلمات را در نظر نمی‌گیرد. بنابراین، نمایش‌های n-gram افزایش یافته [۳۹] برای کاهش ابعاد و کمی بودن داده‌ها توصیه می‌شود. استفاده از یک روش انتخاب ویژگی مؤثر نیز ممکن است منجر به کاهش پیچیدگی محاسباتی و بهبود راندمان زمانی شود [۴۰] از طرح‌های جایگزین مبتنی بر تعبیه نیز می‌توان استفاده کرد [۴۱-۴۸].

۶- نتیجه‌گیری

در این مطالعه، ما مدل پیشنهادی LSTM، CNN، کارآمد با تفکیک شدت احساسات بر روی کلمه WSD برای استخراج نظر و اعتقادکاوی ارائه داده‌ایم. ما با انجام آزمایشات گسترده در مجموعه داده فارسی توپیتر [۴۲]، عملکرد مدل را در مقایسه با نمایش‌های اولیه کلمه ثابت کردیم. ما مدل پیشنهادی را با روش‌های موجود یادگیری عمیق و سایر روش‌های یادگیری ماشین مقایسه کردیم [۴۹-۵۵]. از این رو، اثربخشی مدل ارائه شده نشان داده شد.

- [12] Du, C. and Huang, L., (2018), Text Classification Research with Attention-based Recurrent Neural Networks, *International Journal of Computers Communications & Control*, ISSN 1841-9836, 13(1), pp. 50-61.
- [13] *Dashtipour, K., Gogate, M., Adeel, A., Ieracitano, C., Larijani, H., & Hussain, A. (2018). Exploiting deep learning for Persian sentiment analysis. In Advances in Brain Inspired Cognitive Systems: 9th International Conference, BICS 2018, Xi'an, China, July 7-8, 2018, Proceedings 9 (pp. 597-604). Springer International Publishing.*
- [14] Basiri, M.E. and kabiri, A., (2018), Words Are Important: Improving Sentiment Analysis in the Persian Language by Lexicon Refining, *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, Vol 17(4), pp. 1145-1154.
- [15] S. Urologin, Sentiment Analysis Visualization and Classification of Summarized News Articles: A Novel Approach, (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 9, No. 8, pp. 616-624, 2018.
- [16] Qiu, L. and Li, J., (2018), Sentiment analysis of short texts in microblog based on dependency parsing, *springer: Cluster Computing*, Volume 21, Issue 1, pp 985-995.
- [17] H. Alikarami, A.M. Bidgoli and M. Sadeghzadeh, Text mining of Persian texts based on Cellular Learning Automata and optimizing parameters of SVM, 4th international congress on engineering, technology & applied science, New Zealand-Auckland, 2019.
- [18] A. Onan, Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and Computation: Practice and Experience*, 33(23), e5909, 2020.
- [19] A. Onan, Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification, *Journal of King Saud University - Computer and Information Sciences*, Volume 34, Issue 5, Pages 2098-2117, 2022.
- [20] Y. Wang, M. Wang, W. Xu, A sentiment-enhanced hybrid recommender system for movie recommendation A big data analytics framework. *Wirel. Commun. Mob. Comput.* 2018, 2018.
- [21] K. Wang and X. Wan, Counterfactual Representation Augmentation for Cross-Domain Sentiment Analysis. *IEEE Transactions on Affective Computing*, 2022.
- [22] S. Mai, Y. Zeng, S. Zheng and H. Hu, (2022). Hybrid contrastive learning of tri-modal representation for multimodal sentiment analysis. *IEEE Transactions on Affective Computing*.
- [23] T. Turki and S.S. Roy, Novel Hate Speech Detection Using Word Cloud Visualization and Ensemble Learning Coupled with Count Vectorizer. *Applied Sciences*, 12(13), 6611, 2022.
- [24] Alikarami, H. and Khadem, F., (2016), Data Mining Using Genetic Algorithms and Cellular Learning Automata Based on Factor Analysis and Cluster Analysis, 1st International Conference on New Research Achievements in Electrical and Computer Engineering, Tehran, Iran.
- [25] Hajek P., Barushka A., Munk M. (2020) Opinion Mining of Consumer Reviews Using Deep Neural Networks with Word-Sentiment Associations. In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) *Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology*, vol 583. Springer, Cham, 2020. https://doi.org/10.1007/978-3-030-49161-1_35.
- [26] Joseph Turian, Lev-Arie Ratinov, and Yoshua Bengio. 2010. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 384–394, Uppsala, Sweden. Association for Computational Linguistics.
- نتایج آزمایشات نشان می‌دهد که تفکیک شدت احساسات کلمات فقط بر اساس تعبیه کلمات مؤثر هستند. ادغام WSD با N-gram، پیشرفت بیشتری را ارائه می‌دهد.
- در این پژوهش، علاوه بر تحلیل احساسات بر روی مجموعه داده، احساسات کاربران در مقوله‌های مختلف سیاسی، اقتصادی، فرهنگی و علمی مورد تجزیه و تحلیل قرار گرفت تا تحلیل‌گران بتوانند جهت‌گیری فکری کاربران در آن زمینه و دیدگاه‌های مثبت و منفی را نشان دهند. این مطالعه از WSD کارآمد با تمایز شدت شدت احساس کلمات برای عقیده کاوی و تحلیل باور استفاده کرد. مدل پیشنهادی با روش‌های یادگیری عمیق موجود و سایر روش‌های یادگیری ماشین برای نشان دادن اثربخشی مقایسه شد.

مراجع

- [1] Liu, B., 2012. Sentiment Analysis and Opinion Mining. *Synthesis lectures on human language technologies*, pp. 1-167.
- [2] Cach Dang, N. , Moreno-García, M.N. and De la Prieta, F., (2020), Sentiment Analysis Based on Deep Learning: A Comparative Study, *Electronics* 2020, 9, 483; doi:10.3390/electronics9030483.
- [3] *Alikarami, H., Bidgoli, A. M., & Javadi, H. H. S. (2023). Belief Mining in Persian Texts Based on Deep Learning and Users' Opinions. IEEE Transactions on Affective Computing*, 15(2), 632-643.
- [4] *Yao, Q.Z., Song, Z.L. and Peng, C., (2011), Research on text categorization based on LDA, Computer Engineering and Applications, Vol 47(13), pp. 150–153.*
- [5] Piryani, R., Madhavi, D. and Singh, V.K., (2017), “Analytical mapping of opinion mining and sentiment analysis research during 2000–2015,” *Information Processing & Management*, vol. 53, no. 1, pp. 122–150.
- [6] Fang, Y., Tan, H. and Zhang, J., (2018), Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness, *IEEE. Translations and content mining are permitted for academic research only*, Vol 6, pp.20625-20631.
- [7] Johnson, R., Zhang, T.: Effective use of word order for text categorization with convolutional neural networks. In: *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 103–112 (2015)
- [8] D. Tang, B. Qin, T. Liu.: Document modelling with gated recurrent neural network for sentiment classification. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1422–1432 (2015). A. Onan, Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and Computation: Practice and Experience*, 33(23), e5909, 2020.
- [9] Do, H.H.; Prasad, P.; Maag, A.; Alsadoon, A.J. Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Expert Syst. Appl.* 2019, 118, 272–299.
- [10] *Shams, M., Shakery, A., & Faili, H. (2012, May). A non-parametric LDA-based induction method for sentiment analysis. In The 16th CSI international symposium on artificial intelligence and signal processing (AISP 2012) (pp. 216-221). IEEE.*
- [11] *Basiri, M. E., Naghsh-Nilchi, A. R., & Ghassem-Aghaee, N. (2014). A framework for sentiment analysis in persian. Open transactions on information processing*, 1(3), 1-14.

- [42] Hosseini, P. et al., 2018. SentiPers: A Sentiment Analysis Corpus for Persian.
- [43] Catal, C., Nangir, M.: A sentiment classification model based on multiple classifiers. *Appl. Soft Comput.* 50, 135–141 (2017)
- [44] Dastgheib, M.B. and Koleini, S., (2019), Persian Text Classification Enhancement by Latent Semantic Space, *International Journal of Information Science and Management*, Vol 17(1), pp. 33-46.
- [45] Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A., (2016), “The rise of social bots,” *Commun. ACM*, vol. 59, no. 7, pp. 96–104.
- [46] Chen, Z.; Liu, B. Lifelong machine learning. *Synth. Lect. Artif. Intell. Mach. Learn.* 2018, 12, 1–207. [CrossRef]
- [47] Kim, Y., 2014. Convolutional Neural Networks for Sentence Classification. Doha, Qatar, s.n.
- [48] LeCun, Y., Bengio, Y. & Hinton, G., 2015. Deep learning. *Nature*, Volume 521, pp. 436-444.
- [49] Li, L.; Goh, T.-T.; Jin, D. How textual quality of online reviews affect classification performance: A case of deep learning sentiment analysis. *Neural Comput. Appl.* 2018, 1–29.
- [50] H. Alikarami, A.M. Bidgoli and M. Sadeghzadeh, Text mining of Persian texts based on Cellular Learning Automata and optimizing parameters of SVM, 4th international congress on engineering, technology & applied science, New Zealand-Auckland, 2019.
- [51] Mousavirad, S.J. and Ebrahimpour-Komleh, H., (2014), Wrapper Feature Selection using Discrete Cuckoo Optimization Algorithm, *Austrian E-Journals of Universal Scientific Organization*, Vol. 4(11), Apr, pp. 709-721.
- [52] Shayaa, S. and et al., (2018), Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges, *IEEE. Translations and content mining are permitted for academic research only*, Vol 6, pp. 37807-37827.
- [53] Singh, V.K.; Mukherjee, M.; Mehta, G.K. Combining collaborative filtering and sentiment classification for improved
- [54] Roustaei, A. and Rastegari, H., (2018), Persian question classification using headword and semantic features, *IEEE, Journal of Theoretical and Applied Information Technology*, Vol 96(21), pp. 7206-7214.
- [55] Kausar, S., Huahu, X., Shabir, M.Y., Ahmad, W.: A sentiment polarity categorization technique for online product reviews. *IEEE Access* 8, 3594–3605 (2019)
- [27] Jason Wei and Kai Zou. 2019. EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. arXiv e-prints, page arXiv:1901.11196.
- [28] Available online: <http://alt.qcri.org/semeval2017/> (accessed on 12 March 2020).
- [29] Available online: <http://help.sentiment140.com/site-functionality> (accessed on 12 March 2020).
- [30] Available online: <http://www.cs.cornell.edu/people/pabo/movie-review-data/> (accessed on 12 March 2020).
- [31] Available online: <https://www.kaggle.com/c/word2vec-nlp-tutorial/data> (accessed on 12 March 2020).
- [32] Available online: <https://www.kaggle.com/crowdflower/twitter-airline-sentiment> (accessed on 12 March 2020).
- [33] Abid, F.; Alam, M.; Yasir, M.; Li, C.J. Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter. *Future Gener. Comput. Syst.* 2019, 95, 292–308.
- [34] Alharbi, A.S.M.; de Doncker, E. Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. *Cogn. Syst. Res.* 2019, 54, 50–61.
- [35] Kraus, M.; Feuerriegel, S. Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees. *Expert Syst. Appl.* 2019, 118, 65–79.
- [36] Hassan, A. and Mahmood, A., (2018), Convolutional Recurrent Deep Learning Model for Sentence Classification, *IEEE*, Vol 6, pp. 13949 – 13957.
- [37] C. Yang, H. Zhang, B. Jiang and K.J. Li, Aspect-based sentiment analysis with alternating coattention networks. *Inf. Process. Manag.* 2019, 56, 463–478.
- [38] C. Wu, F. Wu, S. Wu, Z. Yuan, J. Liu and Y. Huang, Semi-supervised dimensional sentiment analysis with variational autoencoder. *Knowl. Based Syst.* 2019, 165, 30–39.
- [39] Chen, X., Xue, Y., Zhao, H., Lu, X., Hu, X., Ma, Z.: A novel feature extraction methodology for sentiment analysis of product reviews. *Neural Comput. Appl.* 31(10), 6625–6642 (2019)
- [40] Barushka, A., Hajek, P.: Spam detection on social networks using cost-sensitive feature selection and ensemble-based regularized deep neural networks. *Neural Comput. Appl.* 1–19 (2020)
- [41] Jeong, B.; Yoon, J.; Lee, J.-M. Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *Int. J. Inf. Manag.* 2019, 48, 280–290.

Energy-Efficient Fixed-Point Hardware Accelerator for Embedded DNNs

Marzie Mastalizade¹, Ali Ansarmohammadi², Najme Nazari³, Mostafa E. Salehi Nasab^{4*}

¹ Faculty of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

² Faculty of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

³ Faculty of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

⁴ Faculty of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

Received: 26 July 2023, Revised: 07 December 2023, Accepted: 17 December 2023

Paper type: Research

Abstract

Deep Neural Networks (DNNs) have demonstrated remarkable performance in various application domains, such as computer vision, pattern recognition, and natural language processing. However, deploying these models on edge-computing devices poses a challenge due to their extensive memory requirements and computational complexity. These factors make it difficult to deploy DNNs on low-power and limited-resource devices. One promising technique to address this challenge is quantization, particularly fixed-point quantization. Previous studies have shown that reducing the bit-width of weights and activations, such as to 3 or 4 bits, through fixed-point quantization can preserve the classification accuracy of full-precision neural networks. Despite extensive research on the compression efficiency of fixed-point quantization techniques, their energy efficiency, a critical metric in evaluating embedded systems, has not been thoroughly explored. Therefore, this research aims to assess the energy efficiency of fixed-point quantization techniques while maintaining accuracy. To accomplish this, we present a model and design an architecture for each quantization method. Subsequently, we compare their area and energy efficiency at the same accuracy level. Our experimental results indicate that incorporating scaling factors and offsets into LSQ, a well-known quantization method, improves DNN accuracy by 0.1%. However, this improvement comes at the cost of a 3× decrease in hardware energy efficiency. This research highlights the significance of evaluating fixed-point quantization techniques not only in terms of compression efficiency but also in terms of energy efficiency when applied to edge-computing devices.

Keywords: Deep Neural Network, Embedded Systems, Energy-Efficiency, Fixed-point Quantization.

* Corresponding Author's email: mersali@ut.ac.ir

بررسی سربارهای سخت‌افزاری و بهره‌وری انرژی در پیاده‌سازی انواع چندی‌سازی ممیز ثابت در شتاب‌دهنده شبکه عصبی عمیق

مرضیه مستعلی‌زاده^۱، علی انصاری‌محمدی^۲، نجمه نظری^۳، مصطفی ارسالی صالحی نسب^۴*

^۱ گروه معماری سیستم‌های کامپیوتری، دانشکده مهندسی برق و کامپیوتر، دانشکده فنی، دانشگاه تهران، تهران، ایران
^۲ گروه معماری سیستم‌های کامپیوتری، دانشکده مهندسی برق و کامپیوتر، دانشکده فنی، دانشگاه تهران، تهران، ایران
^۳ گروه معماری سیستم‌های کامپیوتری، دانشکده مهندسی برق و کامپیوتر، دانشکده فنی، دانشگاه تهران، تهران، ایران
^۴ گروه معماری سیستم‌های کامپیوتری، دانشکده مهندسی برق و کامپیوتر، دانشکده فنی، دانشگاه تهران، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۵/۰۴ تاریخ بازبینی: ۱۴۰۲/۰۹/۱۶ تاریخ پذیرش: ۱۴۰۲/۰۹/۲۶

نوع مقاله: پژوهشی

چکیده

یکی از کارآمدترین راه‌کارهای فشرده‌سازی و کاهش انرژی مصرفی شبکه‌های عصبی عمیق در دستگاه‌های نهفته، چندی‌سازی با استفاده از نمایش اعداد ممیز ثابت است. در سال‌های اخیر، روش‌های متنوعی برای بهبود صحت شبکه‌های چندی‌سازی شده مطرح شده است که اغلب سربارهای محاسباتی زیادی به شبکه تحمیل می‌کنند، اگرچه این موضوع تاکنون از دید طراحان شبکه‌های عصبی عمیق پنهان مانده است. در این پژوهش، روش‌های مختلف چندی‌سازی ممیز ثابت، بر اساس مؤلفه‌های تأثیرگذار در سربارهای سخت‌افزاری، طبقه‌بندی و مدل شده است. پس از آن، معماری‌های سخت‌افزاری ارائه شده برای هر یک از مدل‌ها به صورت عادلانه، با در نظر گرفتن هزینه‌فایده‌ی بین صحت شبکه و بهره‌وری انرژی سخت‌افزار، بررسی و مقایسه می‌شوند. نتایج نشان می‌دهد تکنیک‌هایی که برای کاهش خطای روش‌های چندی‌سازی به کار گرفته می‌شود، اگرچه به افزایش صحت شبکه‌های عصبی منجر می‌شود اما از طرف دیگر بهره‌وری انرژی سخت‌افزار را کاهش می‌دهد. براساس نتایج شبیه‌سازی، افزودن ضریب مقیاس و آفست به چندی‌سازی ممیز ثابت LSQ، صحت شبکه را حدود ۰/۱ افزایش می‌دهد اما بهره‌وری انرژی سخت‌افزار حدود سه برابر کمتر شده است. این موضوع لزوم توجه به سربارهای سخت‌افزاری را به خصوص در سیستم‌های نهفته، بیش از پیش نشان می‌دهد.

کلیدواژه‌گان: شبکه‌های عصبی عمیق، سیستم‌های نهفته، بهره‌وری انرژی، چندی‌سازی ممیز ثابت.

* رایانامه نویسنده مسؤول: mersali@ut.ac.ir

۱- مقدمه

ممیزثابت را تا ۸ بیت کاهش داده‌اند. در نتیجه، در مقایسه با محاسبات ۳۲ بیتی تقریباً تا ۴ برابر در حافظه و محاسبات بهبود حاصل شده است. معماری‌های بسیاری [۱۷، ۱۸] برای اجرای شبکه‌های عصبی بر مبنای چندی‌سازی ممیزثابت ارائه شده است که هدف آنها رسیدن به حداکثر کارایی انرژی بوده است. در این معماری‌ها سربارهای روش نمایش وزن‌ها و فعال‌سازها^۹ به صورت یک عدد ممیزثابت در نظر گرفته نشده است.

محققان برای دستیابی به حداقل خطای چندی‌سازی از روش‌های دیگری نیز در جهت بهبود دقت استفاده کرده‌اند. مقاله‌های [۱۱]، [۱۹] تلاش کرده‌اند تا محدوده‌ی پویای چندی‌سازی را در حین آموزش شبکه بهتر تنظیم کنند و از این طریق توانسته‌اند تعداد بیت مورد نیاز را به ۴ بیت کاهش دهند. نکته‌ای که در این روش‌ها از منظر پیاده‌سازی سخت‌افزاری پنهان مانده، این است که نحوه‌ی نمایش وزن‌ها و فعال‌سازها فراتر از یک عدد ممیزثابت تنها است. بدین معنی که در این روش‌ها برای نمایش وزن‌ها (یا فعال‌سازها^{۱۰}) از حاصل‌ضرب یک عدد ممیزثابت و ضریب مقیاس^{۱۰} استفاده شده است و مقادیر ضریب مقیاس به صورت اعداد ممیزثابت در نظر گرفته شده‌اند که در عملیات سخت‌افزاری هزینه‌ی زیادی دارند و به این هزینه‌ها توجه نشده است.

مقاله‌های [۲۰-۲۲] هر کدام با استفاده از تکنیک‌های مخصوص به خود سعی کرده‌اند که خطای چندی‌سازی و تعداد بیت را کاهش دهند و در نهایت، با استفاده از حدود ۳ یا ۴ بیت، به صحت‌های نزدیک به صحت کامل شبکه برسند. هرچند با استفاده از این روش‌ها، خطای چندی‌سازی با استفاده از نمایش اعداد ممیزثابت به حداقل رسیده اما نیاز است که مقادیر چندی‌سازی شده با استفاده از تعدادی پارامتر با دقت کامل مقیاس شوند و همچنین برای انتقال به بازه‌ی خاصی از اعداد، با یک مقدار آفست جمع شوند. در این روش‌ها نیز، به تأثیر پارامترهای ۳۲ بیتی افزوده‌شده (یعنی ضریب مقیاس و آفست) بر روی هزینه‌ی سخت‌افزار توجه زیادی نشده است. با استفاده از این روش‌ها، شبکه‌های فشرده‌شده برای اجرا روی سیستم‌های نهفته مناسب‌تر خواهند شد البته هم‌چنان حجم محاسبات ضرب‌وانباشت در آنها زیاد و نیازمند بهینه‌سازی است.

همان‌طور که گفته شد، روش‌های مختلف چندی‌سازی ممیزثابت با استفاده از حدود ۳ یا ۴ بیت به صحت‌های نزدیک به صحت کامل

امروزه شبکه‌های عصبی عمیق^۱ کاربردهای مهم و گسترده‌ای در حوزه‌های مختلف به‌دست آورده‌اند که از آن جمله می‌توان به طبقه‌بندی تصاویر^۲، بینایی ماشین^۳ و تشخیص اشیاء^۴ اشاره کرد. عملکرد بسیار خوب این شبکه‌ها در حوزه‌های مختلف باعث شده است که اخیراً در سیستم‌های نهفته^۵ نیز مورد استقبال قرار گیرند.

با این وجود، حجم بسیار زیاد محاسبات شبکه‌های عصبی عمیق و فضای زیادی که برای ذخیره‌سازی داده‌ها نیاز دارند باعث شده که ورود آنها به سیستم‌های نهفته با چالش روبرو شود زیرا این سیستم‌ها بودجه‌ی محدودی برای انرژی مصرفی و فضای ذخیره‌سازی دارند. به‌عنوان مثال، شبکه‌ی ResNet-50 با ۵۰ لایه‌ی مختلف حدود ۲۵/۵ میلیون داده دارد و باید ۳/۹ میلیارد عملیات ضرب‌وانباشت^۶ را انجام دهد [۱]. به همین دلیل، روش‌های فشرده‌سازی شبکه، از جمله چندی‌سازی^۷ و هرس کردن^۸ شبکه [۲-۴] معرفی شدند تا حجم محاسبات و فضای ذخیره‌سازی را کاهش دهند.

در میان روش‌های مختلف فشرده‌سازی شبکه‌های عصبی عمیق، چندی‌سازی محبوبیت بیشتری پیدا کرده و پژوهش‌های بسیاری در این حوزه انجام شده است [۵-۱۲]. چندی‌سازی به بیان ساده، نگاشت اعداد حقیقی به اعداد گسسته‌ی محدود، به منظور کاهش هزینه‌ی حافظه و محاسبات است. در همین راستا، چندی‌سازی‌های مختلف معرفی شده‌اند که از نمایش اعداد متنوعی چون ممیزثابت، باینری، توان ۲ و ممیزثابت استفاده کرده‌اند و برای رسیدن به صحت قابل قبول عرض بیت‌های ۱۶ تا ۱ بیت را به‌کار گرفته‌اند.

نتایج اولین پژوهش‌ها نشان می‌دهد که چندی‌سازی ممیزثابت، برای حفظ صحت شبکه‌های عصبی به ۱۶ بیت نیاز داشت [۱۳]، [۱۴]. اگر چه چندی‌سازی ممیزثابت ۱۶ بیت می‌تواند حجم محاسبات و فضای ذخیره‌سازی موردنیاز را تا ۲ برابر به نسبت ممیزثابت ۳۲ بیتی کاهش دهد، اما به دلیل محدودیت‌هایی که برای کاهش منابع دارد، نمی‌تواند راه‌حل مناسبی برای سیستم‌های نهفته باشد.

مقاله‌های [۱۵، ۱۶] تعداد بیت‌های لازم برای چندی‌سازی

⁶ Multiply Accumulate

⁷ Quantization

⁸ Pruning

⁹ Activation

¹⁰ Scaling Factor

¹ Deep Neural Networks (DNN)

² Image Classification

³ Computer Vision

⁴ Object Recognition

⁵ Embedded Systems

است و معماری مربوط به هر مدل را ارائه کرده‌ایم و در نهایت در بخش ششم نتایج انرژی و توان مصرفی مدل‌های مختلف با در نظر گرفتن تعداد بیت و صحت شبکه ارائه شده است. در پایان، در بخش هفتم نتیجه‌گیری آورده شده است.

۲- مفاهیم اولیه

در این بخش به بررسی مفاهیم اولیه‌ی مورد نیاز در این مقاله می‌پردازیم. این مفاهیم شامل پیش‌زمینه‌هایی از شبکه‌های عصبی عمیق و چندی‌سازی است. سعی کرده‌ایم موضوعاتی را که ارائه می‌کنیم، مفاهیم اصلی این حوزه را در برگیرد تا از ایجاد ابهامات احتمالی جلوگیری شود.

۲-۱- شبکه‌های عصبی عمیق

شبکه‌های عصبی عمیق، به‌عنوان زیرمجموعه‌ای از یادگیری ماشین، دسته‌ای از الگوریتم‌ها هستند که برای شناسایی و تشخیص الگوها استفاده می‌شوند. این شبکه‌ها با استفاده از یک ابزار ادراکی ماشینی^۳، داده‌های خام را دریافت می‌کنند و سعی می‌کنند الگوهای موجود در این داده‌ها را شناسایی کنند. این ویژگی سبب شده است که شبکه‌های عصبی عمیق امروزه کاربردهای گسترده‌ای داشته باشند و با توجه به این کاربردها، امروزه تحقیقات بسیاری در این زمینه و چالش‌های پیاده‌سازی آن صورت می‌گیرد.

ساده‌ترین شبکه‌های عصبی از چند لایه‌ی محدود تشکیل می‌شده‌اند. به عنوان مثال، شبکه‌های پرسپترون چند لایه^۴ فقط می‌توانند تصاویر ساده‌ای را پردازش کنند که تعداد اشیاء موجود در آن محدود باشد. با پردازش این تصاویر، شبکه‌ها می‌توانند اشیاء موجود را طبقه‌بندی کنند. برای مثال، مجموعه داده‌ی MNIST را که شامل دست‌نوشته‌ی اعداد صفر تا ۹ است، می‌توان با این شبکه‌ها دسته‌بندی کرد.

در مسائل واقعی، جزئیات تصاویر و اشیاء از جمله رنگ، موقعیت و نور بسیار متنوع و زیاد است. هم‌چنین ممکن است اشیاء یک تصویر همیشه واضح نباشد، مثلاً در یک تصویر اشیاء بسیار زیاد و نامرتب وجود داشته باشد یا اینکه فقط بخشی از شیء مورد نظر در تصویر وجود داشته باشد، در این شرایط تشخیص و دسته‌بندی اشیاء به سادگی قیل نیست و شبکه‌های پرسپترون چند لایه دیگر نمی‌توانند اشیاء را دسته‌بندی کنند. به همین دلیل، شبکه‌های عصبی کانولوشنی برای پردازش تصاویر پیچیده‌تر معرفی شدند.

شبکه رسیده‌اند. در نمایش اعداد ممیزثابت هرچند هزینه‌ی سربارهای اضافی ضریب مقیاس و آفست در نظر گرفته نشده است، اما به دلیل محبوبیت و رواج زیاد، بهینه‌سازی‌های متنوعی بر روی معماری آن صورت گرفته است. براساس تحقیقات ما، تا کنون هیچ پژوهشی به طور خاص این روش‌ها را از نظر ملاحظات پیاده‌سازی سخت‌افزاری مثل توان مصرفی^۱ و بهره‌وری انرژی^۲ بررسی و مقایسه نکرده است. ما در این مقاله قصد داریم ابتدا چندی‌سازی‌های مختلف را با توجه به مؤلفه‌های مهم و تأثیرگذار در پیاده‌سازی سخت‌افزاری طبقه‌بندی کنیم و به منظور ساده‌تر شدن طبقه‌بندی، یک مدل برای هر طبقه ارائه می‌کنیم. سپس برای هر مدل با توجه به پژوهش‌های انجام‌شده و چندی‌سازی‌های مختلف ارائه‌شده، بهترین چندی‌سازی را از نظر صحت شبکه انتخاب می‌کنیم تا در ادامه، هزینه‌ی سربارهای سخت‌افزاری با توجه به آن چندی‌سازی‌ها مقایسه شوند. سپس برای هر مدل، یک معماری طراحی می‌کنیم و در نهایت نتایج توان و انرژی مصرفی مدل‌های مختلف چندی‌سازی به ازای عرض بیت و صحت‌های مختلف را مقایسه می‌کنیم. به بیان دیگر، چالش اصلی مقاله این است که برای مدل‌های مختلف چندی‌سازی‌شده که در مرحله‌ی آموزش استفاده می‌شوند و سربارهای مختلف و تأثیرگذاری در سخت‌افزار دارند، پارامترهای انرژی و توان سخت‌افزار مقایسه شوند. به طور خلاصه نوآوری‌های این مقاله به شرح زیر است:

- بررسی چندی‌سازی‌های ممیزثابت از نظر خطای چندی‌سازی و صحت شبکه و طبقه‌بندی آنها بر مبنای مؤلفه‌های تأثیرگذار در سخت‌افزار.
- ارائه‌ی یک مدل برای هر طبقه از چندی‌سازی‌ها.
- ارائه‌ی روابط یک‌دست برای عملیات ضرب داخلی دو بردار با توجه به نمایش پارامترها در هر مدل.
- ارائه‌ی یک معماری پایه برای تحقق و ارزیابی سخت‌افزاری هر مدل از چندی‌سازی.
- بررسی ملاحظات پیاده‌سازی سخت‌افزاری برای هر مدل از چندی‌سازی.

در ادامه، در بخش دوم مقاله، ابتدا به بررسی مفاهیم اولیه مرتبط با شبکه‌های عصبی عمیق و سپس روش‌های مختلف چندی‌سازی پرداخته‌ایم. در قسمت سوم کارهای پیشین بررسی شده و در قسمت چهارم انگیزه‌ی انجام این مقاله بیان شده است. سپس در بخش پنجم، برای هر کدام از روش‌های چندی‌سازی یک مدل معرفی شده

³ Machine Perception

⁴ Multi-layer Perceptron

¹ Power Consumption

² Energy Efficiency

معنای دیگر، وقتی کرنل روی ماتریس نگاشت ورودی حرکت می‌کند، هر بار پنجره‌ای از نگاشت ورودی با اندازه‌ی $S \times R$ انتخاب می‌شود. می‌توان کرنل و پنجره‌ی انتخاب‌شده از نگاشت ورودی را به صورت دو بردار در نظر گرفت که اعضای آنها باید نظیر به نظیر در هم ضرب و در نهایت با هم جمع شوند. این، همان تعریف ضرب داخلی است. برای دو بردار دلخواه $\bar{X}, \bar{Y} \in \mathbb{R}^K$ تعریف می‌کنیم:

$$\text{dot}(\bar{X}, \bar{Y}) = \sum_{i=0}^{K-1} x_i \times y_i \quad (2)$$

که در آن، تابع dot ضرب داخلی را نشان می‌دهد و $x_i \in \bar{X}$ و $y_i \in \bar{Y}$.

همان‌طور که در بخش‌های قبل گفته شد، شبکه‌های عصبی عمیق کاربردهای فراوانی در حوزه‌های مختلف پیدا کرده‌اند. برای یادگیری داده‌های پیچیده و صحت بالا، به شبکه‌های عصبی عمیقی نیاز داریم که تعداد لایه‌های آن زیاد باشد و از این‌رو محاسبات بسیار زیادی نیاز است، به همین دلیل بار محاسباتی شبکه و حافظه‌ی مورد نیاز برای ذخیره‌ی پارامترهای شبکه بسیار زیاد است [۱].

معمولاً هر چه تعداد لایه‌های شبکه بیشتر باشد، شبکه به صحت بالاتری دست پیدا می‌کند. شبکه‌های عصبی عمیق شامل تعداد بسیار زیادی عملیات ضرب و انباشت اعداد ممیزشناور است که برای این عملیات، حجم بسیار زیادی از داده باید از حافظه خوانده شود. همچنین ممکن است در محاسبات شبکه‌های عصبی، به علت تکرار عملیات، یک داده چند بار از حافظه خوانده شود. خواندن این حجم زیاد از داده‌ها نیاز به پهنای باند زیاد دارد و باعث ایجاد گلوگاه در ارتباطات حافظه می‌شود.

در حوزه‌ی سیستم‌های نهفته و کم‌مصرف، مصرف انرژی و حافظه یک چالش اساسی است. به همین دلیل، در طراحی شبکه‌های عصبی عمیق برای این حوزه، علاوه بر کارایی و صحت شبکه، معیارهای دیگری مثل منابع محاسباتی مورد نیاز، اندازه‌ی حافظه و مهم‌تر از آنها، مصرف انرژی نیز باید در نظر گرفته شود. برای حل این چالش‌ها، راهکارهای مختلفی معرفی شده که از جمله‌ی آنها می‌توان به کوچک کردن شبکه، کاهش تعداد بیت عملگر و عملوند و کم کردن تعداد عملگر و عملوند اشاره کرد.

۲-۲- چندی‌سازی

راهکارهایی که برای کاهش تعداد بیت عملگر و عملوند به کار

شبکه‌ی عصبی کانولوشنی جزو پرکاربردترین روش‌ها در یادگیری عمیق است. این شبکه‌ها می‌توانند در محیط واقعی و پیچیده که اشیاء بدون الگوی خاصی در کنار هم قرار دارند، ویژگی‌های مؤثر را استخراج و اطلاعات غیر مرتبط را حذف کنند. سپس بر اساس ویژگی‌های مؤثر استخراج‌شده، اشیاء را دسته‌بندی کنند.

شبکه‌های عصبی کانولوشنی عموماً از چند لایه‌ی کانولوشن، نمونه‌برداری^۱، نرمال‌سازی^۲ و تماماً متصل^۳ تشکیل شده‌اند. از میان این لایه‌ها، تمرکز این پژوهش بر لایه‌های کانولوشنی است. البته از آنجایی که عملیات لایه‌ی کانولوشن و تماماً متصل از نظر ماهیت یکسان است و تفاوت‌های جزئی دارد، نتایج این پژوهش را برای حوزه‌ی لایه‌ی تماماً متصل نیز می‌توان در نظر گرفت. در ادامه به توضیح لایه‌ی کانولوشن می‌پردازیم.

یکی از وظایف لایه‌های کانولوشن این است که ویژگی‌های مؤثر داده‌های ورودی را برجسته سازند و آنها را استخراج کنند. به همین منظور، در این لایه تعدادی فیلتر که به آن کرنل^۴ گفته می‌شود، روی تصویر ورودی حرکت می‌کند و عملیات کانولوشن را انجام می‌دهد. این عملیات در رابطه‌ی (۱) نشان داده شده است. سپس یک تابع غیر خطی روی حاصل کانولوشن اعمال می‌شود و نگاشت‌های ویژگی^۵ به دست می‌آید که در واقع همان ویژگی‌های مهم استخراج شده از تصویر است.

$$O[z][u][x][y] = B[u] + \sum_{k=0}^{C-1} \sum_{i=0}^{S-1} \sum_{j=0}^{R-1} I[z][k][Ux+i][Uy+j] \times W[u][k][i][j]$$

$$0 \leq z < N, 0 \leq u < M, 0 \leq x < F, 0 \leq y < E,$$

$$E = \frac{H-R+U}{U}, F = \frac{W-S+U}{U}. \quad (1)$$

در رابطه‌ی (۱)، مقدار O, I, W و B به ترتیب ماتریس نگاشت ویژگی خروجی، نگاشت ویژگی ورودی، فیلترها و بایاس هستند. U اندازه‌ی گام است که در واقع نشان می‌دهد به چه میزان روی ورودی حرکت می‌کند.

کانولوشن یک عملیات اصلی در شبکه‌های عصبی است. با توجه به رابطه‌ی (۱) مشخص می‌شود که این عملیات شامل تعداد بسیار زیادی ضرب داخلی^۶ بین بردار ورودی (I) و وزن (W) است. به

⁴ Kernel

⁵ Feature Maps

⁶ Dot Product

¹ Pooling

² Normalization

³ Fully Connected (FC)

متفاوتی را در نظر گرفته است و با این کار توانسته است میانگین تعداد بیت برای لایه‌ها را به ۸ تا ۱۲ بیت برساند. مقاله‌های [۱۵]، [۱۶] چندی‌سازی ممیزثابت را با یادگیری و تنظیم دقیق بعد از یادگیری شبکه همراه کرده‌اند تا بهترین مقدار برای تعداد بیت چندی‌سازی و همچنین تعداد بیت قسمت اعشار ممیزثابت را پیدا کنند. در این معماری‌ها فرض بر استفاده از تنها یک عدد ممیز ثابت برای چندی‌سازی بوده است و هیچ سربار دیگری برای نمایش اعداد لازم نیست. به عنوان مثال معماری‌های [۱۷، ۱۸] با توجه به چندی‌سازی ممیزثابت جهت رسیدن به حداکثر کارایی انرژی ارائه شده است.

مقاله‌های [۱۱، ۱۹] با استفاده از پیدا کردن بازه‌ی پویای مناسب برای چندی‌سازی حین آموزش شبکه توانسته‌اند تعداد بیت مورد نیاز را به کمتر از ۸ بیت کاهش دهند. نحوه‌ی نمایش وزن‌ها و فعال‌سازها در این چندی‌سازی‌ها فراتر از تنها یک عدد ممیزثابت است. در این روش‌ها برای نمایش وزن‌ها (فعال‌سازها) از حاصل ضرب یک عدد ممیزثابت و ضرب مقیاس استفاده شده است. مقادیر ضرب مقیاس به صورت اعداد ممیزشناور در نظر گرفته شده‌اند که انجام عملیات مختلف روی آنها هزینه‌ی زیادی دارد و به این نکته در این روش‌ها از جهت پیاده‌سازی در سخت‌افزار توجه نشده است.

مقاله‌ی DSQ [۲۱] با استفاده از چندی‌سازی تفاضلی نرم^۱، توانسته است صحت شبکه‌های چندی‌سازی شده با تعداد بیت پایین را به شبکه‌های صحت کامل نزدیک کند. این مقاله می‌تواند به دقیق بودن گرادینان در انتشار به عقب^۲ و کاهش خطای چندی‌سازی در فرآیند رو به جلو^۳ با یک محدوده برش^۴ مناسب کمک کند. اگرچه این روش صحت شبکه‌ی چندی‌سازی شده را افزایش داده است ولی نکته‌ای که در پیاده‌سازی سخت‌افزاری این روش به آن توجه نشده، این است که وزن‌ها و فعال‌سازها به مقادیر مقیاس شده و همراه با آفست تبدیل می‌شوند. این موضوع باعث می‌شود که نتوان عملیات ضرب یک وزن در یک فعال‌ساز را تنها با یک عملیات ضرب انجام داد.

مقاله‌ی QIL [۱۹] با استفاده از ترکیب یک تابع انتقال و گسسته‌ساز توانسته است مشکلات حاصل از چندی‌سازی را کاهش دهد و صحت شبکه چندی‌سازی شده را افزایش دهد. در این مقاله به دو نکته‌ی محدوده‌ی برش و هرس کردن (تبدیل اعداد کوچک به صفر) توجه ویژه شده است. این آستانه‌ها، در حین یادگیری شبکه با استفاده از پارامتری کردن تابع انتقال به صورت خودکار به دست می‌آیند. پس

می‌روند، عموماً داده‌های ورودی و وزن شبکه را با تعداد بیت کمتر نشان می‌دهند و از این طریق تلاش می‌کنند تا فضای ذخیره‌ی موردنیاز را کاهش دهند و محاسبات بین آنها را ساده‌تر سازند.

یکی از اصلی‌ترین روش‌ها در این حوزه چندی‌سازی است که در آن، داده‌های حقیقی که عموماً ممیز شناور ۳۲ بیتی‌اند، با تعداد بیت کمتر نشان داده می‌شوند. در چندی‌سازی، چندین سطح برای تخمین اعداد وجود دارد که هر کدام از اعداد حقیقی، با روش‌های مختلف به یکی از سطوح تعلق می‌گیرد. تعداد این سطوح، دقت و تعداد بیت مورد نیاز برای نمایش داده را تعیین می‌کند. به طور کلی تعداد بیت مورد نظر را از رابطه‌ی $\log_2 N$ به دست می‌آورند که N تعداد سطوح را مشخص می‌کند. با کاهش تعداد بیت لازم برای نمایش، فضای ذخیره‌سازی مورد نیاز کاهش می‌یابد، همچنین محاسبات ساده‌تر می‌شود.

روش‌های مختلف چندی‌سازی، شیوه‌های متفاوتی برای انجام این کار دارند. در این روش‌ها، پارامترهای هدف (وزن، ورودی یا هردو)، الگوریتم به کار رفته، نمادها و نمایش اعداد متفاوت‌اند. برای اینکه در ادامه این مقاله از مفاهیم یکسان استفاده کنیم، تعریف می‌کنیم؛

$$x^q = Q(x) \quad (3)$$

که در آن Q تابع چندی‌سازی است که عدد حقیقی x را به عدد چندی‌سازی شده‌ی x^q تبدیل می‌کند. عرض بیت x و x^q هر عددی می‌تواند باشد و در روش چندی‌سازی مشخص می‌شود، اما عموماً x یک عدد ممیزشناور ۳۲ بیتی است.

۳- بررسی کارهای پیشین

چندی‌سازی یکی از راه‌کارهای مؤثر در فشرده‌سازی شبکه‌های عصبی عمیق است که یکی از اهداف ویژه‌ی آن، امکان پیاده‌سازی این شبکه‌ها بر روی دستگاه‌های نهفته است. در چندی‌سازی، تعداد بیت مورد نیاز برای پارامترهای شبکه کاهش داده می‌شود تا از طریق کاهش حافظه و هزینه انجام عملیات محاسباتی، کارایی انرژی شبکه عصبی بر روی دستگاه نهفته افزایش یابد [۲۳-۲۷]. یکی از مؤلفه‌های مهم در افزایش کارایی انرژی، معماری واحدهای محاسباتی است که ارتباط مستقیم با سیستم نمایش اعداد دارد. متداول‌ترین و مرسوم‌ترین روش چندی‌سازی، نمایش اعداد ممیز ثابت بوده است. در این حوزه، [۹، ۱۴، ۲۸] نشان داده‌اند که با ۱۶ بیت چندی‌سازی، شبکه به صحت قابل قبولی می‌رسد. مقاله‌ی [۹] با توجه به توزیع متفاوت وزن‌ها در لایه‌ها، برای هر لایه تعداد بیت

³ Feed Forward

⁴ Clipping Range

¹ Differentiable Soft Quantization

² Back Propagation

تحمیل شده به سخت‌افزار توجه نشده است.

در مقاله‌هایی شبیه به bitpruning، وزن‌ها و فعال‌سازها به صورت حاصل ضرب یک عدد ممیز ثابت در ضریب مقیاس در نظر گرفته شده‌اند که با یک آفست جمع می‌شوند تا اعداد به یک بازه‌ی مشخص منتقل شوند و از حداکثر بازه‌ی پویای قابل نمایش با اعداد ممیز ثابت استفاده شود. در این حالت، هر پارامتر از مجموع دو مؤلفه‌ی مجزا تشکیل شده است. لذا هنگام ضرب دو پارامتر، خروجی مورد نظر شامل ۴ مؤلفه‌ی ضرب جزئی^۱ است که طبیعتاً هزینه‌ی بسیار بیشتری نسبت به استفاده از تنها یک عدد با نمایش اعداد ممیز ثابت خواهد داشت.

مقاله‌ی CSQ [۲۹] با اشاره به اینکه نابرابر بودن تعداد اعداد مثبت و منفی موجود در نمایش اعداد می‌تواند سبب بهینه نبودن خطای چندی‌سازی باشد، یک چندی‌سازی متقارن معرفی می‌کند که علاوه بر یکسان کردن تعداد سطوح مثبت و منفی، می‌تواند کماکان از سخت‌افزارهای استاندارد موجود برای انجام عملیات ضرب بهره‌بردار.

جدول ۱ چندی‌سازی‌های مختلف ممیز ثابت به همراه سربارهای تأثیرگذار بر سخت‌افزار را نشان می‌دهد. وجود این سربارها در روش‌های مختلف و از طرف دیگر صحت‌های مختلف این روش‌ها سبب شد تا ما در این پژوهش به دنبال مقایسه‌ی این چندی‌سازی‌ها از منظر ملاحظات پیاده‌سازی سخت‌افزاری باشیم.

۴- انگیزه‌ی مقاله و تعریف مسأله

هدف اصلی ما در انجام این پژوهش مقایسه‌ی منصفانه و عادلانه میان روش‌های مختلف چندی‌سازی ممیز ثابت است. در واقع، ما قصد داریم به این سوال پاسخ دهیم که در صحت یکسان، کدام روش چندی‌سازی بهره‌وری انرژی بیشتری دارد؟

صحت روش‌های مختلف چندی‌سازی ممیز ثابت در جدول ۲ نشان داده شده است. روش‌های مختلف چندی‌سازی بر روی شبکه‌های عصبی مختلف از جمله ResNet-18، ResNet-34 و ResNet-50 [۳۲] که به‌وسیله‌ی مجموعه‌داده‌ی ImageNet آموزش داده شده‌اند، اعمال شده است. در این جدول، صحت اصلی شبکه‌ها و صحت شبکه‌های چندی‌سازی شده در عرض بیت‌های ۱ تا ۸ گزارش شده است.

مقایسه‌ی صحت روش‌های مختلف با صحت اصلی شبکه در جدول ۲ نشان می‌دهد که روش‌های مختلف چندی‌سازی ممیز ثابت، با حدود ۲ تا ۴ بیت برای وزن و فعال‌سازها، به صحت‌هایی بسیار

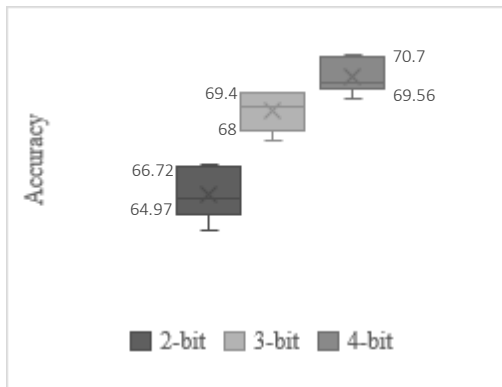
از آن، خروجی تابع انتقال را (که بین بازه‌ی ۱ و ۱- است) به تعداد بیت محدود چندی‌سازی کرده است. نکته‌ی مهم در خروجی چندی‌سازی، نمایش مقدار چندی‌سازی شده با یک عدد ممیز ثابت است که باعث کاهش سربارهای سخت‌افزاری می‌شود.

مقاله‌ی LSQ [۱۱] در حین آموزش شبکه، از یک چندی‌سازی آموزش‌پذیر استفاده می‌کند تا از این طریق، بهترین اندازه‌ی گام را برای سطوح چندی‌سازی به دست آورد. اندازه‌ی گام برای سطوح چندی‌سازی به صورت یک عدد دقت کامل در نظر گرفته شده است و در واقع نقش ضریب مقیاس را در چندی‌سازی دارد. هر چند در این مقاله اشاره شده است که محاسبات مربوط به این ضریب مقیاس می‌تواند با عملیات نرمال‌سازی لایه‌ی بعد ادغام شود اما هزینه‌ی اعمال شده به سخت‌افزار می‌تواند قابل ملاحظه باشد. در ادامه، نویسندگان این مقاله به منظور کاهش خطای چندی‌سازی، با توجه به اینکه سطوح چندی‌سازی را نامتقارن در نظر گرفته‌اند، از مقدار آفستی استفاده می‌کنند تا مقادیر وزن‌ها را با دقت بیشتری به سطوح چندی‌سازی نگاشت کنند. این کار صحت شبکه را بهبود داده است. البته این مقاله به سربار و هزینه‌ی سخت‌افزاری این آفست اشاره‌ای نمی‌کند (LSQ+ [۶]).

مقاله‌ی uL2Q [۲۰] با استفاده از تجزیه و تحلیل جامع داده‌های کمی به دنبال پیدا کردن بهترین چندی‌سازی برای داده‌های با توزیع نرمال است. این مقاله، در ابتدا وزن‌های شبکه را با فرض توزیع نرمال دریافت می‌کند. سپس توزیع آنها را به توزیع نرمال استاندارد تبدیل می‌کند. از این طریق، برای هر تعداد بیت چندی‌سازی بهترین ضرایب مورد نیاز برای مقیاس به توزیع نرمال را مشخص می‌کند. در نهایت، برای استفاده از مقدار چندی‌سازی شده در شبکه نیاز است که عملیات استانداردسازی به صورت معکوس انجام شود. برای این کار، باید دو پارامتر ضریب مقیاس و آفست، به ازای هر بخش از داده به صورت دقت کامل آموزش ببینند که هزینه‌ی آن باید در معماری سخت‌افزار در نظر گرفته شود.

علاوه بر این، مقاله‌ی bitpruning [۲۲] آفست‌هایی را برای انتقال مقادیر وزن و فعال‌ساز تعریف کرده است تا از حداکثر بازه‌ی پویای قابل نمایش با تعداد بیت محدود استفاده کند و از این طریق، خطای چندی‌سازی با تعداد بیت کمتر را کاهش دهد. در این مقاله تعداد بیت مناسب برای هر لایه یا دسته‌ای از اعداد، به صورت خودکار در زمان آموزش به دست می‌آید. هر دو مقدار آفست در این مقاله، به صورت دقت کامل در نظر گرفته شده است و البته به هزینه‌های

^۱ Partial Product



شکل ۱. پراکندگی صحت روش‌های مختلف چندی‌سازی ممیز ثابت در بازه‌ی ۲ تا ۴ بیت بر روی ResNet-18 که با مجموعه‌داده‌ی ImageNet آموزش داده شده است.

روش‌های مختلف چندی‌سازی ممیزثابت، سیستم نمایش اعداد متفاوتی دارند. به همین دلیل، سخت‌افزار مورد نیاز برای اجرای آنها متفاوت خواهد بود.

جدول ۲. مقایسه‌ی صحت روش‌های مختلف چندی‌سازی ممیزثابت در شبکه‌های مختلف بر روی مجموعه‌داده‌ی ImageNet

ImageNet						
Network	Method	Baseline (32 bit)	Accuracy			
			Bit-width			
			2	3	4	8
ResNet-18	LSQ [11] [‡]	70.1	66.7	69.4	70.7	
	LSQ+ [6]		66.8	69.4	70.8	
	FAQ [33]				69.8	70.02
	PACT [8]		64.4	68.1	69.2	
	NICE [5]			67.7	69.8	
	QIL [19]		65.7	69.2	70.1	70
	DSQ [21]		65.17	68.66	69.56	
	uL2Q [20]		65.5		66	65.4
ResNet-34	LSQ	74.1	71.6	73.4	74.1	74.1
	NICE			71.7	73.5	
	FAQ				73.3	73.7
	DSQ		70.02	72.54	72.76	
	QIL		70.6	73.1	73.7	
ResNet-50	LSQ	76.9	73.7	75.8	76.7	76.8
	PACT		72.2	75.3	76.5	
	NICE			75.1	76.5	
	FAQ				76.3	76.5

[‡] برای اینکه بتوانیم نتایج LSQ را با LSQ+ مقایسه کنیم، نتایج صحت هر دو مقاله را از LSQ+ گزارش کرده‌ایم.

نزدیک به صحت اصلی رسیده‌اند. برای بررسی دقیق‌تر، صحت روش‌های چندی‌سازی در بازه‌ی ۲ تا ۴ بیت بر روی شبکه‌ی ResNet-18، به‌عنوان نمونه، در شکل ۱ نشان داده شده‌است. این نمودار^۱ به خوبی پراکندگی صحت روش‌های مختلف چندی‌سازی را نشان می‌دهد.

در شکل ۱، نمودارهای مستطیلی از سمت چپ به راست، به ترتیب عرض بیت‌های ۲، ۳ و ۴ را نشان می‌دهند. در هر بیت، پراکندگی صحت‌ها با استفاده از یک مستطیل و دو خط عمودی مشخص شده است. علامت × داخل مستطیل، نشان‌دهنده‌ی میانه‌ی^۲ صحت‌هاست و مستطیل رنگی نشان‌دهنده‌ی ۵۰ درصد از صحت‌هاست که نزدیک به میانه قرار دارند. هم‌چنین دو خط کوتاه افقی در بالا و پایین هر نمودار، صحت بیشینه و کمینه^۳ را در آن دسته مشخص می‌کند.

طبق نمودار شکل ۱ ارتفاع مستطیل در دسته‌ی ۲، ۳ و ۴ بیت، به ترتیب ۱/۷، ۱/۴ و ۱/۱ درصد است. این موضوع نشان می‌دهد که روش‌های چندی‌سازی ممیزثابت نشان‌داده شده در این نمودار با اختلاف ۱/۱ تا ۱/۷ درصد، به صحت نزدیک به هم رسیده‌اند.

جدول ۱. مقایسه انواع چندی‌سازی ممیز ثابت

نام مقاله	سال	تعداد بیت	نمایش اعداد	ضریب مقیاس	آفست
Dadiannao [14]	2014	16	FxP	-	-
Limited Numerical Precision [28]	2015	16	FxP	-	-
low precision [9]	2015	10-12-20	FxP	-	-
reduced precision [30]	2015	8-12	FxP	-	-
Hardware-oriented approximation [16]	2016	8	FxP	-	-
Ristretto [15]	2018	8	FxP	-	-
QIL [19]	2019	2-5	FxP	-	-
LSQ [11]	2019	2-4	FxP	√	-
LSQ+ [6]	2020	2-4	FxP	√	√
uL2Q [20]	2019	1-8	FxP	√	√
DSQ [21]	2019	1-4	FxP	√	√
BitPruning [22]	2019	1-4	FxP	√	√
LLSQ [31]	2020	2-4	FxP	√	-
CSQ [29]	2022	2-4	FxP	√	√

¹ Box and Whiskers Charts

² Median

³ Maximum and Minimum

و آفست است.

چندی‌سازی ممیزثابت پایه را از نظر اجرای روی سخت‌افزار، شبیه به ممیزثابت با ضریب مقیاس در نظر می‌گیریم. چرا که در چندی‌سازی ممیزثابت با ضریب مقیاس، می‌توان ضریب مقیاس را از رابطه فاکتور گرفت. این موضوع در ادامه به تفصیل بیان خواهد شد. بنابر این توضیحات، در این پژوهش دو مدل معرفی می‌شود؛ چندی‌سازی همراه با ضریب مقیاس بدون آفست و چندی‌سازی همراه با ضریب مقیاس و آفست. برای اختصار، دسته‌ی اول را **ممیزثابت بدون آفست** و دسته‌ی دوم را **ممیزثابت با آفست** می‌نامیم.

۵-۱-۱- ممیز ثابت بدون آفست

در چندی‌سازی ممیزثابت بدون آفست، اعداد حقیقی به اعداد ممیزثابت همراه با ضریب مقیاس تبدیل می‌شوند. فرض کنیم \bar{X} و \bar{W} به ترتیب بردارهای ورودی و وزن باشند به صورتی که

$$\bar{X}, \bar{W} \in \mathbb{R}^K$$

و K مشخص‌کننده‌ی اندازه‌ی ضرب داخلی است. هم چنین فرض کنید اعداد حقیقی $x_k \in \bar{X}$ و $w_k \in \bar{W}$ به ترتیب نشان‌دهنده‌ی یک عدد از ورودی و وزن باشد که $k \in \{1, 2, \dots, K\}$. در چندی‌سازی ممیزثابت، این اعداد حقیقی به x_k^q و w_k^q تبدیل می‌شوند به طوری که

$$\begin{aligned} x_k^q &= Q(x_k) = \alpha_x^q d_{x_k} \\ w_k^q &= Q(w_k) = \alpha_w^q d_{w_k} \end{aligned} \quad (4)$$

در این رابطه α_x^q و α_w^q ضرایب مقیاس q بیتی هستند. براساس بررسی‌های انجام‌شده، انتخاب $q=8$ بیت برای ضریب مقیاس، شبکه را به دقت قابل قبولی می‌رساند، به همین دلیل ما در این پژوهش $q=8$ بیت را انتخاب کرده‌ایم. برای تمام اعداد ورودی و α_x^q برای تمام اعداد α_w^q برای تمام وزن‌ها در بردار K تایی یکسان در نظر گرفته شده است. d_{x_k} و d_{w_k} در این رابطه مقادیر چندی‌سازی شده را به ترتیب برای ورودی و وزن نشان می‌دهند. مجموعه‌ی F سیستم نمایش اعداد ممیزثابت است. در این سیستم، هر عدد از دو بخش تشکیل شده است، یک بخش عدد صحیح^۱ و یک بخش کسری^۲، این دو بخش توسط ممیزثابت از هم جدا می‌شوند.

همانطور که در بخش‌های قبل گفته شد، ضرب داخلی یکی از

تفاوت در سخت‌افزاری که هر روش به آن نیاز دارد، این سوال را ایجاد می‌کند که آیا بهره‌وری انرژی روش‌های مختلف چندی‌سازی ممیزثابت متفاوت است؟ اگر بله، کدام روش کمترین توان مصرفی و بیشترین کارایی انرژی را دارد؟ آیا روش‌های چندی‌سازی‌ای که صحت بالایی به دست می‌آورند، مثل نقاط بیشینه در شکل ۱ ([۱۱])، برای دستیابی به این صحت بالا توان مصرفی بسیار زیادی دارند؟ این سوالات و مشابه آنها ما را برآن داشت تا بهره‌وری انرژی روش‌های مختلف را بررسی کنیم. نتیجه‌ی چنین مقایسه‌ای می‌تواند به انتخاب یک چندی‌سازی کم‌مصرف و کارآمد از نظر انرژی منجر شود. البته می‌دانیم که مقایسه‌ی منصفانه‌ی روش‌های مختلف چندی‌سازی ممیزثابت، به علت سخت‌افزار متفاوت آنها، چالش برانگیز است اما تلاش می‌کنیم تا حد امکان شرایط عادلانه‌ای را برای مقایسه ایجاد کنیم.

۵- روش پیشنهادی

روش پیشنهادی برای انجام این پژوهش شامل دو بخش کلی است. ابتدا شیوه‌های مختلف چندی‌سازی در مدل‌های مشخص دسته‌بندی می‌شود و پس از آن برای هر مدل یک معماری ارائه می‌شود.

۵-۱- دسته‌بندی شیوه‌های مختلف چندی‌سازی به

مدل‌های پایه

چندی‌سازی کردن اعداد ورودی و وزن در شبکه‌های عصبی عمیق، به ناچار منجر به کاهش صحت این شبکه‌ها می‌شود. پژوهشگران مختلف این حوزه تلاش کرده‌اند تا با استفاده از راهکارهای مختلف خطای چندی‌سازی را کاهش دهند و صحت شبکه را به صحت اصلی نزدیک کنند. راهکارهای مختلفی جهت نگاشت بهتر بازه‌ی پارامترهای شبکه به سطوح چندی‌سازی در جهت افزایش دقت چندی‌سازی و به حداقل رساندن اختلاف بین مسیر انتشار به جلو و عقب ارایه شده است که در نتیجه بعضی از این روش‌ها سبب اضافه شدن ضریب مقیاس [۱۱، ۱۲، ۳۳] و آفست [۲۰، ۲۱] به نمایش ممیزثابت می‌شود.

با بررسی سیستم نمایش اعداد ممیزثابت در چندی‌سازی‌های مختلف به این نتیجه می‌رسیم که می‌توان این چندی‌سازی‌ها را در چند دسته‌ی کلی قرار داد. دسته‌ی اول که آن را دسته‌ی پایه می‌نامیم، هیچ‌گونه ضریب مقیاس یا آفستی ندارد. دسته‌ی دوم همراه با ضریب مقیاس است و دسته‌ی سوم همراه با ضریب مقیاس

² Fraction

¹ Integer

β_w^q و β_x^q برای تمام پارامترهای ضرب داخلی یکسان در نظر گرفته شده است.

در این چندی‌سازی، ضرب داخلی به صورت زیر انجام می‌شود؛

$$\begin{aligned} \text{dot}(\bar{X}^q, \bar{W}^q) &= \sum_{k=1}^K (x_k^q w_k^q) = \\ &= \sum_{k=1}^K ((\alpha_x^q d_{x_k} + \beta_x^q)(\alpha_w^q d_{w_k} + \beta_w^q)) \\ &= \sum_{k=1}^K ((\alpha_x^q \alpha_w^q)(d_{x_k} d_{w_k}) + (\alpha_x^q \beta_w^q d_{x_k}) \\ &\quad + (\alpha_w^q \beta_x^q d_{w_k}) + (\beta_x^q \beta_w^q)) \\ &= (\alpha_x^q \alpha_w^q) \sum_{k=1}^K (d_{x_k} d_{w_k}) \\ &\quad + (\alpha_x^q \beta_w^q) \sum_{k=1}^K d_{x_k} \\ &\quad + (\alpha_w^q \beta_x^q) \sum_{k=1}^K d_{w_k} \\ &\quad + \beta_x^q \beta_w^q \end{aligned} \quad (7)$$

همان‌طور که مشاهده می‌شود، حاصل نهایی ضرب داخلی در چندی‌سازی ممیزثابت با آفست ۴ بخش دارد؛ بخش‌های $(\alpha_w^q \beta_x^q) \sum_{k=1}^K d_{w_k}$ و $(\alpha_x^q \beta_w^q) \sum_{k=1}^K d_{x_k}$ ، $(\alpha_x^q \alpha_w^q) \sum_{k=1}^K (d_{x_k} d_{w_k})$ توسط سخت‌افزار و در زمان اجرا محاسبه می‌شوند. بخش $\beta_x^q \beta_w^q$ یک ثابت است که می‌تواند از قبل محاسبه شود و در نهایت با مقادیر بخش‌های قبل جمع شود.

۵-۲- معماری پیشنهادی مدل‌های پایه‌ی چندی‌سازی

در بخش قبل دو مدل پایه برای چندی‌سازی‌های ممیزثابت معرفی شد که شامل ممیزثابت بدون آفست و ممیزثابت با آفست بود. همان‌طور که پیش از این بیان شد، هدف از این پژوهش مقایسه‌ی بی‌طرفانه و عادلانه بین مدل‌های مختلف چندی‌سازی از منظر پیاده‌سازی سخت‌افزاری، از جمله توان مصرفی و مساحت است.

به منظور انجام این مقایسه، در این پژوهش برای هر مدل یک معماری پیشنهاد شده است. ممکن است برای هر کدام از این مدل‌ها معماری‌های دیگری در پژوهش‌های مختلف ارائه شده باشد که از لحاظ توان مصرفی، مساحت یا انرژی نسبت به این معماری پیشنهادی کارآمدتر باشند اما از آنجایی که معماری‌های ارائه شده متنوع و متفاوتند، مقایسه‌ی آنها بسیار سخت است و نمی‌توان معیارهای واحدی برای ارزیابی و مقایسه‌ی آنها در نظر گرفت. از طرفی بیشتر این تکنیک‌ها و معماری‌ها را می‌توان بر روی هر دو مدل چندی‌سازی ممیزثابت اعمال کرد. به همین دلیل، در این پژوهش پایه‌ترین معماری برای هر مدل در نظر گرفته شده است تا

بخش‌های اصلی و سنگین در شبکه‌های عصبی عمیق است و یکی از راه‌های کاهش حجم و انرژی مصرفی این عملیات، چندی‌سازی است. با استفاده از چندی‌سازی ممیزثابت بدون آفست، عملیات ضرب داخلی به صورت زیر انجام می‌شود؛

$$\begin{aligned} \text{dot}(\bar{X}^q, \bar{W}^q) &= \sum_{k=1}^K (x_k^q w_k^q) \\ &= \sum_{k=1}^K ((\alpha_x^q d_{x_k})(\alpha_w^q d_{w_k})) \\ &= \alpha_x^q \alpha_w^q \sum_{k=1}^K (d_{x_k} d_{w_k}) \end{aligned} \quad (5)$$

همان‌طور که گفتیم، ضرایب α_x^q و α_w^q برای تمام اعداد ورودی و وزن در ضرب داخلی یکسان در نظر گرفته شده است. بنابراین می‌توانیم $\alpha_x^q \alpha_w^q$ را از جمع فاکتور بگیریم. ذکر این نکته ضروری است که برای اینکه عملیات لایه‌ی کانولوشن را ساده‌تر کنیم، $\alpha_x^q \alpha_w^q$ را با بخش نرمال‌سازی ادغام می‌کنیم. در این صورت، بخش باقیمانده‌ی $\sum_{k=1}^K (d_{x_k} d_{w_k})$ را می‌توانیم با سخت‌افزاری ساده انجام دهیم که در بخش بعد به تفصیل توضیح داده خواهد شد. هم‌چنین با دقت به بخش $\sum_{k=1}^K (d_{x_k} d_{w_k})$ می‌توان دریافت که این رابطه، دقیقاً عملیات ضرب داخلی را برای چندی‌سازی بدون ضریب نشان می‌دهد، یعنی حالتی که اعداد ورودی و وزن فقط به مقادیر d_{w_k} و d_{x_k} چندی‌سازی می‌شوند. در این شرایط، می‌توانیم از معماری طراحی‌شده برای چندی‌سازی بدون آفست، برای انجام عملیات ضرب داخلی در چندی‌سازی بدون ضریب مقیاس نیز استفاده کنیم.

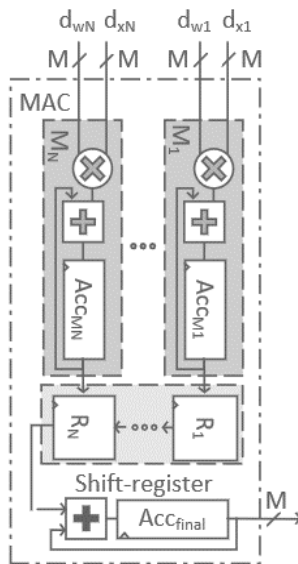
۵-۱-۲- ممیزثابت با آفست

در چندی‌سازی ممیزثابت با آفست، اعداد چندی‌سازی شده علاوه بر ضریب مقیاس یک آفست نیز دارند. اضافه کردن یک آفست با این هدف انجام می‌شود تا بازه‌ی قابل نمایش به پارامترهای شبکه منطبق‌تر شود و خطای چندی‌سازی کاهش یابد.

همانند بخش قبل، x_k و w_k را به ترتیب، یک عدد از ورودی و وزن در نظر می‌گیریم. در چندی‌سازی ممیزثابت با آفست داریم؛

$$\begin{aligned} x_k^q &= Q(x_k) = \alpha_x^q d_{x_k} + \beta_x^q \\ &\quad \alpha_x^q, d_{x_k}, \beta_x^q \in F \\ w_k^q &= Q(w_k) = \alpha_w^q d_{w_k} + \beta_w^q \\ &\quad \alpha_w^q, d_{w_k}, \beta_w^q \in F \end{aligned} \quad (6)$$

که در این رابطه، β_w^q و β_x^q آفست‌های چندی‌سازی هستند که ما ۸ بیتی در نظر گرفته‌ایم. در قسمت قبل دلیل انتخاب ۸ بیت برای ضرایب مقیاس و آفست آورده شده است. همانند ضرایب مقیاس،



شکل ۲. معماری هسته‌ی پردازشی ضرب داخلی در چندی‌سازی ممیز ثابت بدون آفست.

قبلاً گفته شد که تعداد کل ضرب داخلی را K در نظر گرفته‌ایم. در هر سیکل، N عدد ورودی d_{x1}, \dots, d_{xN} و عدد وزن d_{w1}, \dots, d_{wN} که هر کدام M بیتی‌اند وارد می‌شوند و دو به دو به M_i می‌روند. هر M_i یک ضرب‌کننده‌ی M بیتی برای ضرب اعداد ورودی دارد، پس از آن یک جمع‌کننده و ثابت^۴ وجود دارد که برای انباشت نتیجه استفاده می‌شود. پس از گذشت $\frac{K}{N}$ سیکل، تمام ضرب‌های بین بردار ورودی و وزن انجام می‌شود و از این پس باید حاصل ضرب‌های جزئی با هم جمع شود. در این مرحله، پس از رسیدن سیگنال کنترلی مبنی بر تمام شدن K ضرب ورودی در وزن، خروجی ثابت Acc_{Mi} هر کدام از M_i ها وارد ثابت R_i مربوط به خود می‌شود.

مجموعه‌ی R_N, \dots, R_1 در واقع یک ثابت انتقالی^۵ است که به همراه جمع‌کننده و ثابت بعد از خودش، بخش کاهش مدار را تشکیل می‌دهد. در هر سیکل، مقدار R_i وارد R_{i+1} می‌شود و مقدار آخرین ثابت، یعنی R_N که به جمع‌کننده متصل است، با مقدار ثابت Acc_{final} جمع و مجدداً انباشته می‌شود.

برای اینکه بتوان دو بخش ضرب و انباشت و کاهش را به صورت خط لوله پیش برد، سیگنال‌های کنترلی طوری تنظیم شده‌اند که تا پایان ضرب و انباشت تمام ورودی‌ها و وزن‌ها، مانع از ورود مقدار Acc_{Mi} به R_i می‌شوند. پس از آن یکبار تمام مقادیر Acc_{Mi} وارد R_i می‌شود و مجدداً راه ارتباطی بین این دو گروه ثابت مسدود می‌شود.

بتوان جزئیات را تا حد امکان شبیه به هم طراحی کرد و معماری‌ها را منصفانه مقایسه کرد.

برای اینکه جایگاه این پژوهش در شبکه‌های عصبی عمیق مشخص شود لازم به یادآوری است که اصلی‌ترین عملیات در این شبکه‌ها، عملیات کانولوشن است و همان‌طور که در بخش‌های قبل بیان شد (روابط (۱) و (۲))، کانولوشن را می‌توان به عنوان تعداد بسیار زیادی ضرب داخلی بین بردارهای ورودی و وزن در نظر گرفت. در بخش قبل، برای هر کدام از مدل‌های پایه‌ی چندی‌سازی سیستم نمایش اعداد و ضرب داخلی را نشان دادیم، در این بخش معماری مربوط به ضرب داخلی هر چندی‌سازی بر اساس نمایش اعداد بخش قبل معرفی می‌شود. در ادامه جزئیات هر معماری به تفصیل بیان خواهد شد.

۵-۲-۱- ممیز ثابت بدون آفست

با توجه به رابطه‌ی (۵) مشخص می‌شود ضرب داخلی در چندی‌سازی ممیز ثابت بدون آفست، شامل یک عملیات ضرب و انباشت و سپس ضرب در یک ضریب است. قبلاً گفته شد که برای ساده کردن ضرب داخلی و بخش کانولوشن، ضریب را با لایه‌ی بعد ادغام می‌کنیم بنابراین فقط ضرب و انباشت در این لایه باقی می‌ماند.

شکل ۲ معماری هسته‌ی پردازشی^۱ ضرب داخلی در چندی‌سازی ممیز ثابت بدون آفست را نشان می‌دهد. هسته‌ی اصلی این معماری، واحد ضرب و انباشت است که در شکل به صورت بخش‌های رنگی با شناسه‌ی M_i نشان داده شده که $i \in \{1, \dots, N\}$ است. N بیان‌گر تعداد کل ضرب و انباشت‌هایی است که در هر سیکل، به صورت موازی اجرا می‌شوند. در واقع N ضریب موازی‌سازی در سطح واحد پردازشی است.

به طور کلی، این معماری شامل دو بخش است که می‌توانند به صورت موازی با هم اجرا شوند و یک خط لوله^۲ را تشکیل دهند. در بخش اول که آن را ضرب و انباشت می‌نامیم و شامل N واحد MAC موازی است، اعداد ورودی و وزن دو به دو در هم ضرب و انباشت می‌شوند. در بخش دوم که بخش کاهش^۳ است، نتیجه‌ی ضرب‌های بخش اول با هم جمع می‌شود و حاصل نهایی محاسبه می‌شود.

⁴ Register

⁵ Shift Register

¹ Processing Engine (PE)

² Pipeline

³ Reduction

برای اینکه موضوع واضح شود، با یک مثال عددی پیش می‌رویم. فرض کنیم $K=1024$ ، در این صورت طبق رابطه‌ی (۸)، باید N را ۳۲ انتخاب کنیم. معماری این حالت در شکل ۳ (الف) نشان داده شده است. در این شرایط بخش ضرب و انباشت و کاهش هر دو ۳۲ سیکل طول می‌کشد و توازن در خط لوله حفظ می‌شود. اکنون موازی‌سازی را ۲ برابر و N را ۶۴ انتخاب می‌کنیم. معماری این حالت در شکل ۳ (ب) مشاهده می‌شود. در این حالت، بخش ضرب و انباشت ۱۶ سیکل طول می‌کشد و اگر فقط از یک جمع‌کننده در بخش کاهش استفاده کنیم، این بخش پس از ۶۴ سیکل تمام خواهد شد.

از آنجاییکه بخش ضرب و انباشت بسیار سریع تمام می‌شود، تا پایان کار بخش کاهش باید منتظر بماند، این بیکاری باعث کاهش کارایی معماری می‌شود. برای اینکه توازن خط لوله را حفظ کنیم، لازم است یک جمع‌کننده دیگر به بخش کاهش اضافه کنیم تا در هر سیکل ۴ مقدار ثابت با هم جمع شوند و تمام این بخش در ۱۶ سیکل تمام شود. علاوه بر این، به یک جمع‌کننده دیگر نیاز داریم تا در یک ساختار درختی، مقدار دو جمع قبل را محاسبه کند. هم‌چنین یک جمع‌کننده دیگر نیاز است تا به کمک یک ثابت، نتایج را انباشته کند.

در شرایطی که $N=64$ باشد، نسبت به حالت قبل، دو برابر ضرب داخلی بیشتری در زمان مشابه انجام شده است (توان عملیاتی ۲ برابر شده است)، اما با مقایسه‌ی قسمت (الف) و (ب) در شکل ۳ به این نتیجه می‌رسیم که توان مصرفی از دو برابر بیشتر شده است. (تعداد واحدهای ضرب و انباشت و ثابت‌های انتقالی دو برابر شده است اما علاوه بر آن معماری شکل (ب)، نسبت به شکل (الف) سه جمع‌کننده بیشتر دارد و همین موضوع باعث می‌شود توان مصرفی از دو برابر بیشتر شود.)

در این پژوهش بهره‌وری انرژی، توان عملیاتی نسبت به توان مصرفی تعریف شده است و توان عملیاتی تعداد ضرب داخلی در واحد زمان را نشان می‌دهد. در این مثال، توان عملیاتی دقیقاً دو برابر شده است اما توان مصرفی از دو برابر بیشتر است، بنابراین بهره‌وری انرژی نسبت به حالت قبل کاهش می‌یابد. این موضوع که از نظر تئوری بررسی شد، در فصل نتایج، با استفاده از نتایج شبیه‌سازی نیز نشان داده خواهد شد.

در این مرحله بخش کاهش مدار نتایج ضرب و انباشت‌ها را با هم جمع می‌کند و بخش ضرب و انباشت می‌تواند ورودی‌های دیگری را از یک عملیات ضرب داخلی جدید دریافت و آنها را پردازش کند.

مرحله‌ی کاهش، دارای N ثابت است که در هر سیکل مقدار هر یک از آنها به ثابت بعدی منتقل و در نهایت مجموع مقدار ثابت‌ها محاسبه می‌شود. بنابراین این بخش N سیکل طول می‌کشد. گفتیم که بخش ضرب و انباشت کار خود را در $\frac{K}{N}$ سیکل انجام می‌دهد. برای اینکه توازن این دو بخش حفظ شود و خط لوله بالاترین کارایی را داشته باشد، لازم است N را طوری انتخاب کنیم که

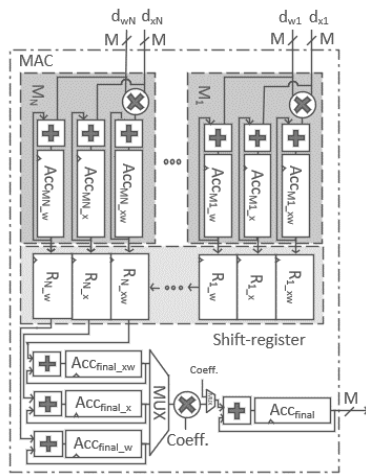
$$\frac{K}{N} = N \rightarrow N_{opt} = \sqrt{K} \quad (8)$$

بر اساس روابط تئوری مشاهده می‌شود که برای ضریب موازی‌سازی یک مقدار بهینه به ازای هر K وجود دارد. اگر $N < N_{opt}$ باشد، آنگاه مدار به بالاترین حد موازی‌سازی خودش نمی‌رسد، تعداد سیکل بیشتری برای محاسبه‌ی ضرب داخلی نیاز دارد و در نتیجه، بهره‌وری انرژی از مقدار بیشینه‌ی خود کمتر می‌شود.

اگر $N > N_{opt}$ باشد، یعنی تعداد واحدهای MAC در بخش ضرب و انباشت بیشتر می‌شود و کل عملیات ضرب و انباشت در تعداد سیکل کمتری انجام می‌شود. از طرف دیگر، تعداد ثابت‌های رجیستر انتقالی R_i هم بیشتر می‌شود، لذا به تعداد سیکل بیشتری برای جمع کردن نتایج نیاز دارد. در این شرایط، بخش ضرب و انباشت مدار سریع‌تر از بخش کاهش به اتمام می‌رسد و تا زمان اتمام کار بخش کاهش بیکار می‌ماند. ایجاد شدن این حالت‌های بیکاری در خط لوله، کارایی آن را پایین می‌آورد و در نتیجه بهره‌وری انرژی را کاهش می‌دهد.

اکنون اگر بخواهیم موازی‌سازی را از N_{opt} بالاتر ببریم و هم‌چنان تعادل خط لوله را حفظ کنیم، نیاز داریم که در بخش کاهش مدار، جمع‌کننده‌های جدیدی را اضافه کنیم تا حاصل ثابت‌ها را به صورت موازی با هم و سریع‌تر جمع کنند. علاوه بر آن، به جمع‌کننده‌های دیگری نیز نیاز داریم تا در یک ساختار درختی، تمام نتایج را با هم جمع کنند و حاصل نهایی را محاسبه کنند. پس با بالا بردن موازی‌سازی از حد بهینه‌ی آن، گرچه می‌توانیم تعداد ضرب داخلی را در زمان مشابه بالاتر ببریم (توان عملیاتی^۱ را افزایش دهیم) اما سرباری که جمع‌کننده‌های جدید به مدار تحمیل می‌کنند، توان مصرفی را بیش از توان عملیاتی افزایش می‌دهد و در نتیجه بهره‌وری انرژی کاهش می‌یابد.

^۱ Throughput

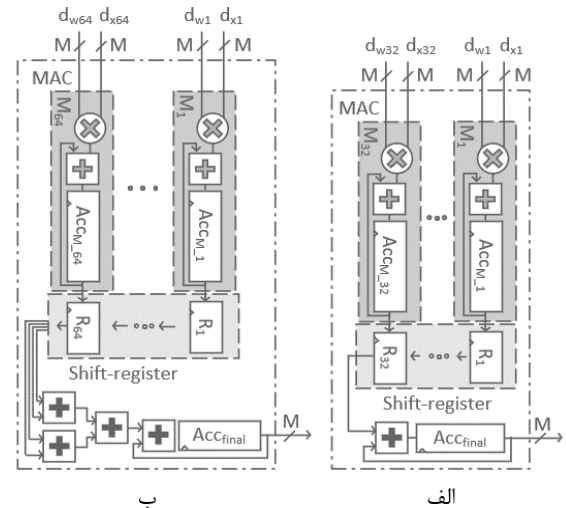


شکل ۴. معماری هسته‌ی پردازشی ضرب داخلی در چندی‌سازی ممیز ثابت با آفست

در هر سیکل، d_{wN}, \dots, d_{w1} و d_{xN}, \dots, d_{x1} وارد می‌شوند و دو به دو به M_i می‌روند. در هر M_i یک ضرب‌کننده وجود دارد که حاصل $(d_{x_i} d_{w_i})$ را محاسبه و در ثبات $Acc_{M_i, xw}$ انباشته می‌کند. علاوه بر این، دو ثبات $Acc_{M_i, w}$ و $Acc_{M_i, x}$ وجود دارند که به ترتیب مقدار d_{w_i} و d_{x_i} را انباشته می‌کنند. بخش ضرب‌وانباشت، پس از گذشت $\frac{K}{N}$ سیکل، تمام ضرب‌وانباشت‌های بردار ورودی و وزن را انجام می‌دهد. در این مرحله مقدار این ثبات‌ها وارد ثبات‌های $R_{i, x}$ ، $R_{i, xw}$ و $R_{i, w}$ می‌شود. مجموعه‌ی این ثبات‌ها یک ثبات انتقالی را تشکیل می‌دهد که به N دسته‌ی سه‌تایی تقسیم شده‌اند. هر کدام از دسته‌ها مربوط به یک M_i است.

پس از اتمام ضرب‌وانباشت و ورود مقادیر M_i به ثبات انتقالی، بخش کاهش مدار شروع می‌شود. در این مرحله، در هر سیکل مقدار $R_{i, xw}$ به $R_{i+1, xw}$ ، مقدار $R_{i, x}$ به $R_{i+1, x}$ و مقدار $R_{i, w}$ به $R_{i+1, w}$ منتقل می‌شود. آخرین مجموعه‌ی سه‌تایی از ثبات انتقالی به سه انباشت‌گر (شامل جمع‌کننده و ثبات) متصل است که مقادیر ثبات‌های مربوط به خود را انباشته می‌کنند. به عبارت دیگر، پس از گذشت N سیکل، مقدار $(\sum_{k=1}^K d_{x_k} d_{w_k})$ ، $\sum_{k=1}^K d_{x_k}$ و $\sum_{k=1}^K d_{w_k}$ به ترتیب در ثبات‌های $Acc_{final, xw}$ ، $Acc_{final, x}$ و $Acc_{final, w}$ ذخیره شده است.

در این مرحله باید مقادیر ثبات‌های نهایی در ضریب مربوط ضرب شوند. این کار در چهار سیکل انجام می‌شود، در سه سیکل اول، به ترتیب هر کدام از مقادیر ثبات‌های $Acc_{final, xw}$ ، $Acc_{final, x}$ و $Acc_{final, w}$ انتخاب شده و در ضریب‌های $(\alpha_x^q \alpha_w^q)$ و $(\alpha_w^q \beta_x^q)$ ضرب شده و در ثبات Acc_{final} انباشته می‌شوند. در سیکل آخر نیز، مقدار $\beta_x^q \beta_w^q$ با مقادیر قبل جمع شده و حاصل نهایی محاسبه می‌شود. لازم به ذکر است در صورتی که مقدار K به اندازه‌ی



شکل ۳. ضرب داخلی در ممیز ثابت بدون آفست با فرض $K=1024$. (الف) $N=N_{opt}=32$ و (ب) $N=2N_{opt}=64$.

معماری نشان داده شده در شکل ۲ مقدار عبارت $\sum_{k=1}^K (d_{x_k} d_{w_k})$ را محاسبه می‌کند. پس از این مرحله می‌توان ضریب $\alpha_x^q \alpha_w^q$ را در لایه نرمال‌سازی اعمال کرد تا حاصل ضرب داخلی به طور کامل و صحیح محاسبه شود. هم‌چنین می‌توان تنها همین معماری را برای ضرب داخلی چندی‌سازی بدون ضریب استفاده کرد.

۵-۲-۲- ممیز ثابت با آفست

شکل ۴ معماری هسته‌ی پردازشی ضرب داخلی را در چندی‌سازی ممیز ثابت با آفست نشان می‌دهد. همانند معماری ممیز ثابت بدون آفست، این معماری نیز دو بخش کلی دارد؛ بخش ضرب‌وانباشت و بخش کاهش. این دو بخش می‌توانند موازی با هم اجرا شوند و خط لوله را تشکیل دهند.

همان‌طور که در رابطه‌ی (۷) دیدیم، حاصل ضرب داخلی در چندی‌سازی ممیز ثابت با آفست چهار بخش دارد که سه بخش آن در زمان اجرا و یک بخش آن از قبل محاسبه می‌شود. هر کدام از سه بخشی که باید در زمان اجرا محاسبه شوند، یعنی $(\alpha_x^q \alpha_w^q) \sum_{k=1}^K d_{w_k}$ ، $(\alpha_x^q \beta_w^q) \sum_{k=1}^K d_{x_k}$ ، $(\alpha_x^q \alpha_w^q) \sum_{k=1}^K (d_{x_k} d_{w_k})$ شامل یک ضرب‌وانباشت و یک ضریب است. بر همین اساس، معماری طوری طراحی شده‌است که ابتدا حاصل ضرب‌وانباشت‌ها محاسبه شود، سپس ضریب مربوط در هر کدام ضرب شود.

قسمت اصلی در این معماری، واحدهای ضرب‌وانباشت است که با اندیس‌های M_i در شکل ۴ مشخص است که $i \in \{1, \dots, N\}$ و همانند معماری ممیز ثابت بدون آفست، در اینجا نیز N ، ضریب موازی‌سازی و نشان‌دهنده‌ی تعداد واحدهای ضرب‌وانباشتی است که به صورت موازی با هم اجرا می‌شوند.

رسیده‌اند (جدول ۲) در شبیه‌سازی‌ها، ۳ بیت را به‌عنوان نمونه انتخاب کرده‌ایم.

پارامتر K ، بسیار به مشخصات شبکه‌های عصبی عمیق مختلف و ابعاد لایه‌های کانولوشن آنها بستگی دارد. به‌منظور انتخاب مقادیر مناسب، شبکه‌های عصبی عمیق مختلفی از جمله ResNet-18 بررسی شده‌اند و در نهایت برای K اعدادی در نظر گرفته شده است که به ابعاد واقعی شبکه‌های عصبی بررسی شده نزدیک باشد.

۶-۳- نتایج شبیه‌سازی و سنتز

همان‌طور که از ابتدای این مقاله گفته شد، هدف از انجام این پژوهش مقایسه‌ی روش‌های مختلف چندی‌سازی ممیزثابت شبکه‌های عصبی از لحاظ پارامترهای سخت‌افزاری از جمله توان مصرفی و بهره‌وری انرژی بوده است. در این بخش با استفاده از نتایج شبیه‌سازی‌ها و سنتز سخت‌افزار به این مقایسه می‌پردازیم.

این بخش شامل ۴ گروه نتایج است؛ یافتن بهترین ضریب موازی‌سازی، بهره‌وری انرژی معماری‌های پیشنهادی، بهره‌وری انرژی روش‌های چندی‌سازی مدرن^۱ و تأثیر افزودن ضریب و آفست بر سخت‌افزار.

۶-۳-۱- یافتن بهترین ضریب موازی‌سازی

همان‌طور که در بخش ۵-۲- معماری پیشنهادی مدل‌های پایه‌ی چندی‌سازی گفته شد، در معماری‌های ممیزثابت بدون آفست و با آفست، N ضریب موازی‌سازی در سطح واحد پردازشی است. همچنین در رابطه‌ی (۸) دیدیم که با توجه به سایر پارامترها، از جمله K ، برای ضرایب موازی‌سازی یک مقدار بهینه وجود دارد. از نظر تئوری بررسی شد که اگر ضرایب موازی‌سازی بیشتر یا کمتر از مقدار بهینه انتخاب شوند، بهره‌وری انرژی کاهش می‌یابد. در این بخش این موضوع را با استفاده از نتایج سنتز معماری بررسی خواهیم کرد.

شکل ۵ بهره‌وری انرژی معماری‌های ممیزثابت بدون آفست و با آفست را در ضرایب موازی‌سازی مختلف نشان می‌دهد. محور عمودی در این نمودار بهره‌وری انرژی است. در این شبیه‌سازی $K=9 \times 256$ و $M=3$ در نظر گرفته شده است. بنابراین، طبق رابطه‌ی (۸)، مقدار ضریب موازی‌سازی بهینه ۴۸ به‌دست می‌آید. همان‌طور که در شکل مشخص است، هر دو معماری ممیزثابت بدون آفست و با آفست بیشترین بهره‌وری خود را در $N=48$ به دست آورده‌اند.

کافی بزرگ انتخاب شود که $\frac{K}{N} \gg 4$ باشد، می‌توان از این چهار سیکل که به بخش کاهش اضافه شده است صرف‌نظر کرد و خط لوله را هم‌چنان متوازن در نظر گرفت.

همانند معماری ممیزثابت بدون آفست، در معماری ممیزثابت با آفست نیز دو بخش ضرب و انباشت و کاهش می‌توانند یک خط لوله را تشکیل دهند. همچنین برای اینکه تعادل این خط لوله حفظ شود، لازم است N طبق رابطه‌ی (۸) محاسبه شود. در این معماری نیز، اگر N کمتر یا بیشتر از N_{opt} قرار داده شود، مشابه، دلایلی که در قسمت قبل گفته شد بهره‌وری انرژی کاهش می‌یابد. این موضوع نیز در بخش نتایج نشان داده خواهد شد.

۶- نتایج

در این بخش، نتایج شبیه‌سازی برای معماری‌های پیشنهادی ارائه و بررسی می‌شوند. پیش از ارائه‌ی نتایج به معرفی ابزارهای استفاده‌شده می‌پردازیم. در این مقاله بهره‌وری انرژی، توان عملیاتی نسبت به توان مصرفی تعریف شده است و توان عملیاتی تعداد ضرب داخلی در واحد زمان را نشان می‌دهد.

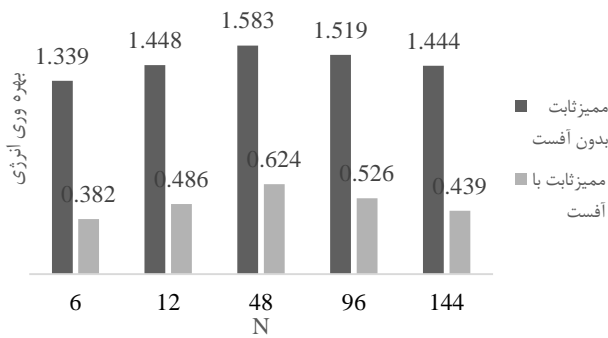
۶-۱- ابزارهای استفاده‌شده

در انجام این پژوهش برای مقایسه‌ی سخت‌افزاری مدل‌های مختلف چندی‌سازی، معماری‌های مناسب هر مدل طراحی، پیاده‌سازی و شبیه‌سازی شده‌اند. پیاده‌سازی معماری‌ها با استفاده از زبان سخت‌افزاری Verilog بوده و با کمک ابزار Modelsim شبیه‌سازی شده‌اند تا از صحت عملکرد آنها مطمئن شویم. پس از آن توان مصرفی و مساحت معماری‌ها با استفاده از ابزار Design Compiler و با تکنولوژی 45nm محاسبه شده است.

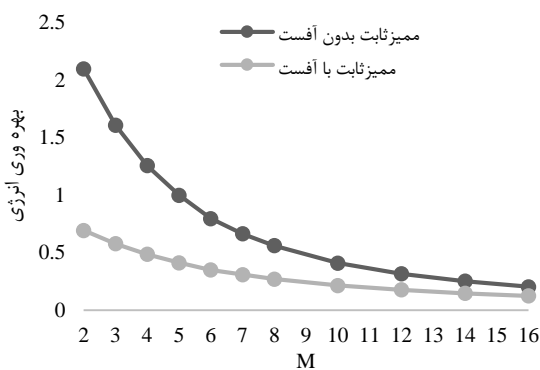
۶-۲- مقدار انتخاب‌شده برای پارامترها

همان‌طور که در بخش روش پیشنهادی توضیح داده شد، در معماری‌های طراحی‌شده‌ی این مقاله، سه پارامتر K ، N و M نقش مهمی دارد که به ترتیب نشان‌دهنده‌ی اندازه‌ی ضرب داخلی، ضریب موازی‌سازی در معماری و عرض بیت چندی‌سازی است. در شبیه‌سازی‌هایی که در ادامه نتایج آنها را بررسی می‌کنیم، برای M ، عموماً بازه‌ی ۱ تا ۱۶ بیت را در نظر گرفته‌ایم چرا که اغلب چندی‌سازی‌های ممیز ثابت در این عرض بیتی ارائه شده‌اند. همچنین از آنجایی که روش‌های اخیر چندی‌سازی ممیز ثابت در ۲ تا ۴ بیت به صحت‌هایی نزدیک به صحت شبکه‌ی ۳۲ بیتی

^۱ State of the Art



شکل ۵. بهره‌وری انرژی معماری ممیز ثابت بدون آفست و با آفست در ضرایب موازی سازی مختلف. $N_{opt}=48$ و $K=9 \times 256$.



شکل ۶. بهره‌وری انرژی برای معماری‌های پیشنهادی، $K=9 \times 512$ بر حسب تعداد بیت‌های مختلف برای چندی سازی.

با توجه به شکل ۶ مشخص می‌شود بهره‌وری انرژی در معماری بدون آفست، با شیب بسیار بیشتری نسبت به معماری با آفست کاهش می‌یابد. به طوری که در ۲ بیت، بهره‌وری انرژی معماری بدون آفست، حدوداً ۳ برابر بهره‌وری انرژی معماری با آفست است اما در ۱۶ بیت، تقریباً هر دو معماری به یک مقدار رسیده‌اند. دلیل این موضوع این است که با افزایش تعداد بیت، مدار ضرب کننده به صورت توان ۲ و جمع کننده به صورت خطی پیچیده می‌شود.

در معماری بدون آفست بین یک ضرب کننده و یک جمع کننده هزینه‌فایده^۱ وجود دارد اما در معماری با آفست هزینه‌فایده بین یک ضرب کننده و ۳ جمع کننده است. لذا معماری بدون آفست شیب بیشتری دارد.

۳-۳-۶- بهره‌وری انرژی روش‌های چندی سازی مدرن

پیش از این گفته شد که هدف از انجام این پژوهش، مقایسه‌ی بی‌طرفانه‌ی مدل‌های مختلف چندی سازی ممیز ثابت از نظر ویژگی‌های سخت‌افزاری به خصوص بهره‌وری انرژی است. در بخش

به طور کلی بهره‌وری انرژی در چندی سازی ممیز ثابت بدون آفست بیشتر از ممیز ثابت با آفست است. دلیل این موضوع این است که معماری ممیز ثابت با آفست نسبت به ممیز ثابت بدون آفست پیچیده‌تر است. این موضوع از مقایسه‌ی شکل ۲ و شکل ۴ نیز به دست می‌آید. لذا برای انجام یک ضرب داخلی مشابه (وقتی K یکسان باشد) معماری ممیز ثابت با آفست به توان مصرفی بیشتری نیاز دارد. همان طور که در شکل ۵ مشخص است، اگر مقدار N را از نقطه‌ی بهینه کمتر انتخاب کنیم (ما در این شبیه‌سازی مقادیر ۶ و ۱۲ را برای مثال در نظر گرفتیم) موازی سازی به حداکثر خود نمی‌رسد، بنابراین مدت زمان محاسبه‌ی ضرب داخلی بیشتر شده (توان عملیاتی کم می‌شود) و بهره‌وری انرژی کاهش می‌یابد. اگر N را بیشتر از نقطه‌ی بهینه قرار دهیم (برای مثال ما ۹۶ و ۱۴۴ را در نظر گرفتیم)، بنابر دلایلی که در فصل قبل گفته شد، بهره‌وری انرژی از مقدار بیشینه‌ی خود کمتر می‌شود.

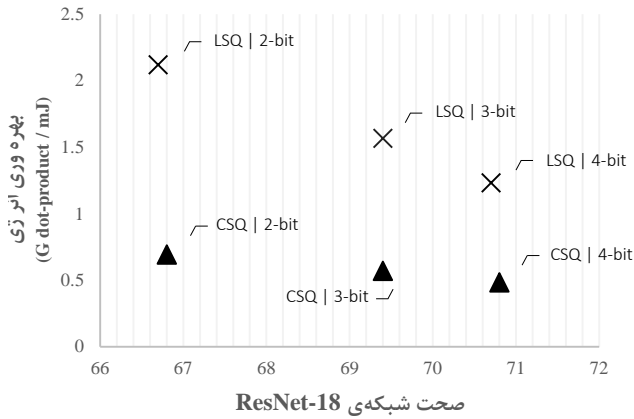
۳-۳-۶- بهره‌وری انرژی معماری‌های پیشنهادی به ازای عرض بیت‌های مختلف

شکل ۶ بهره‌وری انرژی را برای دو معماری ممیز ثابت بدون آفست و ممیز ثابت با آفست نشان می‌دهد. نتایج این نمودار برای $K=9 \times 512$ و $M = \{2, \dots, 16\}$ به دست آمده است.

بر اساس نتایج شکل ۶ مشخص می‌شود که معماری ممیز ثابت بدون آفست، بهره‌وری انرژی بیشتری نسبت به ممیز ثابت با آفست دارد. این موضوع را قبلاً در شکل ۵ نیز دیده بودیم. دلیل این موضوع نیز این است که معماری ممیز ثابت با آفست نسبت به ممیز ثابت بدون آفست پیچیده‌تر است و در توان عملیاتی یکسان، توان مصرفی بیشتری دارد لذا بهره‌وری انرژی آن در تمام عرض بیت‌ها کمتر است.

نکته‌ی دیگری که در شکل ۶ می‌توان مشاهده کرد، این است که در هر دو معماری بدون آفست و با آفست، با افزایش تعداد بیت چندی سازی از ۲ تا ۱۶ بیت، به دلیل زیاد شدن توان مصرفی معماری، بهره‌وری انرژی کاهش می‌یابد چرا که توان عملیاتی تغییر نکرده است. این موضوع می‌تواند در انتخاب عرض بیت چندی سازی نقش مهمی ایفا کند.

^۱ Trade-off



شکل ۷. بهره‌وری انرژی روش‌های مدرن چندی‌سازی در تعداد بیت مختلف. شکل‌های \times و Δ به ترتیب مدل‌های ممیزثابت بدون آفست و ممیزثابت با آفست را نشان می‌دهند. عرض بیت هر مدل به همراه نام آن روی نمودار مشخص است.

نکته‌ی دیگر نمودار شکل ۷ این است که در تمام عرض بیت‌های ۲، ۳ و ۴، مدل ممیزثابت با آفست، به صحت برابر یا بیشتری از ممیزثابت بدون آفست رسیده است. اما، بهره‌وری انرژی آن بسیار کمتر است. به بیان دیگر، بین صحت شبکه و بهره‌وری انرژی در روش‌های مختلف چندی‌سازی ممیزثابت یک هزینه‌فایده‌ی مهم وجود دارد. در مثال شکل ۷ حداکثر افزایش صحت شبکه ۰/۲ درصد است (در ۴ بیت، CSQ به ۰/۲ صحت بیشتری دست یافته است) اما بهره‌وری انرژی روش ممیزثابت با آفست، بسیار کمتر از روش بدون آفست است (به ترتیب ۲/۵، ۲/۷ و ۳/۱ برابر کمتر در ۲، ۳ و ۴ بیت). هزینه‌فایده‌ی میان صحت شبکه و معیارهای سخت‌افزاری مثل بهره‌وری انرژی، در حوزه‌ی چندی‌سازی شبکه‌های عصبی، به خصوص برای سیستم‌های نهفته موضوع بسیار مهمی است که تا کنون کسی به آن توجه نکرده است.

۶-۳-۴- تأثیر افزودن ضریب و آفست بر توان مصرفی و مساحت

در بخش انگیزه مقاله عنوان کردیم که یکی از سوالاتی که منجر به انجام این پژوهش شده است، این است که راهکارهای کاهش خطای چندی‌سازی و افزایش صحت شبکه، نظیر افزودن ضریب و آفست، چه تأثیری بر سخت‌افزار دارد. در این بخش قصد داریم این موضوع را بررسی کنیم.

شکل ۸ توان مصرفی و مساحت را برای معماری‌های ممیزثابت بدون آفست و ممیزثابت با آفست نشان می‌دهد که به نتایج ممیزثابت بدون آفست نرمال شده‌اند. نتایج برای حالتی محاسبه شده‌اند که $M=3$ و $N=24$ ، $K=9 \times 64$ است.

قبل بهره‌وری انرژی را برای معماری‌های پیشنهادی به صورت مستقل بررسی کردیم. در این بخش این موضوع را بررسی می‌کنیم که روش‌های چندی‌سازی موجود چه بهره‌وری انرژی خواهند داشت. برای این کار، از میان روش‌های چندی‌سازی مطرح که در بخش کارهای پیشین مرور شدند، در هر کدام از مدل‌های ممیزثابت بدون آفست و ممیزثابت با آفست، مدلی را انتخاب کردیم که به بالاترین صحت روی شبکه عصبی ResNet-18 رسیده است. با این توضیحات، در گروه ممیزثابت بدون آفست و با آفست، به ترتیب روش‌های LSQ و CSQ انتخاب شده‌اند.

پس از آن بهره‌وری انرژی را برای هر روش چندی‌سازی، با توجه به مدل (بدون آفست یا با آفست) و عرض بیت آن (۲، ۳ یا ۴) محاسبه کردیم. ذکر این نکته ضروری است که ممکن است برای هر کدام از روش‌های چندی‌سازی گفته شده، شتاب‌دهنده‌های سخت‌افزاری دیگری وجود داشته باشد که از لحاظ انرژی مصرفی نسبت به مدل پیشنهادی این پژوهش کارآمدتر بوده و نتایج بهتری به دست آورده باشند اما از آنجا که شتاب‌دهنده‌های موجود بسیار با هم متفاوت‌اند و عملاً امکان مقایسه‌ای منصفانه میان آنها وجود ندارد و همچنین بیشتر تکنیک‌ها و معماری‌های ارائه شده را می‌توان به هر دو مدل چندی‌سازی اعمال کرد، بهره‌وری انرژی روش‌های مختلف را با استفاده از معماری‌های پیشنهادی این پژوهش محاسبه کرده‌ایم تا بستری فراهم کنیم که تمام روش‌های چندی‌سازی ممیزثابت در شرایط یکسان ارزیابی شوند.

شکل ۷ نتایج بهره‌وری انرژی را برای روش‌های چندی‌سازی مختلف نشان می‌دهد. محور افقی در این نمودار صحت شبکه‌ی عصبی ResNet-18 است که با مجموعه‌داده‌ی ImageNet آموزش دیده و روش‌های مختلف چندی‌سازی ممیزثابت روی آن اعمال شده است. این نتایج برای عرض بیت ۲، ۳ و ۴ محاسبه شده‌اند.

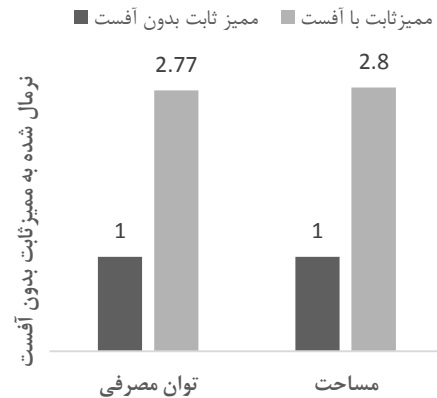
همان‌طور که در نتایج بخش بهره‌وری انرژی معماری‌های پیشنهادی دیدیم، هرچه عرض بیت در چندی‌سازی بیشتر شود، از یک طرف صحت شبکه بالاتر می‌رود و از طرف دیگر بهره‌وری انرژی کاهش می‌یابد. این موضوع را در شکل ۷ نیز می‌توان مشاهده کرد. برای مثال روش LSQ در ۲ بیت به دقت ۶۶/۷ می‌رسد و بهره‌وری انرژی آن ۲/۱۲ است. با افزایش تعداد بیت به ۴ بیت، صحت شبکه تا ۷۰/۷ افزایش می‌یابد (۳ درصد افزایش) اما از طرف دیگر، بهره‌وری انرژی به ۱/۲۳ کاهش می‌یابد (۱/۷ برابر کمتر).

ممیزثابت با آفست، همواره بهره‌وری انرژی کمتری نسبت به ممیزثابت بدون آفست در بیت یکسان دارد که این موضوع ناشی از معماری پیچیده‌تر آن است. در آخر، تأثیر افزودن ضریب و آفست را بر سخت‌افزار بررسی کردیم که مشخص شد افزودن ضریب و آفست به چندی‌سازی ممیزثابت، باعث افزایش توان مصرفی و مساحت تا حدود ۳ برابر می‌شود.

یکی از نتایج مهمی که در بررسی‌ها به‌دست آمده و از ابتدا انگیزه‌ی انجام این پژوهش بوده است، هزینه‌فایده‌ی میان صحت شبکه‌های عصبی چندی‌سازی‌شده و سربارهای سخت‌افزاری هم‌چون بهره‌وری انرژی است. براساس نتایج شبیه‌سازی، مشخص شد که روش‌های ممیزثابت با آفست، می‌توانند نسبت به روش‌های بدون آفست، به صحت‌های بالاتری (در حدود ۰/۱) در شبکه‌های عصبی دست پیدا کنند اما از طرف دیگر، بهره‌وری انرژی آنها بسیار کاهش پیدا می‌کند (به‌طور میانگین ۲/۸ برابر کمتر) این موضوع در حوزه‌ی سیستم‌های نهفته که با محدودیت توان مصرفی و انرژی روبرو هستند، نقش بسیار مهمی دارد.

مراجع

- [1] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and review," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295-2329, 2017.
- [2] Y. He, P. Liu, Z. Wang, Z. Hu, and Y. Yang, "Filter pruning via geometric median for deep convolutional neural networks acceleration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4340-4349.
- [3] Z. Zhuang et al., "Discrimination-aware channel pruning for deep neural networks," *Advances in neural information processing systems*, vol. 31, 2018.
- [4] Z. Liu, M. Sun, T. Zhou, G. Huang, and T. Darrell, "Rethinking the value of network pruning," *arXiv preprint arXiv:1810.05270*, 2018.
- [5] C. Baskin et al., "Nice: Noise injection and clamping estimation for neural network quantization," *Mathematics*, vol. 9, no. 17, p. 2144, 2021.
- [6] Y. Bhalgat, J. Lee, M. Nagel, T. Blankevoort, and N. Kwak, "Lsq+: Improving low-bit quantization through learnable offsets and better initialization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 696-697.
- [7] S.-E. Chang et al., "RMSMP: A Novel Deep Neural Network Quantization Framework with Row-wise Mixed Schemes and Multiple Precisions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5251-5260.
- [8] J. Choi, Z. Wang, S. Venkataramani, P. I.-J. Chuang, V. Srinivasan, and K. Gopalakrishnan, "Pact: Parameterized clipping activation for quantized neural networks," *arXiv preprint arXiv:1805.06085*, 2018.
- [9] M. Courbariaux, Y. Bengio, and J.-P. David, "Training deep neural networks with low precision multiplications," *arXiv preprint arXiv:1412.7024*, 2014.
- [10] Z. Dong, Z. Yao, A. Gholami, M. W. Mahoney, and K. Keutzer, "Hawq: Hessian aware quantization of neural networks with mixed-



شکل ۸. تأثیر افزودن ضریب و آفست بر توان مصرفی و مساحت سخت‌افزار. نتایج به ممیزثابت بدون آفست نرمال شده‌اند.

در نتایج بخش‌های قبل دیدیم وقتی که روش‌های چندی‌سازی از آفست استفاده می‌کنند، با کاهش خطای چندی‌سازی، صحت شبکه را افزایش می‌دهند (شکل ۷). از طرف دیگر، بررسی نتایج مختلف (شکل ۵ و شکل ۶) نشان داد که بهره‌وری انرژی برای معماری ممیزثابت با آفست، کمتر از ممیزثابت بدون آفست است. در توضیح دلیل این اتفاق، گفتیم که معماری ممیزثابت با آفست، نسبت به بدون آفست، پیچیدگی بیشتری دارد و در توان عملیاتی یکسان، توان مصرفی بیشتری دارد. این موضوع را که پیش از این به‌صورت تئوری بیان کردیم، می‌توان در شکل ۸، براساس نتایج شبیه‌سازی مشاهده کرد. براساس این نتایج، مشخص است که افزودن ضریب و آفست، توان مصرفی و مساحت را به ترتیب ۲/۷۷ و ۲/۸ برابر افزایش می‌دهد.

۷- نتیجه‌گیری

روش‌های مختلف چندی‌سازی سیستم نمایش اعداد متفاوتی دارند و عملیات مختلف را به شیوه‌های مختلف انجام می‌دهند، به همین دلیل به نظر می‌رسد اجرای هر کدام از روش‌ها به سخت‌افزار متفاوتی نیاز داشته باشد. به همین منظور، در این مقاله، روش‌های مختلف چندی‌سازی ممیزثابت در مدل‌های پایه دسته‌بندی شدند که شامل ممیزثابت بدون آفست و ممیزثابت با آفست بود و برای هر مدل یک معماری پیشنهاد داده شد تا بتوان به صورت عادلانه آنها را مقایسه کرد و در نهایت معماری‌های پیشنهادی ارزیابی شدند.

براساس نتایج ارزیابی‌ها، مشخص شد که هر کدام از معماری‌های پیشنهادی برای چندی‌سازی ممیزثابت بدون آفست و با آفست، یک ضریب موازی‌سازی بهینه دارند که اگر کمتر یا بیشتر از آن مقدار باشد، بهره‌وری انرژی کاهش می‌یابد. پس از آن با مقایسه‌ی بهره‌وری انرژی معماری‌های ممیزثابت به این نتیجه رسیدیم که معماری

- [22] M. Nikolić *et al.*, "Bitpruning: Learning bitlengths for aggressive and accurate quantization," *arXiv preprint arXiv:2002.03090*, 2020.
- [23] N. Nazari, M. Loni, M. E. Salehi, M. Daneshtalab, and M. Sjodin, "Tot-net: An endeavor toward optimizing ternary neural networks," in *2019 22nd Euromicro Conference on Digital System Design (DSD)*, 2019: IEEE, pp. 305-312.
- [24] S. A. Mirsalari, N. Nazari, S. A. Ansarmohammadi, M. E. Salehi, and S. Ghiasi, "E2BNet: MAC-free yet accurate 2-level binarized neural network accelerator for embedded systems," *Journal of Real-Time Image Processing*, vol. 18, pp. 1285-1299, 2021.
- [25] S. A. Mirsalari, N. Nazari, S. A. Ansarmohammadi, S. Sinaei, M. E. Salehi, and M. Daneshtalab, "ELC-ECG: Efficient LSTM Cell for ECG classification based on quantized architecture," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021: IEEE, pp. 1-5.
- [26] M. E. Salehi, "Binary neural networks," 2020.
- [27] N. Nazari, S. A. Mirsalari, S. Sinaei, M. E. Salehi, and M. Daneshtalab, "Multi-level binarized lstm in eeg classification for wearable devices," in *2020 28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, 2020: IEEE, pp. 175-181.
- [28] S. Gupta, A. Agrawal, K. Gopalakrishnan, and P. Narayanan, "Deep learning with limited numerical precision," in *International conference on machine learning*, 2015: PMLR, pp. 1737-1746.
- [29] F. Asim, J. Park, A. Azamat, and J. Lee, "Centered Symmetric Quantization for Hardware-Efficient Low-Bit Neural Networks," 2022: British Machine Vision Association (BMVA).
- [30] P. Judd *et al.*, "Reduced-precision strategies for bounded memory in deep neural nets," *arXiv preprint arXiv:1511.05236*, 2015.
- [31] X. Zhao, Y. Wang, X. Cai, C. Liu, and L. Zhang, "Linear symmetric quantization of neural networks for low-precision integer hardware," 2020.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [33] J. L. McKinstry *et al.*, "Discovering low-precision networks close to full-precision networks for efficient inference," in *2019 Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing-NeurIPS Edition (EMC2-NIPS)*, 2019: IEEE, pp. 6-9.
- precision," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 293-302.
- [11] S. K. Esser, J. L. McKinstry, D. Bablani, R. Appuswamy, and D. S. Modha, "Learned step size quantization," *arXiv preprint arXiv:1902.08153*, 2019.
- [12] M. Ghasemzadeh, M. Samragh, and F. Koushanfar, "ReBNet: Residual binarized neural network," in *2018 IEEE 26th annual international symposium on field-programmable custom computing machines (FCCM)*, 2018: IEEE, pp. 57-64.
- [13] T. Chen *et al.*, "Diannao: A small-footprint high-throughput accelerator for ubiquitous machine-learning," *ACM SIGARCH Computer Architecture News*, vol. 42, no. 1, pp. 269-284, 2014.
- [14] Y. Chen *et al.*, "Dadiannao: A machine-learning supercomputer," in *2014 47th Annual IEEE/ACM International Symposium on Microarchitecture*, 2014: IEEE, pp. 609-622.
- [15] P. Gysel, J. Pimentel, M. Motamedi, and S. Ghiasi, "Ristretto: A framework for empirical study of resource-efficient inference in convolutional neural networks," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5784-5789, 2018.
- [16] P. Gysel, M. Motamedi, and S. Ghiasi, "Hardware-oriented approximation of convolutional neural networks," *arXiv preprint arXiv:1604.03168*, 2016.
- [17] S. Sharify *et al.*, "Laconic deep learning inference acceleration," in *2019 ACM/IEEE 46th Annual International Symposium on Computer Architecture (ISCA)*, 2019: IEEE, pp. 304-317.
- [18] S. Ghodrati, H. Sharma, C. Young, N. S. Kim, and H. Esmailzadeh, "Bit-parallel vector composability for neural acceleration," in *2020 57th ACM/IEEE Design Automation Conference (DAC)*, 2020: IEEE, pp. 1-6.
- [19] S. Jung *et al.*, "Learning to quantize deep networks by optimizing quantization intervals with task loss," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4350-4359.
- [20] C. Gong *et al.*, "μl2q: An ultra-low loss quantization method for DNN compression," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019: IEEE, pp. 1-8.
- [21] R. Gong *et al.*, "Differentiable soft quantization: Bridging full-precision and low-bit neural networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4852-4861.

Dynamic Load Balancing Improvement in Software-Defined Networks Using Fuzzy Multi-Objective Programming Algorithms

Mohammadreza Forghani¹, Mohammadreza Soltanaghaei^{1*}, Farsad Zamani Boroujeni²

¹ Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

² Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Received: 25 December 2023, Revised: 25 April 2024, Accepted: 28 April 2024

Paper type: Research

Abstract

Software-Defined Networking (SDN) has been recognized as an efficient approach in the field of communication technology, aiming to improve the performance and efficiency of computer networks, thus reducing costs. One of the key challenges in SDN is load balancing among nodes. Solving this challenge leads to improved response time and network performance. Nowadays, various methods have been proposed for load balancing in SDN, but they have not yet reached the ideal state. In this article, a new method is presented to enhance load balancing and reduce response time. This method utilizes multi-objective evolutionary algorithms and fuzzy weighting. In the proposed method, factors such as bandwidth, traffic status, link buffer, and desired router are taken into account, and the best path and router with desired load balancing for information flows are selected with the minimum time. One prominent advantage of this method is the possibility of performing load balancing automatically without the need for human intervention. Experimental results demonstrate that the proposed method shows a significant improvement of approximately 14.8% in response time compared to other methods, while maintaining load balancing in SDNs. By using the proposed method, in addition to improving service quality and user satisfaction, response time will also be enhanced. In summary, the proposed method is introduced as a viable approach in SDNs and exhibits superiority over existing methods.

Keywords: Software-Defined Networks (SDN), Load Balancing, Multi-Objective Optimization, Fuzzy Voting Algorithm, Response Time.

* Corresponding Author's email: soltan@khuisf.ac.ir

بهبود توازن بار پویا و زمان پاسخ در شبکه‌های نرم‌افزارمحور با بهره‌گیری از الگوریتم‌های برنامه‌ریزی آرمانی چند منظوره فازی

محمدرضا فرقانی^۱، محمدرضا سلطان آقایی^{۱*}، فرساد زمانی بروجنی^۲

^۱ گروه مهندسی کامپیوتر، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران

^۲ گروه مهندسی کامپیوتر، واحد علوم و تحقیقات تهران، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۱۰/۰۴ تاریخ بازبینی: ۱۴۰۳/۰۲/۰۶ تاریخ پذیرش: ۱۴۰۳/۰۲/۰۹

نوع مقاله: پژوهشی

چکیده

شبکه نرم‌افزارمحور به عنوان یک رویکرد کارآمد در حوزه فناوری ارتباطات شناخته شده و هدف آن بهبود عملکرد و افزایش بهره‌وری و کاهش هزینه در شبکه می‌باشد. یکی از چالش‌های اساسی در شبکه‌های نرم‌افزارمحور، توازن بار بین گره‌ها است. حل این چالش باعث بهبود زمان پاسخ و عملکرد شبکه می‌شود. امروزه روش‌های متعددی برای توازن بار در شبکه‌های نرم‌افزارمحور ارائه شده است، اما هنوز به وضعیت ایده‌آل نرسیده‌اند. در این مقاله، یک روش جدید برای بهبود توازن بار و کاهش زمان پاسخ ارائه می‌شود. این روش از الگوریتم‌های برنامه‌ریزی آرمانی چند منظوره و وزن‌دهی فازی بهره می‌برد. در روش پیشنهادی، فاکتورهایی مانند پهنای باند، وضعیت ترافیک و گذردهی گره و مسیرها، بافر و مسیریاب مد نظر قرار می‌گیرند و بهترین مسیر و مسیریاب با توازن بار مطلوب برای جریان‌های اطلاعات با کمترین زمان پاسخ انتخاب می‌شوند. یکی از مزایای بارز این روش، امکان انجام توازن بار به صورت خودکار و بدون نیاز به مداخله انسان است. نتایج تجربی نشان می‌دهد که روش پیشنهادی نسبت به روش‌های دیگر، بهبود قابل توجهی در زمان پاسخ حدود ۱۴٫۸ درصد را نشان می‌دهد و همچنین توازن بار شبکه‌های نرم‌افزارمحور را حفظ می‌کند. با استفاده از روش پیشنهادی، علاوه بر بهبود کیفیت سرویس و رضایت کاربران، زمان پاسخ نیز بهبود خواهد یافت. به طور خلاصه، روش پیشنهادی به عنوان یک رویکرد قابل استفاده در شبکه‌های نرم‌افزارمحور مطرح است و نسبت به روش‌های موجود برتری دارد.

کلیدواژگان: شبکه نرم‌افزارمحور، توازن بار پویا، بهینه‌سازی چند منظوره، الگوریتم رأی‌گیری فازی، زمان پاسخ.

* رایانامه نویسنده مسؤول: soltan@khuisf.ac.ir

۱- مقدمه

بنیاد ONF^۲ و عضویت بیش از ۸۰ شرکت بزرگ صنعت شبکه در آن و تدوین استاندارد OpenFlow، این فناوری وارد فاز جدیدی شد. اولین محصولات کاربردی این شبکه از سال ۲۰۱۳ به بازار عرضه شد [۸، ۹].

در سال‌های اخیر، توسعه شبکه‌های نرم‌افزارمحور شتاب بیشتری گرفته است و اکثر تجهیزات تجاری در حال حاضر پروتکل OpenFlow را که اساس شبکه‌های نرم‌افزارمحور است، پشتیبانی می‌کنند [۱۰، ۱۱]. این توسعه سریع شبکه‌های نرم‌افزارمحور منجر به تأسیس سازمان ONF توسط شرکت‌های بزرگی مانند گوگل، فیسبوک، یاهو، مایکروسافت، Verizon و Deutsche Telekom شد. هدف اصلی این سازمان، توسعه شبکه‌های نرم‌افزارمحور با استفاده از استانداردهای متن‌باز بوده است [۱۲، ۱۳].

تحقیق حاضر بهبود کیفیت سرویس در شبکه‌های نرم‌افزارمحور با تمرکز بر مهندسی ترافیک، توازن بار ترافیکی و کاهش زمان پاسخ را مورد بررسی قرار می‌دهد. در این تحقیق، با استفاده از تکنیک‌های بهینه‌سازی ابتکاری و توسعه الگوریتم‌های برنامه‌ریزی آرمانی چند منظوره فازی، تلاش می‌شود تا توازن بار ترافیکی به صورت پویا بهبود یابد و زمان پاسخ در شبکه‌های نرم‌افزارمحور کاهش یابد [۱۴، ۱۵].

در شبکه‌های نرم‌افزارمحور، نقش لایه‌های نرم‌افزاری در طراحی و مدیریت شبکه بسیار حائز اهمیت است. با افزایش نقش لایه‌های نرم‌افزاری و کاهش نقش سخت‌افزار، امکان انعطاف‌پذیری و پیکربندی آسان‌تر شبکه‌ها فراهم می‌شود. این امر به شبکه‌ها اجازه می‌دهد تا با توجه به نیازهای مختلف، ترافیک را بهبود دهند و بار ترافیکی را به طور هوشمندانه توزیع کنند [۶، ۱۶].

در این تحقیق، با بهره‌گیری از الگوریتم‌های برنامه‌ریزی آرمانی چند منظوره فازی و تکنیک‌های بهینه‌سازی ابتکاری، تلاش می‌شود تا توازن بار ترافیکی در شبکه‌های نرم‌افزارمحور به صورت پویا بهبود یابد. همچنین، با اعمال این الگوریتم‌ها و روش‌ها، زمان پاسخ در شبکه‌های نرم‌افزارمحور بهبود می‌یابد. این بهبودها می‌تواند منجر به کاهش تأخیر در ارائه سرویس‌ها، افزایش بهره‌وری شبکه و بهبود تجربه کاربران گردد [۶، ۱۷].

الگوریتم برنامه‌ریزی آرمانی چند منظوره فازی^۳ یک روش محاسباتی است که برای حل مسائل بهینه‌سازی چند هدفه، از منطق فازی و

در سال‌های اخیر، با توسعه پیشرفت‌های فناوری اطلاعات و ارتباطات مانند تلفن همراه، سرویس‌های ابری، شبکه‌های اجتماعی، داده‌های حجیم و چندرسانه‌ای، مدیریت و پیکربندی این فناوری‌ها بسیار پیچیده، چالش‌برانگیز و زمان‌بر شده است. در حال حاضر، دسترسی به پهنای باند بالا و قابلیت گسترش و مدیریت پویا نیز از اهمیت ویژه برخوردار است [۱، ۲]. بسیاری از شبکه‌های سنتی به صورت سلسله‌مراتبی ساختاردهی می‌شوند، که با استفاده از گره‌های اترنت در یک ساختار شبکه‌ای درختی، تشکیل می‌شوند. این معماری در مورد ارتباطات مشتری یا سرویس‌دهنده قابل مشاهده است، اما برای ارتباطات پویا و نیازهای شرکت‌ها، مراکز داده و رسانه‌های سرویس‌دهنده کافی نیست [۳-۵]. معماری فعلی شبکه‌ها طوری طراحی نشده است که بتواند نیازهای کنونی شرکت‌ها، سرویس‌گیرنده‌های مخابراتی و کاربران را برآورده کند.

محدودیت‌هایی مانند پیچیدگی، سیاست‌های متناقض، عدم مقیاس‌پذیری و وابستگی به تولیدکننده محصول، هماهنگی نامناسب بین نیازهای بازار و قابلیت‌های شبکه را محدود می‌کند و صنعت فناوری اطلاعات را به یک نقطه انحراف می‌برد. به منظور جلوگیری از این روند، معماری شبکه‌های نرم‌افزارمحور^۱ (SDN) معرفی شده‌اند و استانداردها، پروتکل‌ها و نرم‌افزارهای مرتبط با آنها توسعه یافته‌اند [۶].

با توجه به افزایش کاربران و نیاز به سرویس‌های جدید و با کیفیت بالاتر و همچنین گسترش و پیچیده‌تر شدن روزافزون سخت‌افزارها و نرم‌افزارها، فناوری شبکه‌های نرم‌افزارمحور به عنوان یک رویکرد کارآمد در حوزه فناوری ارتباطات مطرح می‌شود. هدف اصلی این فناوری بهینه‌سازی استفاده از منابع، بهبود عملکرد و بهره‌وری و کاهش هزینه‌ها در شبکه‌های کامپیوتری است [۶، ۱۷].

این فناوری به منظور افزایش هوشمندی شبکه‌ها و ارتقای عملکرد آنها، با انتقال قسمت کنترل داده‌ها از سوئیچ و مسیریاب سخت‌افزاری به لایه‌های نرم‌افزار مجازی شبکه و استفاده از یک کنترل‌کننده نرم‌افزاری متمرکز، قابلیت‌هایی نظیر برنامه‌ریزی، مقیاس‌پذیری، انعطاف‌پذیری، اتوماسیون، هوشمندی و توسعه نرم‌افزاری شبکه را برای سازمان‌ها فراهم می‌کند. این ایده ابتدا در سال ۲۰۰۵ مطرح شد و از سال ۲۰۱۰ به شدت رشد کرد. با تشکیل

³ Fuzzy Multi-Objective Optimization Algorithm

¹ Software Defined Networks

² Open Networking Foundation

روش، با استفاده از انتخاب تصادفی ماشین‌های مجازی، توازن بار صورت می‌پذیرد. این مقاله راهکاری برای توازن بار بر اساس ترافیک بیش از حد ارائه می‌دهد. با جابجایی بار به ماشین مجازی که بیشترین کارایی را دارد، می‌توان زمان پاسخ کلی روش پیشنهادی را کاهش داد. نتایج حاصل نشان می‌دهد که رویکرد پیشنهادی باعث کاهش تأثیر بار حدود ۱۰ تا ۱۵ درصد می‌شود. همچنین، روش پیشنهادی باعث کاهش مدت زمان اجرا و زمان پاسخ ماشین‌های مجازی می‌شود.

مطالعه‌ی [۲۶]، به منظور توازن بار ترافیک، استفاده از کنترل‌کننده مجازی شبکه نرم‌افزارمحور (vSDN) به عنوان یک تابع شبکه مجازی (VNF) را هدف قرار داده است. با استفاده از vSDN، هنگامی که بار ترافیک نامتوازن و افزایش یابد، کنترل‌کننده‌های vSDN فرعی برای به اشتراک گذاری این بار اضافی، اضافه می‌شوند. نیاز به کنترل‌کننده‌های vSDN فرعی تعیین می‌شود و یک کپی vSDN با تنظیمات دقیقاً یکسان با کنترل‌کننده اصلی vSDN ایجاد می‌شود که با دقت عمل کرده و وظایف توازن بار ترافیک را با کنترل‌کننده اصلی vSDN به اشتراک می‌گذارد. نتایج نشان داد که بهبود ۵۰ درصدی در بار میانگین، بهبود ۴۱ درصدی در تأخیر میانگین و بهبود قابل توجهی در زمینه پاسخ ping، استفاده پهنای باند و توانایی سیستم وجود دارد. پژوهش [۲۷] چارچوب‌های DALB و CAMD را برای حل چالش عدم توازن بار در شبکه‌های SDN معرفی می‌کند. چارچوب CAMD کارآمدتر از DALB است، اما در صورت وجود بار فراگیر کارایی آن کاهش می‌یابد. مطالعه حاکی است که الگوریتم ISMDA با استفاده از مدل مهاجرت کارآمد، بهبود قابل توجهی در توازن بار دارد. این الگوریتم با افزایش توان و کاهش زمان پاسخ و فضای مهاجرت، عملکرد بهتری نسبت به CAMD ارائه می‌دهد. نتایج نشان می‌دهد که ISMDA در برابر بار فراگیر عملکرد بهتری دارد.

تحقیق [۲۸]، با استفاده از شبکه تعریف شده با نرم‌افزار، یک چارچوب مسیریابی آنلاین با صرفه‌جویی در انرژی و توازن بار پیشنهاد می‌شود. سپس، مسئله بهینه‌سازی مصرف برق و توازن بار مدل‌سازی شده و الگوریتمی برای حل آن ارائه می‌شود. همچنین الگوریتمی برای به‌روزرسانی مسیرها به علت برنامه‌ریزی جریان‌ها پیشنهاد شده است. نتایج آزمایش‌ها نشان می‌دهد که الگوریتم پیشنهادی در صرفه‌جویی برق، توازن بار و کاهش احتمال اختلال در لینک‌ها عملکرد بهتری نسبت به روش‌های مقایسه‌ای دارد. پژوهش [۲۹] راهکاری برای توازن بار و کنترل اختلال در شبکه‌های مرکز داده ارائه می‌دهد. با توجه به تغییرات پویا در الگوهای ترافیک،

مجموعه‌های فازی بهره می‌برد. این الگوریتم‌ها قادر به پیدا کردن تعادل مطلوب بین اهداف متفاوت در یک مسئله بهینه‌سازی چند هدفه هستند. الگوریتم برنامه‌ریزی آرمانی چند منظوره فازی توانایی مدل‌سازی و حل مسائلی را دارد که در آن هدف‌های متفاوت و مغایر مدنظر قرار دارند. این مسائل می‌توانند شامل بهینه‌سازی همزمان چندین هدف مانند کاهش هزینه، افزایش کیفیت، کاهش زمان، حداکثر سود و موارد دیگر باشند [۱۸-۲۱].

در الگوریتم برنامه‌ریزی آرمانی چند منظوره فازی، ابتدا مسئله بهینه‌سازی چند هدفه به شکل یک مسئله فازی مدل می‌شود. سپس با استفاده از روش‌های بهینه‌سازی مانند الگوریتم‌های تکاملی، الگوریتم‌های گرانشی، الگوریتم‌های جستجوی مستعمل و غیره، جواب بهینه‌ای برای مسئله بهینه‌سازی چند هدفه با در نظر گرفتن تعادل بین اهداف به دست می‌آید. در هر مرحله از اجرای الگوریتم برنامه‌ریزی آرمانی چند منظوره فازی، جمعیتی از نقاط کاندیدا (نقاطی که جواب ممکن به مسئله را نمایندگی می‌کنند) تولید می‌شود و سپس با استفاده از معیارهای تقریبی فازی و عملکردهای هدف، نقاط بهتر و بهینه‌تر انتخاب می‌شوند. در نهایت، جواب بهینه‌ای که حاوی تعادل مطلوب بین اهداف است، به عنوان جواب نهایی ارائه می‌شود [۲۲، ۲۳].

سازماندهی مقاله به این صورت است که، در بخش ۲، کارهای انجام شده در گذشته مورد بررسی قرار گرفته‌اند. در بخش ۳، روش پیشنهادی و جزئیات مربوط به آن تشریح شده‌اند. در نهایت، در بخش ۴، نتایج به‌دست‌آمده تحلیل می‌شوند و در نهایت در بخش ۵ نتیجه‌گیری ارائه می‌گردد.

۲- مرور ادبیات و کارهای گذشته

در این بخش به ارائه تحقیقات انجام شده در زمینه بهبود توازن بار در شبکه‌های نرم‌افزار محور پرداخته می‌شود.

پژوهش [۲۴]، روشی برای حل چالش‌های شبکه‌های تعریف شده نرم‌افزاری SDN توزیع شده معرفی می‌کند. در این روش، با استفاده از آستانه متغیری بر اساس کنترل‌کننده‌ها، از ترافیک غیرمتعادل و شلوغی در شبکه استفاده می‌شود و سپس با استفاده از ترکیب هوشمندانه الگوریتم ژنتیک و الگوریتم OPSO، بهترین کنترل‌کننده با ظرفیت مناسب برای مهاجرت انتخاب می‌شود. نتایج پیاده‌سازی و ارزیابی روش GOP-SDN، بهبود ۲۴،۷۲ درصدی در ظرفیت شبکه و کاهش ۱۳،۹۶ درصدی در تعداد مهاجرت‌ها را نشان داد. تحقیق [۲۵]، راهکاری کارآمد برای توازن بار با استفاده از الگوریتم الهام‌گرفته از زنبور عسل در محیط‌های ابری معرفی می‌کند. در این

تحقیق [۳۳]، الگوریتم جستجوی مبتنی بر چندین رگرسیون (MRBS) برای انتخاب بهینه سرویس دهنده و مسیریابی در شبکه مرکز داده پیشنهاد شده است. MRBS بر اساس تحلیل رگرسیون، سرویس دهنده را بر اساس پیش‌بینی نوع ترافیک و زمان پاسخ مبتنی بر پارامترهای داده سرویس دهنده مانند بار، زمان پاسخ و پهنای باند و استفاده از سرویس دهندگان انتخاب می‌کند. الگوریتم پیشنهادی با ترکیب الگوریتم هیوریستیک و مدل رگرسیون، انتخاب سرویس دهنده و مسیر را بهبود می‌بخشد. الگوریتم پیشنهادی با تخمین وزن‌های کاهنده گرادیان تصادفی، تأخیر و زمان را بیش از ۴۵ درصد کاهش می‌دهد و نشان می‌دهد که در مقایسه با الگوریتم‌های سنتی، استفاده از سرویس دهنده بهتری به نسبت ۸۳ درصد به دلیل برآورد وزن کاهنده گرادیان تصادفی را دارد. تحقیق [۳۴] یک روش احتمالی برای توازن بار بر اساس تحلیل واریانس ارائه می‌دهد. این روش قادر است به طور پویا جریان‌های ترافیک را مدیریت کند تا از تقاضاهای بازیابی توسط کاربران بزرگ جابجایی پشتیبانی کند. در این روش، از فناوری سوئیچینگ مجازی OpenFlow به جای فناوری سوئیچینگ سخت‌افزاری سنتی استفاده می‌شود. کنترل‌کننده SDN با استفاده از تحلیل واریانس ترافیک داده را در هر درگاه نظارت کرده و با استفاده از الگوریتم انتخاب مبتنی بر احتمال، ترافیک را به طور پویا با تکنولوژی OpenFlow هدایت می‌کند. در مقایسه با روش‌های توازن بار موجود که برای پشتیبانی از شبکه‌های سنتی طراحی شده‌اند، این روش هزینه کمتر، اعتمادپذیری بالاتر و قابلیت مقیاس‌پذیری بیشتری دارد.

پژوهش [۳۵] یک طرح تعاملی را در کنترل‌کننده SDN معرفی کرده است که از الگوریتم فراابتکاری میگو برای برنامه‌نویسی مبتنی بر بهینه‌سازی استفاده می‌کند. روش پیشنهادی بر تعامل بین متعادل‌سازی بار و بهره‌وری انرژی، استفاده از داده‌های صفحه کنترل و به‌روزرسانی‌های ماشین مجازی برای استخراج راه‌حل‌های بهینه تأکید دارد. نتایج شبیه‌سازی بهبود حدود ۳۰ درصدی راندمان انرژی و افزایش عملکرد تقریبی ۲۰ درصدی را نشان می‌دهد.

پژوهش [۳۶] درباره طرحی برای استقرار چندین کنترلر پویا با استفاده از توازن بار است. در روش پیشنهادی، درخواست‌های جریان به یک صف مدل تبدیل می‌شوند و تأخیر در انتشار ترافیک و ظرفیت کنترلرها به عنوان دو عامل اصلی در استقرار چندین کنترلر در نظر

رویکردهای موجود نیازمند تغییرات بزرگ در تجهیزات شبکه مرکز داده هستند. در روش پیشنهادی، با تشخیص دقیق جریان‌های بزرگ و ارائه روش‌های مسیریابی متفاوت برای انواع مختلف جریان‌ها، راهکاری بر اساس OpenFlow برای توازن بار در شبکه‌های مرکز داده ارائه شده است. نتایج نشان می‌دهد که استراتژی پیشنهادی قوانین جریان را به طور کافی تولید کرده و عملکرد استفاده از پهنای باند را به طرز قابل توجهی نسبت به راهکارهای دیگر در ادبیات افزایش می‌دهد.

تحقیق [۳۰] درباره نوآوری‌ها در مدیریت ترافیک شبکه در شهرهای هوشمند با استفاده از اینترنت اشیاء است. پیاده‌سازی چندین کنترل‌کننده در SDN به همراه اینترنت اشیاء، امنیت، حریم خصوصی و انعطاف‌پذیری سیستم کنترل متمرکز را در برنامه‌های شهر هوشمند بهبود می‌بخشد. نتایج آزمایشات نشان می‌دهد که استفاده از الگوریتم پیشنهادی موجب بهبود استفاده از واحد پردازش مرکزی با توازن بار شده و باعث کاهش ۲۰ درصدی در استفاده از CPU نسبت به عدم توازن بار می‌شود. مطالعه‌ی [۳۱] رویکردی برای بهینه‌سازی مسئله اختصاص سوئیچ به کنترل‌کننده با پشتیبانی از توازن بار ارائه شده است. مسئله به عنوان یک مسئله بهینه‌سازی اختصاص حداقل هزینه تخصیص دو جانبه^۱ فرموله شده است. الگوریتم جدید بر اساس مفهوم جریمه مبتنی بر بار معرفی شده است که هدف آن تعادلی بین زمان سفر رفت و برگشت و بار کنترل‌کننده است. در نهایت، پروتکل جدیدی به نام پروتکل (DHAP) به عنوان پیاده‌سازی راه‌حل پیشنهادی در محیط‌های چند کنترل‌کننده استفاده شده است. نتایج نشان می‌دهد که راه‌حل پیشنهادی در مقایسه با روش‌های موازی، بهترین عملکرد را در زمینه زمان برپا شدن جریان و توازن بار ارائه می‌دهد.

پژوهش [۳۲] راهکار جابجایی سوئیچ‌ها را برای مدیریت ترافیک ارائه کرده‌اند. اما جابجایی سوئیچ‌ها می‌تواند تأثیری بر کیفیت خدمات کاربران و زمان پاسخ پیام‌های کنترلی داشته باشد. راهکار RTSM برای بهینه‌سازی زمان پاسخ در جابجایی سوئیچ‌ها ارائه شده است. معماری SDWAN جدید با کنترل‌کننده لایه ۲، وابستگی به ارتباطات WAN را کاهش می‌دهد. شرایط Karush-Kuhn-Tucker برای انتخاب کنترل‌کننده هدف استفاده شده است. روش انتخاب سوئیچی به کمک کاربران در جابجایی دستگاه‌ها کمک می‌کند. نتایج ارزیابی نشان می‌دهد که الگوریتم پیشنهادی عملکرد بهتری در موازنه بار و زمان پاسخ شبکه دارد.

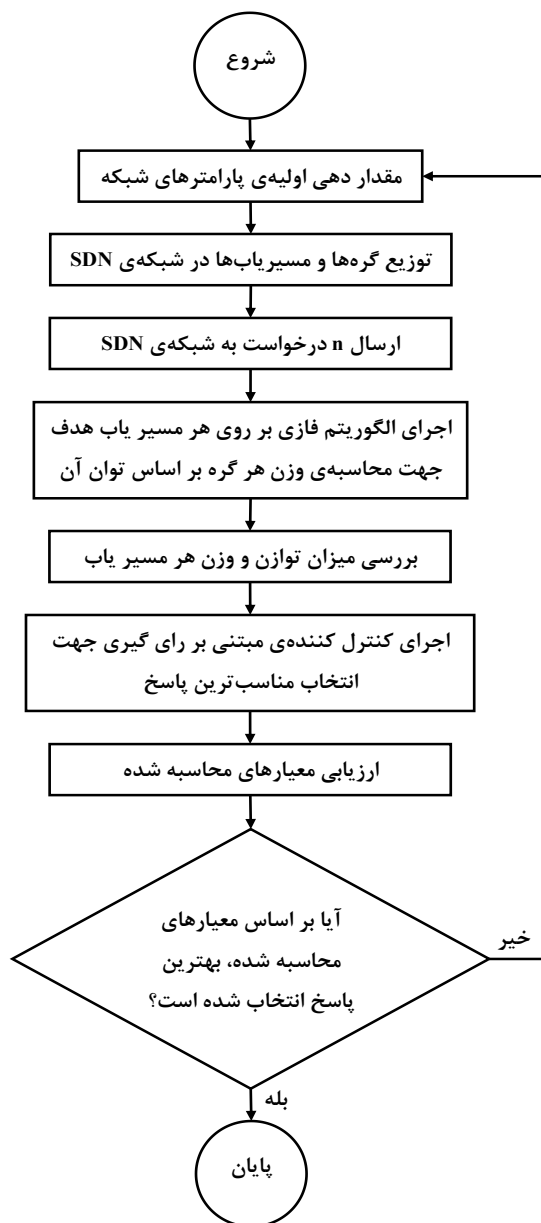
^۱ Minimum Cost Bipartite Assignment

۳-۱- مراحل کلیدی روش پیشنهادی

روش پیشنهادی شامل چهار مرحله‌ی کلیدی زیر است:

۱. اجرای الگوریتم فازی جهت تعیین وزن مطلوب
۲. فرموله کردن مدل فازی پیشنهادی
۳. مکانیسم توازن بار مسیر (Path Load)
۴. اجرای تکنیک رأی‌گیری آرمانی

در ادامه هر مرحله بصورت کامل توضیح داده می‌شود و جزئیات آن مورد بررسی قرار می‌گیرد.



شکل ۱. فلوچارت روش پیشنهادی

گرفته می‌شوند. الگوریتم اصلاح شده انتشار پیوسته^۱ (PSOAP) بر اساس بهینه‌سازی ذرات، در شبکه استاتیک اولیه برای حل مشکلات خوشه‌بندی (مشکلات تنظیم مقادیر اولیه پارامترهای بایاس و ضرایب همگرایی) و برنامه‌ریزی مناسب شبکه پیشنهاد شده است. با استفاده از شبکه ترافیک پویا، راه‌گزینی‌ها در زیردامنه‌های مختلف با استفاده از الگوریتم جستجوی Breadth-First Search (BFS) مجدداً تنظیم می‌شوند تا به توازن بار کنترل‌رها برسند. ارزیابی نشان می‌دهد که این طرح می‌تواند در مقایسه با انتشار پیوسته (AP) و الگوریتم‌های مبتنی بر ژنتیک، استقرار چندین کنترلر با ثبات، دقت و توازن بهتری داشته باشد.

تحقیقات گسترده‌ای در زمینه توازن بار انجام شده است؛ با این حال، هنوز تا دستیابی به وضعیت مطلوب فاصله زیادی وجود دارد و نیاز به تحقیقات جامع‌تری می‌باشد [۳۵، ۳۷]. در بخش بعدی روش پیشنهادی برای توازن بار پویا مورد بررسی قرار خواهد گرفت.

۳- روش پیشنهادی

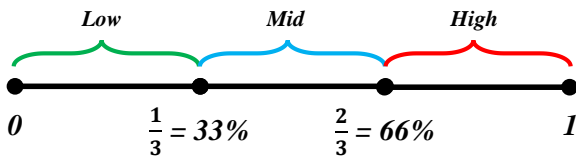
در این بخش، روش پیشنهادی برای بهبود توازن بار پویا و زمان پاسخ در SDN با استفاده از الگوریتم‌های برنامه‌ریزی آرمانی چند منظوره فازی را بررسی می‌کنیم.

با توجه به فلوچارت روش پیشنهادی در شکل ۱، در اولین مرحله، پارامترهای شبکه SDN تنظیم می‌شوند. این پارامترها ابعاد شبکه، پهنای باند مسیریاب‌ها و توان پردازشی آن‌ها می‌باشند. سپس تعداد گره‌ها و مسیریاب‌ها (G) تعیین می‌شود و این گره‌ها در کل شبکه‌ی SDN توزیع می‌شوند. سپس، تعداد درخواست‌ها (n) تعیین می‌شود و این درخواست‌ها به مسیریاب‌ها و گره‌ها ارسال می‌شوند. الگوریتم فازی بر روی هر مسیریاب اجرا می‌شود و وزن گره‌ها بر اساس توان آن محاسبه می‌شود. در نهایت، با توجه به وزن‌های محاسبه شده، بهترین پاسخ بر اساس ارزیابی معیارها انتخاب می‌شود و مسیریاب مورد نظر، عملیات درخواست یا درخواست‌های دریافت شده را انجام می‌دهد. در این حالت، مسیریابی که وزن بیشتری دارد، به عنوان مسیریاب مطلوب انتخاب می‌شود. بنابراین با انتخاب بهترین و کوتاه‌ترین مسیر برای انجام هر عمل (Task) زمان پاسخ کاهش پیدا می‌کند و با کاهش زمان پاسخ می‌توان ادعا کرد که توازن بار نیز بهبود خواهد یافت. در ادامه، مراحل مختلف روش پیشنهادی توضیح داده شده و با جزئیات بیشتری مورد بررسی قرار می‌گیرند.

¹ Particle Swarm Optimization base on Affinity Propagation

جدول ۱. قوانین و شرایط موجود برای انتخاب گره مطلوب

میزان گذردهی (Gateway) (Throughput)	پهنای باند / سربار داده (Overload/Bandwidth)		
	Low	Mid	High
Low	Mid	Worst	Worst
Mid	Normal	Mid	Bad
High	Good	Normal	Mid



شکل ۲. فاکتور قطبیت بر اساس پهنای باند یا سربار داده و همچنین فاکتور قطبیت بر اساس میزان گذردهی

انتخاب گره مطلوب بر اساس شروط زیر انجام می‌شود:

R1=if (Overloading/BW = Low & Gateway = Low) then select_ R node = Mid

R2=if (Overloading/BW = Mid & Gateway = Low) then select_ R node = Worst

R3=if (Overloading/BW = High & Gateway = Low) then select_ R node = Worst

R4=if (Overloading/BW = Low & Gateway = Mid) then select_ R node = Normal

R5=if (Overloading/BW = Mid & Gateway = Mid) then select_ R node = Mid

R6=if (Overloading/BW = High & Gateway = Mid) then select_ R node = Bad

R7=if (Overloading/BW = Low & Gateway = High) then select_ R node = Good

R8=if (Overloading/BW = Mid & Gateway = High) then select_ R node = Normal

R9=if (Overloading/BW = High & Gateway = High) then select_ R node = Mid

این نکته حائز اهمیت است که در صورتی که دو گره با شرایط مساوی وجود داشته باشند، برای انتخاب یکی از آنها باید از تابع برازندگی (Fitness) که از رابطه ۱ محاسبه می‌شود استفاده نمود.

$$Fitness = \frac{Overloading}{BW} + \frac{1}{Gateway} \quad (1)$$

$$Selectnode = Max (Fitness1, Fitness2)$$

در اینجا Overloading بار پردازشی، BW پهنای باند و Gateway میزان گذردهی یا نرخ ترافیک آزاد در درگاه شبکه است. با استفاده از روش پیشنهادی، می‌توان بهبود قابل توجهی در عملکرد SDN

۳-۱-۱- اجرای الگوریتم فازی جهت تعیین وزن مطلوب

پس از اینکه گره‌ها در شبکه توزیع شدند و درخواست‌ها نیز به شبکه جهت اجرا ارسال شدند، می‌بایست فرآیند پردازش این درخواست‌ها صورت پذیرد. بدین منظور، ابتدا با کمک منطق فازی اقدام به محاسبه وزن هر مسیریاب به منظور اجرای درخواست‌ها می‌گردد. زمانی که کنترلر می‌خواهد یک گره مطلوب را انتخاب کند تا سایر گره‌ها بتوانند اطلاعات مربوطه را ارسال نمایند، سیستم فازی نیز اعمال می‌گردد. جهت انتخاب گره مطلوب از یک سری قوانین فازی به صورت زیر استفاده می‌شود. دو پارامتر اصلی که برای انتخاب گره‌های مطلوب، استفاده می‌شود عبارتند از میزان گذردهی و پهنای باند یا سربار داده.

برای انتخاب درست، باید قوانین فازی را تشکیل داده و بر اساس آن‌ها عمل کنیم. جدول ۱، قوانین انتخاب گره مطلوب بر اساس میزان نرخ گذردهی و همچنین بر اساس پهنای باند / سربار داده در سه حالت $\{Low, Mid, High\}$ را مشخص می‌کند. همچنین در جدول ۱، مقدار Bad انتخاب ضعیف گره از نظر پهنای باند / سربار داده و مقدار Worst انتخاب نامطلوب و غیرقابل قبول گره از بدترین میزان از پهنای باند / سربار داده را از نظر فازی نشان می‌دهد.

با توجه به جدول ۱ و قوانین فازی ارائه شده، توابع عضویت مربوط به میزان گذردهی و میزان پهنای باند / سربار داده و فاکتور قطبیت بر اساس میزان گذردهی، به صورت شکل ۱ تعریف می‌گردد.

انتخاب یک گره مطلوب بر اساس قوانین بالا قابل انجام است. به عنوان مثال، فرض کنید گره ۱ دارای پهنای باند / سربار داده‌ای برابر با ۸۰٪ باشد و گره ۲ دارای پهنای باند / سربار داده‌ای برابر با ۶۰٪ باشد. همچنین، میزان گذردهی از گره ۱ به گره بعدی حدوداً ۵۰ کیلوبایت است و میزان گذردهی از گره ۲ به گره بعدی حدوداً ۷۰ کیلوبایت است. بنابراین، احتمال انتخاب هر گره به شرح زیر است: به دلیل داشتن پهنای باند / سربار داده بالا (۸۰٪) و میزان گذردهی تا گره بعدی معادل ۵۰ کیلوبایت، احتمال انتخاب گره ۱ در حالت خوب (Good) است. همچنین، به دلیل داشتن پهنای باند / سربار داده متوسط (۶۰٪) و میزان گذردهی تا گره بعدی معادل ۷۰ کیلوبایت، احتمال انتخاب گره ۲ در حالت ضعیف است. در نهایت، گره شماره ۱ انتخاب می‌شود.

مقصد، برای اطمینان از کارآیی، طرح پیشنهادی به دو مرحله تقسیم شده است.

مرحله اول: الگوریتم انتخاب مسیرهای Top-K

مرحله دوم: ارزیابی ترکیبی فازی^۴

بنابراین ابتدا مسیرهای Top-K انتخاب می‌شوند و سپس هنگام ورود جریان جدید یا تغییر وضعیت شبکه، با استفاده از روش ارزیابی ترکیبی فازی، بهترین مسیر از بین مسیرهای Top-K به صورت پویا انتخاب می‌شود.

مرحله اول: الگوریتم انتخاب مسیرهای Top-K

معمول است که بیش از یک مسیر بین دو گره در یک شبکه وجود داشته باشد، به ویژه برای گره‌های مهم یا با تقاضای اطمینان بالا. در صورت وجود تعداد زیادی مسیر، انتخاب بهترین مسیر در زمان واقعی به صورت بهینه غیرکارآمد خواهد بود. بنابراین، در ابتدا مسیرهای Top-K که به عنوان کوتاه‌ترین مسیرها در نظر گرفته می‌شوند، انتخاب می‌شوند. این انتخاب می‌تواند به دو صورت انجام شود: یکی به صورت اولیه و دیگری به صورت دوره‌ای.

در حالت اولیه، هیچ ترافیکی در شبکه وجود ندارد، بنابراین مسیرهای Top-K کوتاه‌ترین مسیرها خواهند بود. از الگوریتم کوتاه‌ترین مسیر K برای پیدا کردن K مسیر کوتاه‌تر استفاده می‌شود. این الگوریتم، یک توسعه از الگوریتم کوتاه‌ترین مسیر است که K مسیر کوتاه‌تر را مشخص می‌کند. در این مقاله، از الگوریتم کوتاه‌ترین مسیر K بر اساس الگوریتم Floyd که یک الگوریتم مبتنی بر برنامه نویسی پویا برای محاسبه کوتاه‌ترین مسیر بین تمامی گره‌ها در توپولوژی شبکه می‌باشد، استفاده شده است. اصل اساسی این الگوریتم به این صورت است که یک نقشه ماتریسی را برای جستجوی کوتاه‌ترین مسیر ثبت می‌کند. برای هر جفت از گره‌ها i و j ، الگوریتم به تدریج گره‌های جدیدی را به عنوان گره میانی به مسیر اصلی اضافه می‌کند. اگر طول مسیر جدید کوتاه‌تر از مسیر اصلی باشد، مسیر اصلی و نقشه ماتریسی با مسیر جدید به‌روزرسانی می‌شوند. این فرایند با استفاده از رابطه‌ی ۲ انجام می‌شود.

$$map[i, j] = \min((map[i, k] + map[k, j]), map[i, j]) \quad (2)$$

در اینجا i و j به ترتیب مقادیر سطر و ستون نقشه‌ی ماتریسی هستند. اگرچه مسیرهای Top-K در ابتدا بر اساس تعداد واسطه‌ها

داشت. در جدول ۲ نمادهای پرکاربرد مورد استفاده در این مقاله و توضیحات آنها آورده شده است.

۳-۱-۲- فرموله کردن مدل فازی پیشنهادی

در یک محیط شبکه پویا، انتخاب بهترین مسیر برای جریان داده‌ها چالشی پیچیده است، زیرا تأثیرات فازی متعددی که بر این انتخاب تأثیر می‌گذارند، وجود دارد. هدف‌ها و محدودیت‌ها در این مسئله به طور طبیعی فازی هستند. توازن بار در مسیریابی، یک مسئله تصمیم‌گیری در محیط شبکه با خصوصیات فازی است. برای مقابله با ابهامات، می‌توان از تئوری فازی بهره‌برداری کرده و اهداف و محدودیت‌های فازی را به عنوان مجموعه‌های فازی دقیق تعریف کرده و با استفاده از ترکیب اهداف و محدودیت‌های داده‌شده، تصمیم‌گیری فازی انجام داد. بنابراین، در این مقاله، استفاده از مکانیسم ارزیابی ترکیبی فازی^۱ (FSEM) برای کنترل توازن بار در مسیریابی شبکه‌های نرم‌افزارمحور پیشنهاد شده است.

جدول ۲. نمادهای پرکاربرد

نماد	توضیحات
Overloading	بار پردازشی در شبکه
BW	پهنای باند ^۲
Gateway	میزان گذردهی در درگاه شبکه
Top-K	مسیرهای برتر برای انتخاب بهترین مسیر
h	تعداد واسطه‌ها یا پرش‌ها ^۳ در مسیر
p	تعداد بسته‌های انتقال داده شده
b	تعداد بایت‌های سوئیچ بحرانی
r	نرخ ارسال و دریافت سوئیچ (مسیریاب) مورد نظر
U	دامنه فاکتور
V	دامنه رتبه
A	بردار وزن
R	ماتریس رابطه فازی
N	مجموعه فازی منفی
t	نرخ ارسال و دریافت مسیریاب یا سوئیچ
M	جابجایی یا انتقال داده‌ها
ZE	مجموعه فازی صفر
P	مجموعه فازی مثبت
r_{ij}	درجه وابسته به فاکتور u_i در دامنه‌ی U و فاکتور v_j در دامنه‌ی V
b_k	ارزیابی نهایی بهترین مسیر k

با توجه به تعداد زیادی مسیر ممکن بین یک گره مبدأ و یک گره

³ Hop

⁴ Fuzzy Synthetic Evaluation

¹ Fuzzy Synthetic Evaluation Mechanism

² Band Width

در روش پیشنهادی، دامنه رتبه به وسیله یک یا چند مجموعه فازی تشکیل می‌شود که نشان‌دهنده رتبه‌بندی گزینه‌های مختلف است. این مجموعه‌ها شامل منفی (N)، صفر (ZE) و مثبت (P) می‌باشند. با توجه به اینکه هدف اصلی، انتخاب بهترین مسیر است، می‌توان $V = (P)$ تعریف کرد، که در اینجا P نشان‌دهنده مجموعه فازی مثبت است. در روش پیشنهادی، فاکتورها به وسیله ماتریس رابطه فازی R ارزیابی شده و از رابطه ۶ محاسبه می‌شود.

$$R = \begin{bmatrix} r_{11} & \dots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \dots & r_{nm} \end{bmatrix} \quad (6)$$

در این ماتریس، r_{ij} نشان‌دهنده درجه‌ای است که به فاکتور u_i در دامنه‌ی U و فاکتور v_j در دامنه‌ی V وابسته است. در روش پیشنهادی، عضویت در مجموعه P برای هر ویژگی U محاسبه می‌شود. عملکرد عضویت h ، p ، b و t از رابطه ۷ بدست می‌آید [۳۸].

$$\begin{aligned} r_h &= 1.0/e^h \\ r_p &= 1.0/\log(p + 0.1) \\ r_b &= 1.0/\log(b + 0.1) \\ r_t &= 1.0/\left(1 + e^{\frac{-t}{50.0}}\right) \end{aligned} \quad (7)$$

در اینجا r_t ، نرخ ارسال و دریافت مسیریاب یا سوئیچ است. برای محاسبه r_t ، از ماتریس رابطه فازی R استفاده می‌شود که در آن هر درایه ماتریس مقداری را برای درجه ارتباط بین فاکتورهای مختلف نشان می‌دهد. در این روش، ابتدا عضویت در مجموعه فازی مثبت (P) برای هر ویژگی U محاسبه می‌شود، سپس با استفاده از ماتریس رابطه فازی R ، مقادیر درجه ارتباط بین ویژگی‌ها و وابستگی آن‌ها به هم، به عنوان مقادیر r_t محاسبه می‌شوند.

سپس، ماتریس R می‌تواند به عنوان بردار ستونی برای یک مسیر مشخص از رابطه‌ی ۸ محاسبه شود.

$$R = \begin{pmatrix} r_h \\ r_p \\ r_b \\ r_t \end{pmatrix} \quad (8)$$

برای تعیین بردار وزن $A = (a_1, a_2, a_3, \dots, a_n)$ ، باید درجه اهمیت هر عامل در دامنه U مشخص شود. در اینجا، برای متعادل‌سازی بار مسیر، A به صورت $(0.4, 0.15, 0.15, 0.3)$ تعریف

(ه‌ا‌پ) انتخاب می‌شوند، اما مسیر ترافیکی در شبکه برای مسیره‌های Top-K با استفاده از روش ارزیابی ترکیبی فازی انتخاب می‌شود. فاصله زمانی می‌تواند به مدت پنج دقیقه باشد. به این ترتیب، مسیره‌های Top-K قادر به سازگاری با تغییرات ترافیک شبکه می‌باشند.

مرحله دوم: ارزیابی ترکیبی فازی

برای ارزیابی مسیره‌ها در شبکه‌های نرم‌افزارمحور، عوامل متعددی تأثیر می‌گذارند. ارزیابی مسیر می‌تواند به عنوان یک مسئله تصمیم‌گیری چندمعیاری در نظر گرفته شود. بنابراین، یک مدل ارزیابی ترکیبی فازی با ویژگی‌های چندگانه به عنوان یک مدل تصمیم‌گیری فازی برای توازن بار مسیر پیشنهاد شده است. از دیدگاه عوامل چندگانه، درجات رضایت از ویژگی‌ها به وسیله مجموعه‌های فازی مرتبط به طور کامل از طریق تابع عضویت ارزیابی می‌شود. علاوه بر این، این امکان را فراهم می‌کند تا درجه خاصی از اهمیت به هر معیار اختصاص داده شود. یک تصمیم بهینه به عنوان نقطه‌ای در فضای گزینه‌های ممکن تعریف شده است که در آن عملکرد یک تصمیم فازی به حداکثر مقدار خود می‌رسد. فاکتور مقدار دامنه U برای ارزیابی مسیر از رابطه ۳ محاسبه می‌شود.

$$U = (u_1, u_2, u_3, \dots, u_n) \quad (3)$$

برای ارزیابی مسیره‌ها، دو جنبه در نظر گرفته می‌شود. یکی طول مسیر است که از شمارش تعداد واسطه‌ها بدست می‌آید و دیگری، بار ترافیکی در طول مسیر است که از ترافیک سوئیچ‌ها و پیوندهای مسیر استخراج می‌شود. از آنجا که یک مسیر از تعداد زیادی سوئیچ و پیوند تشکیل شده است، میانگین یا کل ترافیک نمی‌تواند وضعیت واقعی مسیر را نشان دهد، بنابراین سوئیچ‌ها و پیوندها برای نشان دادن وضعیت مسیر انتخاب می‌شوند. ترافیک سوئیچ با استفاده از تعداد بسته‌ها و تعداد بایت‌هایی که توسط سوئیچ ارسال می‌شود، اندازه‌گیری می‌گردد. بنابراین میزان ترافیک پیوند و میزان ارسال درگاه مربوطه اندازه‌گیری می‌شود. دامنه فاکتور u برای ارزیابی مسیر از رابطه‌ی ۴ محاسبه می‌شود.

$$u = (h, p, b, t) \quad (4)$$

در اینجا، h تعداد واسطه‌ها (ه‌ا‌پ)، p تعداد بسته‌های انتقال داده شده و b تعداد بایت‌های سوئیچ بحرانی و t نرخ ارسال و دریافت مسیریاب یا سوئیچ مورد نظر است. برای تعیین دامنه رتبه‌ی V برای هر عامل از رابطه‌ی ۵ استفاده می‌شود.

$$V = (v_1, v_2, v_3, \dots, v_n) \quad (5)$$

در ادامه هر یک از ماژول‌ها توضیح داده می‌شوند.

۱. ماژول جمع‌آوری داده‌ها

این ماژول شامل بررسی توپولوژی و جمع‌آوری داده‌های ترافیک است. بررسی توپولوژی به کنترلر کمک می‌کند تا با داشتن اطلاعات توپولوژی شبکه، آگاهی کاملی از موقعیت میزبان‌ها در سراسر شبکه داشته باشد. این اطلاعات با استفاده از پیوندهایی که با ارسال بسته‌های داده LLDP^۲ تشکیل شده‌اند، جمع‌آوری می‌شود. پروتکل LLDP یک پروتکل استاندارد در لایه ۲ مدل OSI با هدف شناسایی شبکه است که به کلیه تجهیزات شبکه (مستقل از برند) اجازه می‌دهد اطلاعاتی درباره خود را به دستگاه‌های همسایه ارسال و مبادله کنند. این پروتکل به‌ویژه در شبکه‌های بزرگ که شامل تجهیزات مختلف از برندهای گوناگون هستند، کاربرد دارد و به مدیریت و نگهداری شبکه کمک می‌کند. هنگامی که پیوندها بین گره‌ها ایجاد یا خراب می‌شوند، رویدادهای پیوند ایجاد می‌شوند. جمع‌آوری داده‌های ترافیک اطلاعات آماری مربوط به تعداد بسته‌ها، تعداد بایت‌های ارسال شده توسط سوئیچ و میزان ارسال درگاه مربوطه را از سوئیچ‌های OpenFlow جمع‌آوری می‌کند.

۲. ماژول ارزیابی مسیر

این ماژول از الگوریتم مسیرهای Top-K و الگوریتم ارزیابی ترکیبی فازی استفاده می‌کند. با توجه به اینکه ممکن است تعداد زیادی مسیر اضافی در شبکه وجود داشته باشد، کنترلر ابتدا باید مسیرهای بهینه Top-K را انتخاب کند. این الگوریتم بر اساس معیارهایی مانند پهنای باند، تأخیر و هزینه‌های مسیریابی، مسیرهای بهینه را انتخاب می‌کند. سپس بر اساس این مسیرهای بهینه، مسیرهای بیشتری را برای ارزیابی انتخاب می‌کند. الگوریتم ارزیابی ترکیبی فازی بر اساس منطق فازی و قوانین قابلیت دسترسی به منابع شبکه، بارگذاری مسیر و توازن بار مسیر، مسیرهای بهینه را ارزیابی می‌کند. با استفاده از این الگوریتم، کنترلر می‌تواند تصمیم بگیرد که کدام مسیرها برای بارگذاری مسیر استفاده شوند.

۳. ماژول نصب جدول جریان

این ماژول بر اساس مسیرهایی که توسط ماژول ارزیابی مسیر انتخاب شده‌اند، جداول جریان را در سوئیچ‌ها نصب می‌کند. این جداول جریان مشخص می‌کنند که بسته‌هایی که از یک مسیر خاص

شده است، به این معنی که طول مسیر اهمیت بیشتری دارد و حجم کار سوئیچ و پیوند بحرانی مشترک هستند. نتایج ارزیابی نهایی، با ترکیب بردار A و ماتریس R از رابطه‌ی ۹ قابل محاسبه است.

$$B = (b_1, b_2, b_3, \dots, b_i)$$

$$b = A \otimes R$$

$$b = (a_1, a_2, a_3, \dots, a_n) \otimes \begin{bmatrix} r_{11} & \dots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \dots & r_{nm} \end{bmatrix} \quad (9)$$

در اینجا، نماد \otimes ضرب تک تک درایه‌های ماتریس خطی A در تک تک ستون‌های ماتریس R و مقدار ارزیابی است. در نتیجه ارزیابی مسیرهای K از رابطه‌ی ۱۰ محاسبه می‌شود.

$$b_k = (0.4, 0.15, 0.15, 0.3) \otimes \begin{bmatrix} r_{h1} & r_{h2} & \dots & r_{hk} \\ r_{p1} & r_{p2} & \dots & r_{pk} \\ r_{b1} & r_{b2} & \dots & r_{bk} \\ r_{t1} & r_{t2} & \dots & r_{tk} \end{bmatrix} \quad (10)$$

در اینجا b_k ارزیابی نهایی مسیر k است که بصورت زیر محاسبه می‌شود:

$$b_k = (0.4 \times r_{hk}) + (0.15 \times r_{pk}) + (0.15 \times r_{bk}) + (0.3 \times r_{tk})$$

۳-۱-۳- مکانیسم توازن بار مسیر^۱

راه‌حل پیشنهادی در این مقاله برای توازن بار مسیر در شبکه‌های نرم‌افزارمحور از معماری لایه‌بندی شبکه‌های نرم‌افزارمحور استفاده می‌کند. این معماری شامل لایه‌ی کنترل و لایه‌ی داده است. مسیری که ترافیک شبکه را از گره منبع به گره مقصد منتقل می‌کند، باید بهینه باشد تا بار را به طور متعادل بین مسیرها تقسیم کند و همچنین از منابع شبکه به طور کامل استفاده کند. برای این منظور، کنترلر با استفاده از الگوی ارزیابی مسیر، مسیرهایی را انتخاب می‌کند که برای بارگذاری مسیر مفید هستند. این عملیات توسط سه ماژول انجام می‌شود:

۱. ماژول جمع‌آوری داده‌ها

۲. ماژول ارزیابی مسیر

۳. ماژول نصب جدول جریان

¹ Path load balancing mechanism

² Link Layer Discovery Protocol

هاپ، تعداد بسته‌ها، تعداد بایت و نرخ ارسال درگاه سوئیچ هستند. این شاخص‌ها از اطلاعات سراسری شبکه، شامل اطلاعات گره‌ها و پیوندها، جمع‌آوری می‌شوند. بردار وزن A برابر با $(0.4, 0.15, 0.15)$ (0.3) است که عناصر آن به ترتیب با شاخص‌های هاپ، تعداد بسته‌ها، تعداد بایت و نرخ ارسال بسته مطابقت دارند. سپس چهار عامل عملکرد با عضویت تعریف شده است. در نهایت، بردار وزن A با ماتریس R ترکیب می‌شود تا به نتایج ارزیابی برسیم. نمره ارزیابی نهایی مسیر بهینه را نشان می‌دهد. مسیریابی روش پیشنهادی در الگوریتم ۲ نشان داده شده است.

الگوریتم ۲: مسیریابی روش پیشنهادی Top-K

Input: Top-K paths K_PATH determined
Output: The optimal path $BEST_PATH$

- 1: Define the factor domain U (hops, bytes, packets, port_rate) and rank domain $V(P)$ for a path
- 2: For each i in the range $\{0, 1, 2, \dots, K-1\}$:
- 3: Calculate $hops[i]$ as the length of $K_PATH[i]$
- 4: Calculate $bytes[i]$ as the total bytes of $K_PATH[i]$
- 5: Calculate $packets[i]$ as the total packets of $K_PATH[i]$
- 6: Calculate $port_rate[i]$ as the transfer rate of $K_PATH[i]$
- 7: EndFor
- 8: For each j in $\{0, 1, 2, \dots, K-1\}$:
- 9: Transform $hops[j]$ by applying the function $1.0 / \exp(hops[j])$
- 10: Transform $bytes[j]$ by applying the function $1.0 / \log(bytes[j] + 0.1)$
- 11: Transform $packets[j]$ by applying the function $1.0 / \log(packets[j] + 0.1)$
- 12: Transform $port_rate[j]$ by applying the function $1.0 / (1 + \exp(-port_rate[j]/50.0))$
- 13: EndFor
- 14: Define weight vector A ($weight_hops$, $weight_bytes$, $weight_packets$, $weight_port_rate$)
- 15: For each k in $\{0, 1, 2, \dots, K-1\}$:
- 16: Calculate $temp_hops[k]$ as the product of $weight_hops$ and $hops[k]$
- 17: Calculate $temp_bytes[k]$ as the product of $weight_bytes$ and $bytes[k]$
- 18: Calculate $temp_packets[k]$ as the product of $weight_packets$ and $packets[k]$
- 19: Calculate $temp_port_rate[k]$ as the product of $weight_port_rate$ and $port_rate[k]$
- 20: Calculate $score[k]$ as the sum of $temp_hops[k]$, $temp_bytes[k]$, $temp_packets[k]$, and $temp_port_rate[k]$
- 21: EndFor
- 22: $BEST_PATH$ is equal to the path with the maximum score among all scores

این الگوریتم برای مسیریابی با استفاده از روش Top-K طراحی شده است. در این روش، مجموعه‌ای از K بهترین مسیر از بین مسیرهای موجود انتخاب می‌شود. مراحل این الگوریتم به شرح زیر است: ابتدا، دامنه عوامل (مانند تعداد واسطه‌ها، حجم داده‌ها، تعداد بسته‌ها و نرخ درگاه) و دامنه رتبه‌بندی برای مسیرها تعیین می‌شود. سپس برای هر مسیر از 0 تا $K-1$ ، ویژگی‌های مختلف محاسبه شده و مقادیر

عبور می‌کنند، به کدام درگاه‌ها هدایت شوند. روش مسیریابی پیشنهادی Top-K در الگوریتم ۱ نشان داده شده است.

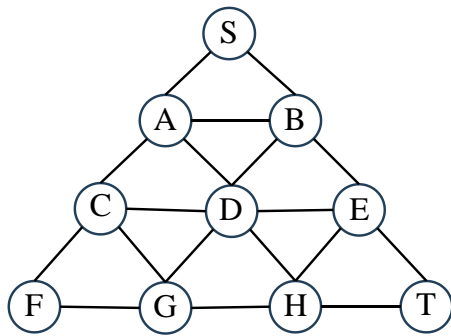
الگوریتم ۱: یافتن کوتاه‌ترین مسیر از مبدا به مقصد با روش پیشنهادی

Input: Graph G represented by a set of nodes and edges, starting node S , destination node T

Output: Shortest path from node S to node T

- 1: Initialize $A[0]$ to represent the shortest path from node S to node T
- 2: For each k in the range $\{1, 2, \dots, K-1\}$:
- 3: For each i in the range $\{0, 1, 2, \dots, \text{length}(A[k-1]) - 2\}$:
- 4: Set $currentNode$ as the i -th node in $A[k-1]$
- 5: For each j in the range $\{0, 1, 2, \dots, K-1\}$:
- 6: Set $previousNode$ as the j -th node in $A[k-1]$
- 7: If $currentNode$ is equal to $previousNode$:
- 8: Set the distance between $currentNode$ and the $(i+1)$ -th node as infinity
- 9: End if
- 10: EndFor inner loop
- 11: Calculate $S[i]$ representing the shortest path from $currentNode$ to node T
- 12: Calculate $R[i]$ as the shortest path from node S to $currentNode$
- 13: Calculate $B[i]$ as the sum of $R[i]$ and $S[i]$
- 14: EndFor outer loop
- 15: Update $A[k]$ to represent the shortest path with the minimum length among all $B[i]$
- 16: Restore the distance between $currentNode$ and $previousNode$ to its original value
- 17: EndFor outer loop

در این الگوریتم، برای یافتن کوتاه‌ترین مسیر از یک گره به گره دیگر در یک گراف، مراحل زیر انجام می‌شود: ابتدا، کوتاه‌ترین مسیر از گره شروع (S) به گره مقصد (T) در یک لیست با نام $A[0]$ ذخیره می‌شود. در مراحل بعدی، برای هر مرحله k (که نشان‌دهنده تعداد مراحل محاسبه کوتاه‌ترین مسیر است)، به ازای هر گره i در مسیر قبلی ($A[k-1]$)، مسیر جدیدی که از گره فعلی تا گره مقصد (T) را محاسبه می‌کنیم (که با $S[i]$ نشان داده می‌شود). همچنین، کوتاه‌ترین مسیر از گره S به گره فعلی را نیز محاسبه می‌کنیم و در لیست $R[i]$ ذخیره می‌کنیم. سپس مجموع این دو مسیر، به عنوان کوتاه‌ترین مسیر از گره S به گره مقصد (T) را در لیست $B[i]$ ذخیره می‌کنیم. پس از پایان مراحل برای هر k ، مسیری که کوتاه‌ترین مسیر از گره S به گره مقصد (T) است را در لیست $A[k]$ ذخیره می‌کنیم. در پایان هر مرحله، مقادیر فاصله‌های میان گره‌های متصل را به حالت اولیه برمی‌گردانیم و مراحل برای مرحله k بعدی را انجام می‌دهیم. این مراحل با استفاده از ارتباطات میان مسافت مسیره‌ها و محاسبات فاصله‌ی بین مسیرهای مختلف، در نهایت منجر به یافتن کوتاه‌ترین مسیر از یک گره به گره دیگر در گراف می‌شود. الگوریتم ۲ برای ارزیابی ترکیبی فازی برای ارزیابی مسیرهای بهینه Top-K در شبکه استفاده می‌شود. شاخص‌های بارگذاری شامل



شکل ۳. مثالی از عملکرد الگوریتم‌های مسیریابی ۱ و ۲ برای پیدا کردن یک مسیر از S به T

حالا با استفاده از الگوریتم ۱، مراحل اجرای الگوریتم به صورت زیر است:

- در مرحله‌ی اول، مسیر از S به T به صورت (S, A, B, E, T) محاسبه می‌شود.
- در مرحله دوم، مسیری جدید از A به T به صورت (A, B, E, T) به عنوان کوتاه‌ترین مسیر محاسبه می‌شود.
- در مرحله سوم، مسیری جدید از B به T به صورت (B, E, T) به عنوان کوتاه‌ترین مسیر محاسبه می‌گردد.
- در نهایت، کوتاه‌ترین مسیر از S به T برابر با (S, A, B, E, T) به دست می‌آید.

اما با استفاده از الگوریتم ۲، مراحل اجرای الگوریتم به صورت زیر است:

ابتدا ویژگی‌های هر مسیر محاسبه می‌شود. برای مسیرهای ممکن، ویژگی‌هایی مانند تعداد پرش‌ها، حجم داده‌ها، تعداد بسته‌ها و نرخ درگاه محاسبه می‌شود. سپس ویژگی‌ها با توجه به وزن‌های مشخص شده تبدیل می‌شوند. از بین مسیرهای محاسبه شده، مسیری که بیشترین امتیاز را دارد به عنوان مسیر بهینه انتخاب می‌شود. بنابراین ممکن است با توجه به ویژگی‌ها و وزن‌های مشخص شده، مسیر (S, A, C, D, E, T) به عنوان مسیر بهینه انتخاب شود.

در مثال بالا، الگوریتم ۱ به صورت دقیق و تکاملی مسیرهای کوتاه را محاسبه می‌کند، در حالی که الگوریتم ۲ بر اساس امتیازدهی به ویژگی‌های مسیرها عمل می‌کند و ممکن است مسیری که به طور دقیق کوتاه‌ترین نباشد را به عنوان مسیر بهینه انتخاب کند.

تکنیک رای‌گیری آرمانی، آخرین تکنیک بهبود توازن بار پویا و کاهش زمان پاسخ در روش پیشنهادی است که در ادامه توضیح داده می‌شود.

آنها تبدیل می‌شوند. سپس بردار وزن برای ارزیابی مسیرها با استفاده از وزن‌های مختلف مانند تعداد پرش‌ها، حجم داده‌ها، تعداد بسته‌ها و نرخ درگاه تعیین می‌شود. سپس برای هر مسیر، با استفاده از وزن‌های تعیین شده و مقادیر ویژگی‌های محاسبه شده، یک امتیاز محاسبه می‌شود. در نهایت، بهترین مسیر برابر با مسیری که بیشترین امتیاز را دارد، انتخاب می‌شود. این الگوریتم با استفاده از یک ترکیب خطی از ویژگی‌های مختلف مسیرها و وزن‌های مشخص شده، بهترین مسیر را از بین مسیرهای Top-K انتخاب می‌کند.

در الگوریتم ۲، $K_PATH[i]$ نشان‌دهنده مسیرهای برتر (Top-K paths) می‌باشد. این مسیرها در واقع کوتاه‌ترین مسیرها از یک گره به گره دیگر در گراف است که با در نظر گرفتن ویژگی‌های مختلف مانند تعداد پرش‌ها، حجم داده‌ها، تعداد بسته‌ها و نرخ درگاه محاسبه می‌شوند. مقدار $K_PATH[i]$ ، می‌تواند تا K مسیر مختلف باشد که در لیست K_PATH ذخیره شده‌اند. به عبارت دیگر، $K_PATH[i]$ می‌تواند مجموعه‌ای از مسیرهای مختلف باشد که به عنوان مسیرهای برتر از نظر کوتاه‌ترین مسیرها در گراف در نظر گرفته شده‌اند و K ، تعداد مسیرهای برتر است.

تفاوت الگوریتم ۱ و الگوریتم ۲ در حل مسئله یافتن مسیر بهینه در یک گراف در این است که، در الگوریتم ۱، به صورت تکاملی و دقیق، کوتاه‌ترین مسیر از گره شروع به گره مقصد محاسبه می‌شود. در هر گام از محاسبه، از نتایج قبلی استفاده می‌شود و مسیرها به صورت دقیق و بر اساس داده‌های قبلی محاسبه می‌شوند. اما در الگوریتم ۲، ابتدا ویژگی‌های مسیرهای مختلف، مانند تعداد پرش‌ها، حجم داده‌ها، تعداد بسته‌ها و نرخ درگاه، محاسبه می‌شود و سپس با توجه به وزن‌های مشخص شده، مسیر بهینه انتخاب می‌شود. این روش انتخاب مسیر بر اساس امتیازدهی به ویژگی‌های مسیرها انجام می‌شود، به طوری که مسیری با بیشترین امتیاز به عنوان مسیر بهینه انتخاب می‌شود. به عنوان مثال گراف شکل ۳ را در نظر بگیرید. فرض کنید می‌خواهیم مسیر بهینه از گره S به گره T را پیدا کنیم. این گراف شامل مجموعه از گره‌های S, A, B, C, D, E, F, G, H و T می‌باشد.

۳-۱-۴- اجرای تکنیک رای گیری آرمانی

روش پیشنهادی در این مقاله با استفاده از تکنیک رای گیری آرمانی، بهبود توازن بار پویا و کاهش زمان پاسخ در شبکه‌های نرم‌افزارمحور را هدف قرار می‌دهد. در این روش، توازن بار بین مسیریاب‌ها بر اساس وزن‌هایی که از آنها بدست می‌آید، تنظیم می‌شود. میزبان هدف، سیگنال‌هایی به کنترلر ارسال می‌کند و مسیریاب‌های همسایه وزن‌هایی بین صفر تا یک بر اساس بازدهی خود ارسال می‌کنند. سپس با رأی گیری آرمانی بر اساس حداکثر آرا، مسیریاب با کمترین وزن به عنوان مسیریاب مطلوب انتخاب می‌شود.

معادله ۱۱ نحوه رأی گیری مبتنی بر حداکثر آرا را نشان می‌دهد. با استفاده از این معادله، بازدهی مسیریاب مورد نظر بر اساس وزن‌ها تعیین می‌شود و مسیریاب مطلوب انتخاب می‌شود.

$$B_s = \text{Max}(w_1, w_2, w_3, \dots, w_n) \quad (11)$$

در اینجا B_s وزن ماکزیمم و w_n وزن مسیریاب می‌باشند. با استفاده از تکنیک رای گیری آرمانی، این روش قادر است بهبود قابل توجهی در توازن بار و زمان پاسخ شبکه‌های نرم‌افزارمحور را ایجاد کند. با انتخاب مسیریابی مناسب بر اساس بازدهی و وزن‌ها، بار شبکه بهینه توزیع می‌شود و زمان پاسخ به درخواست‌ها بهبود می‌یابد. این روش می‌تواند در شبکه‌های SDN و شبکه‌های نرم‌افزارمحور کارآمدی بالا را ارائه دهد.

۳-۲- معیارهای ارزیابی در روش پیشنهادی

در این بخش برخی از مهمترین معیارهای ارزیابی عملکرد روش پیشنهادی مورد بررسی قرار می‌گیرد. این معیارهای ارزیابی عبارتند از:

الف- نرخ ارسال داده در شبکه: میزان داده‌هایی است که از یک دستگاه به دستگاه دیگر در شبکه ارسال می‌شود، معمولاً به صورت بیت یا بایت در ثانیه اندازه‌گیری می‌شود، طبق رابطه ۱۲.

$$S = \frac{A(mb)}{T(s)} \quad (12)$$

در اینجا S نرخ ارسال داده بر حسب مگابایت در ثانیه، مقدار داده A بر حسب مگابایت و زمان انتقال T بر حسب ثانیه است.

ب- نرخ دریافت داده شبکه نسبت به ارسال: نسبت داده‌هایی که یک دستگاه یا یک رابط شبکه دریافت می‌کند به داده‌هایی که

از همان دستگاه یا رابط شبکه به سایر دستگاه‌ها یا رابط‌های شبکه ارسال می‌شود را نرخ دریافت داده شبکه نسبت به ارسال می‌گویند و مقدار آنرا با R/S بیان می‌کنند. این نرخ معمولاً به صورت تعداد بیت‌ها یا بایت‌های دریافتی به تعداد بیت‌ها یا بایت‌های ارسالی در یک بازه زمانی خاص محاسبه می‌شود، و به عنوان یک نسبت یا درصد بیان می‌شود و مقدار آن از رابطه‌ی ۱۳ محاسبه می‌گردد. در شبکه‌های مختلف، این نرخ‌ها ممکن است متفاوت باشند و بستگی به نوع و تنظیمات شبکه دارند.

$$R/S = \frac{R(mb)}{S(mb)} \times 100 \quad (13)$$

در اینجا R/S نسبت دریافت به ارسال بر حسب درصد، $R(mb)$ و $S(mb)$ به ترتیب حجم داده‌های ارسالی و دریافتی بر حسب مگابایت است.

تأخیر تبادلات انتها به انتها: زمانی است که طول می‌کشد تا یک پیام از یک دستگاه ارسال کننده به دستگاه گیرنده برسد، شامل زمان پردازش، انتقال و تأخیر در دستگاه‌های مختلف است و از رابطه‌ی ۱۴ محاسبه می‌شود.

$$D = \sum_{i=1}^{i=n} (RT_i - ST_i) \quad (14)$$

در اینجا D میزان تأخیر تبادل، RT زمان دریافت و ST زمان ارسال داده در شبکه است.

زمان پاسخ شبکه: زمان پاسخ شبکه به مدت زمانی اشاره دارد که یک بسته داده برای رسیدن از یک دستگاه به دستگاه دیگر در شبکه نیاز دارد. این زمان معمولاً به صورت میلی‌ثانیه (ms) یا میکروثانیه (μs) اندازه‌گیری می‌شود که از رابطه‌ی ۱۵ محاسبه می‌شود.

$$RT = MT + PT + DT + \delta T \quad (15)$$

در اینجا RT زمان پاسخ، MT زمان جابجایی یا انتقال، PT زمان پردازش، DT زمان تأخیر در پردازش داده و δT زمان‌های ناشی از اختلالات شبکه هستند.

با بررسی این معیارها، کوتاه‌ترین مسیرها برای SDN شناسایی شده و با اجرای آنها توازن بار در شبکه بهبود یافته و زمان پاسخ شبکه کاهش خواهد یافت.

¹ End-to-End Delay

۴- نتایج تجربی

در این بخش، نتایج روش پیشنهادی و شبیه‌سازی آن ارائه می‌شود و با یک روش مشابه دیگر مقایسه می‌شود.

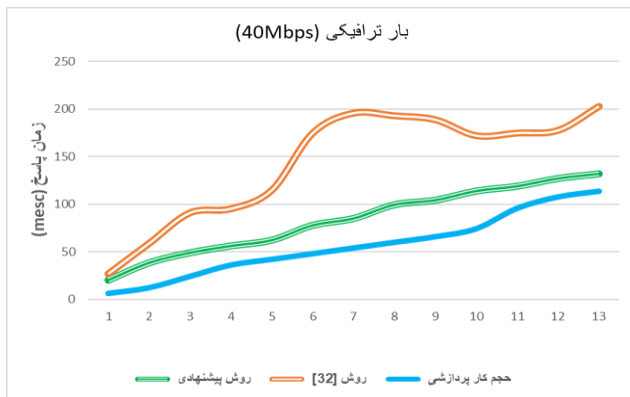
۴-۱- نتایج شبیه‌سازی روش پیشنهادی

در این بخش، نتایج شبیه‌سازی روش پیشنهادی جهت بهبود توازن بار و زمان پاسخ با یک روش مشابه دیگر مورد مقایسه قرار گرفته است. برای توصیف توازن بار، از عوامل مختلفی شامل نرخ ارسال و دریافت داده در شبکه، تأخیر تبادلات و زمان پاسخ شبکه استفاده شده است. در این تحلیل، حجم درخواست‌های ورودی به عنوان ترافیک بار مد نظر قرار گرفته و حجم پردازش CPU نشان می‌دهد چند درخواست ورودی در حال پردازش هستند. برای توصیف توازن بار، یک مسئله کاهش زمان پاسخ مورد تعریف قرار گرفته است و برای بهترین مقایسه با روش پیشنهادی، از مدل صف درخواست‌های M/M/c با توزیع بار مشخص استفاده شده است [۳۲].

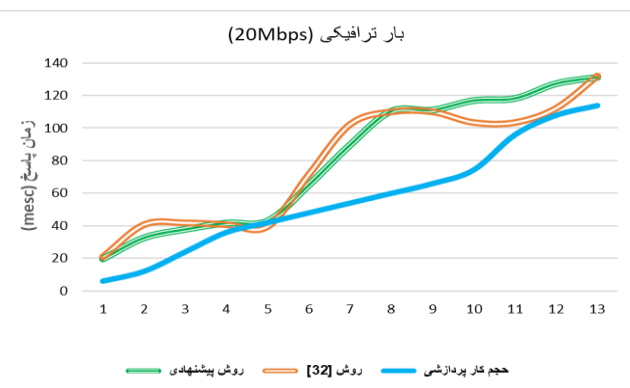
در این شبیه‌سازی، از آخرین نسخه‌های Mininet V2.3.0 و Floodlight V1.2 به عنوان دو ابزار برجسته در زمینه شبیه‌سازی شبکه‌های SDN استفاده شده است. این ابزارها به خوبی قابلیت تنظیم پارامترهای مختلف شبیه‌سازی براساس سناریوهای روش پیشنهادی را دارند. همچنین، برای تست عملکرد و توازن بار پویا، از ابزار Locust به عنوان پایگاه داده استفاده شده است که یک ابزار متن‌باز با قابلیت تعریف سناریوهای تست با تنظیم پارامترهای مربوطه است و برای ارزیابی و تجزیه و تحلیل نتایج تست، امکان گزارش‌دهی جامع را فراهم می‌کند. اعتبار و قابلیت‌های این ابزارها، در انجام شبیه‌سازی‌ها و آزمایشات مربوط به توازن بار و عملکرد شبکه و SDN بسیار مفید و حائز اهمیت هستند. سناریویی که در شبیه‌سازی روش پیشنهادی استفاده شده به این صورت است که در یک SDN شامل سوئیچ‌ها و یک کنترل‌کننده مرکزی، با اعمال یک بار ترافیک پویا و متغیر، میزان بار و زمان پاسخ در نقاط مختلف شبکه اندازه‌گیری شده است. تنظیمات شبکه شامل تعداد 100 عدد سوئیچ و یک کنترل‌کننده مرکزی و پارامترهای برنامه‌ریزی SDN شامل پروتکل ارتباط بین کنترل‌کننده و سوئیچ^۱ gRPC و الگوریتم مسیریابی^۲ OSPF می‌باشد. همچنین، ترافیک با الگوی پویا و متغیر شبیه‌سازی شده و نرخ ارسال درخواست‌ها و حجم داده‌ها تعیین شده است. با استفاده از ماژول‌های اندازه‌گیری بار و زمان پاسخ در نقاط مختلف شبکه، میزان بار و زمان پاسخ در سوئیچ‌ها و

کنترل‌کننده‌ها اندازه‌گیری شده است. در نهایت، با تحلیل داده‌های جمع‌آوری شده، توازن بار در سوئیچ‌ها و کنترل‌کننده‌ها مورد بررسی قرار گرفته و زمان پاسخ در نقاط مختلف شبکه مقایسه و تحلیل شده است. نتایج با استفاده از روش پیشنهادی و روش‌های مقایسه‌ای [۳۲] مورد ارزیابی قرار گرفته است.

در شکل ۴، زمان پاسخ روش پیشنهادی و روش [۳۲] با سه بار ترافیکی مختلف نشان داده شده است. نتایج نشان می‌دهند که در بار ترافیکی (40Mbps)، در روش پیشنهادی زمان پاسخ کمتری نسبت به روش [۳۲] دارد. همچنین، در بار ترافیکی (20Mbps)، نیز روش پیشنهادی زمان پاسخ کمتری دارد. اما در شرایط بار ترافیکی خیلی کم و نزدیک به (0Mbps)، روش [۳۲] زمان پاسخ کمتری نسبت به روش پیشنهادی دارد. بررسی دقیق‌تر نشان می‌دهد روش پیشنهادی در مقابله با بار ترافیکی بالا و درخواست‌های پردازشی سنگین عملکرد بهتری دارد. در حالی که روش [۳۲] در شرایط بار ترافیکی کم می‌تواند زمان پاسخ کمتری داشته باشد. از آنجایی که در اکثر موارد، بار ترافیکی بالاست می‌توان ادعا کرد که روش پیشنهادی عملکرد بهتری نسبت به روش مشابه دارد.



(الف)



(ب)

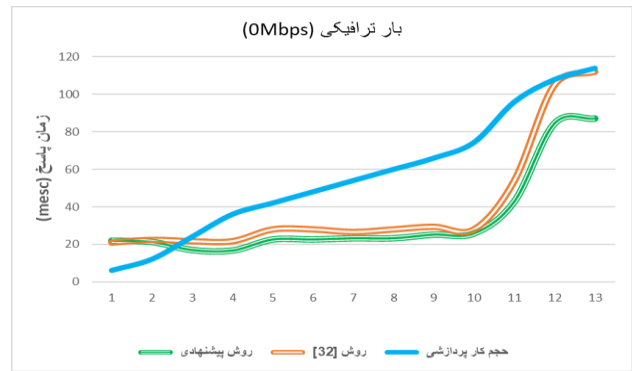
² Open Shortest Path First

¹ Google Remote Procedure Call

SDN عملکرد بهتری نسبت به سایر روش‌ها از نظر زمان پاسخ، میانگین زمان پاسخ، تعادل بار و حجم درخواست‌های پردازش شده دارد. این نتایج نشان می‌دهند که روش پیشنهادی می‌تواند به عنوان یک روش بهینه در این حوزه معرفی شود.

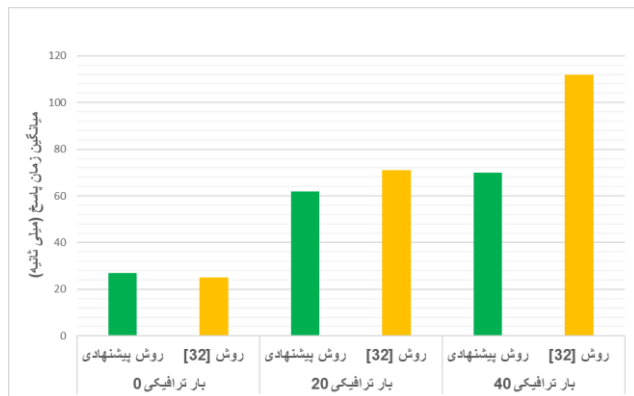
۵- نتیجه‌گیری

با توجه به نتایج و پیشنهادات این مطالعه، مشخص می‌شود که مدیریت حجم درخواست‌های پردازش شده در SDN امری بحرانی و حیاتی است که نیازمند ابزارها و روش‌های بهینه و کارآمد می‌باشد.

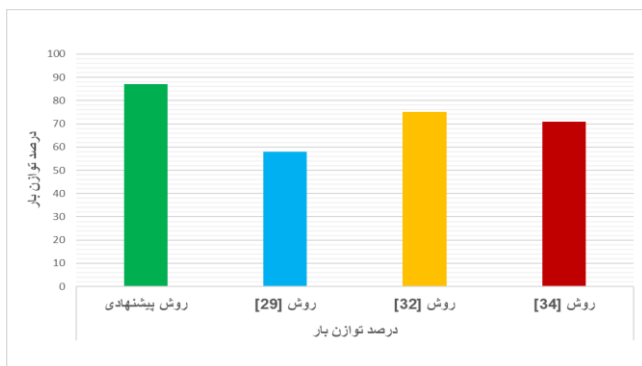


(ج)

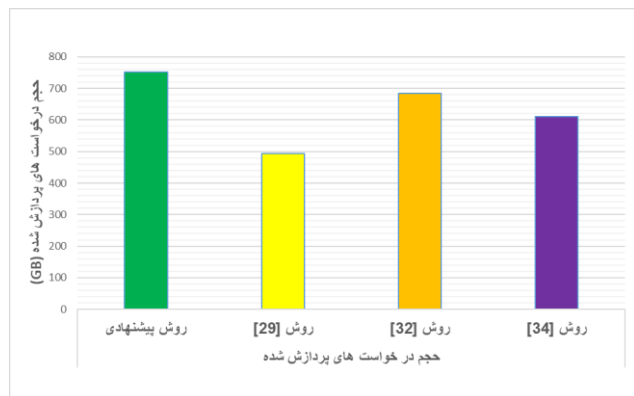
شکل ۴. مقایسه‌ی زمان پاسخ روش پیشنهادی با روش [۳۲] در سه بار ترافیکی (الف) (۴۰Mbps)، (ب) (۲۰Mbps) و (ج) (۰Mbps)



شکل ۵. مقایسه میانگین زمان پاسخ روش پیشنهادی و روش [۳۲] با بار ترافیکی (۰، ۲۰، ۴۰).



شکل ۶. مقایسه درصد تعادل بار روش پیشنهادی با سه روش [۲۹]، [۳۲] و [۳۴].



در شکل ۵، نتایج مقایسه‌ای از میانگین زمان پاسخ روش پیشنهادی و روش [۳۲] با بارهای ترافیکی (۰، ۲۰، ۴۰) نمایش داده شده است. نتایج آماری نمودار شکل ۶ نشان می‌دهد که در شرایط بار ترافیکی کم، روش [۳۲] با کمی تفاوت، عملکرد بهتری نسبت به روش پیشنهادی دارد. اما در شرایط بار ترافیکی متوسط و بالا، روش پیشنهادی عملکرد بهتری نسبت به روش [۳۲] ارائه می‌دهد. این نتایج همچنین برتری روش پیشنهادی را نسبت به روش مشابه، همانند جدول ۱ نشان می‌دهند. در آزمایشی دیگر، درصد توازن بار روش پیشنهادی با سه روش [۲۹]، [۳۲] و [۳۴]، مقایسه شده است. نمودار مقایسه درصدها در شکل ۶، نشان داده شده است. بر اساس نتایج نمودار، روش پیشنهادی دارای درصد توازن بار بالایی برابر با ۸۷ درصد است. در مقابل، روش [۲۹] با درصد توازن بار ۵۸ درصد، روش [۳۲] با درصد توازن بار ۷۵ درصد و روش [۳۴] با درصد توازن بار ۷۱ درصد عملکرد خود را نشان داده‌اند. روش پیشنهادی در توازن بار بهبود قابل توجهی نسبت به روش [۲۹] دارد و به طور معمول از روش‌های [۳۲] و [۳۴] عملکرد بهتری را ارائه می‌دهد. این نتایج نیز برتری روش پیشنهادی را نسبت به سایر روش‌های مقایسه شده در این تحقیق نشان می‌دهد.

آخرین آزمایش حجم درخواست‌های پردازش شده در یک زمان ثابت در روش پیشنهادی و روش‌های [۲۹]، [۳۲] و [۳۴]، است که نتایج مقایسه در شکل ۷، نشان داده شده است. با توجه به نمودار نشان داده شده، روش پیشنهادی در پردازش حجم درخواست‌ها عملکرد بهتری نسبت به روش‌های [۲۹] و [۳۴] دارد و همچنین از روش‌های [۳۲] حجم بیشتری را پردازش می‌کند. این نتایج می‌توانند به عنوان راهنمایی برای انتخاب مناسب‌ترین روش در مدیریت حجم درخواست‌ها در شبکه و SDN استفاده شوند.

با توجه به نتایج تجربی به دست آمده، روش پیشنهادی در زمینه

- [6] J. Gong and A. Rezaeipanah, "A fuzzy delay-bandwidth guaranteed routing algorithm for video conferencing services over SDN networks," (in eng), *Multimed Tools Appl*, pp. 1-30, Jan 23 2023, doi: 10.1007/s11042-023-14349-6.
- [7] Z. Liu, X. Dong, L. Wang, J. Feng, C. Pan, and Y. Li, "Satellite Network Task Deployment Method Based on SDN and ICN," (in eng), *Sensors (Basel)*, vol. 22, no. 14, Jul 21 2022, doi: 10.3390/s22145439.
- [8] D. Yang and W. T. Tsai, "SDN-Based Congestion Control and Bandwidth Allocation Scheme in 5G Networks," (in eng), *Sensors (Basel)*, vol. 24, no. 3, Jan 24 2024, doi: 10.3390/s24030749.
- [9] H. Xu, W. D. Liu, L. Li, and Q. Zhou, "An IoT-based low-cost architecture for smart libraries using SDN," *Sci Rep*, vol. 14, no. 1, p. 7022, Mar 25 2024, doi: 10.1038/s41598-024-57484-2.
- [10] S. E. dashti and S. Shabooei, "Improving resource allocation in mobile edge computing using gray wolf and particle swarm optimization algorithms," *Journal of Information and Communication Technology*, vol. 59, no. 16, pp. 108-124, 2024. [Online]. Available: <http://rimag.ir/fa/Article/39782>.
- [11] S. Naderi, "Application identification through intelligent traffic classification," *Journal of Information and Communication Technology*, vol. 59, no. 16, pp. 264-278, 2024. [Online]. Available: <http://rimag.ir/fa/Article/42244>.
- [12] D. Wang et al., "DoSDefender: A Kernel-Mode TCP DoS Prevention in Software-Defined Networking," (in eng), *Sensors (Basel)*, vol. 23, no. 12, Jun 8 2023, doi: 10.3390/s23125426.
- [13] Z. B. Zuo, R. Y. He, X. W. Zhu, and C. W. Chang, "A novel software-defined network packet security tunnel forwarding mechanism," (in eng), *Math Biosci Eng*, vol. 16, no. 5, pp. 4359-4381, May 17 2019, doi: 10.3934/mbe.2019217.
- [14] Y. Guo et al., "Traffic Management in IoT Backbone Networks Using GNN and MAB with SDN Orchestration," (in eng), *Sensors (Basel)*, vol. 23, no. 16, Aug 10 2023, doi: 10.3390/s23167091.
- [15] L. Li, K. Li, X. Meng, Y. Wang, and X. Wang, "Dynamic weight routing and optical-code algorithm based on SDN," (in eng), *Heliyon*, vol. 9, no. 1, p. e12407, Jan 2023, doi: 10.1016/j.heliyon.2022.e12407.
- [16] M. Hussain, N. Shah, R. Amin, S. S. Alshamrani, A. Alotaibi, and S. M. Raza, "Software-Defined Networking: Categories, Analysis, and Future Directions," (in eng), *Sensors (Basel)*, vol. 22, no. 15, Jul 25 2022, doi: 10.3390/s22155551.
- [17] M. Hamzei, S. Khandagh, and N. Jafari Navimipour, "A Quality-of-Service-Aware Service Composition Method in the Internet of Things Using a Multi-Objective Fuzzy-Based Hybrid Algorithm," (in eng), *Sensors (Basel)*, vol. 23, no. 16, Aug 17 2023, doi: 10.3390/s23167233.
- [18] G. Yuan, Y. Yang, G. Tian, and A. M. Fathollahi-Fard, "Capacitated multi-objective disassembly scheduling with fuzzy processing time via a fruit fly optimization algorithm," (in eng), *Environ Sci Pollut Res Int*, Jan 31 2022, doi: 10.1007/s11356-022-18883-y.
- [19] H. Li, D. Ou, and Y. Ji, "An Environmentally Sustainable Software-Defined Networking Data Dissemination Method for Mixed Traffic Flows in RSU Clouds with Energy Restriction," (in eng), *Int J Environ Res Public Health*, vol. 19, no. 22, Nov 16 2022, doi: 10.3390/ijerph192215112.
- [20] K. Salimi and M. Mollamotalebi, "Improving the load balancing in Cloud computing using a rapid SFL algorithm (R-SFLA)," (*Journal of Information and Communication Technology*, vol. 57, no. 57, p. 191, 2023.
- [21] A. Ghorbannia Delavar and R. Akraminejad, "WSTMOS: A Method For Optimizing Throughput, Energy, And Latency In Cloud

شکل ۷. مقایسه حجم درخواست‌های پردازش شده در روش پیشنهادی و سایر روش‌ها.

در این مطالعه، یک روش پیشنهادی با عملکرد برتر در زمینه زمان پاسخ، میانگین زمان پاسخ، توازن بار و حجم درخواست‌ها ارائه شده است که می‌تواند به عنوان یک راهکار مناسب برای مدیریت SDN مورد استفاده قرار گیرد.

با اجرای روش پیشنهادی، امکان بهبود کارایی و بهره‌وری در شبکه و SDN وجود دارد. این بهبود در نتایج عملکرد و اجرای شبکه‌ها، توازن بار بین منابع، کاهش زمان پاسخ و ارتقای تجربه کاربران تأثیر مثبتی خواهد داشت. بنابراین، روش پیشنهادی می‌تواند به عنوان یک انتخاب مناسب برای مدیران شبکه و متخصصان SDN در بهبود کارایی و عملکرد شبکه و SDN مورد استفاده قرار گیرد. برای پیشرفت و توسعه بیشتر در این حوزه، پژوهش‌های آینده می‌توانند به بهبود و گسترش روش پیشنهادی، بررسی تأثیرات و عملکرد آن در شرایط مختلف، و بررسی قابلیت اعمال آن در شبکه‌های بزرگتر و پیچیده‌تر متمرکز شوند. همچنین، با بررسی ترکیب روش پیشنهادی با روش‌های دیگر و ارائه روش‌های ترکیبی، می‌توان دقت و کارایی را بهبود بخشید. این تحقیقات آینده می‌توانند به پیشرفت و توسعه حوزه مدیریت حجم درخواست‌ها در SDN کمک کرده و بهبودی مداوم در عملکرد و کارایی این سیستم‌ها را فراهم نمایند.

مراجع

- [1] B. Isyaku, K. B. Abu Bakar, N. M. Yusuf, M. Abaker, A. Abdelmaboud, and W. Nagmeldin, "Software defined wireless sensor load balancing routing for internet of things applications: Review of approaches," (in eng), *Heliyon*, vol. 10, no. 9, p. e29965, May 15 2024, doi: 10.1016/j.heliyon.2024.e29965.
- [2] M. Hussain, R. Amin, R. Gantassi, A. H. Alshehri, J. Frnda, and S. M. Raza, "Efficient handling of ACL policy change in SDN using reactive and proactive flow rule installation," (in eng), *Sci Rep*, vol. 14, no. 1, p. 14976, Jun 28 2024, doi: 10.1038/s41598-024-65721-x.
- [3] K. Rui, H. Pan, and S. Shu, "Secure routing in the Internet of Things (IoT) with intrusion detection capability based on software-defined networking (SDN) and Machine Learning techniques," *Scientific Reports*, vol. 13, no. 1, p. 18003, 2023/10/21 2023, doi: 10.1038/s41598-023-44764-6.
- [4] K. Luo, "A distributed SDN-based intrusion detection system for IoT using optimized forests," (in eng), *PLoS One*, vol. 18, no. 8, p. e0290694, 2023, doi: 10.1371/journal.pone.0290694.
- [5] I. Smolka and J. Stój, "Utilization of SDN Technology for Flexible EtherCAT Networks Applications," *Sensors*, vol. 22, no. 5, p. 1944, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/5/1944>.

- internet of things in SDN for smart cities," *Sustainability*, vol. 13, no. 17, p. 9587, 2021.
- [31] A. El Kamel and H. Youssef, "Improving switch-to-controller assignment with load balancing in multi-controller software defined WAN (SD-WAN)," *Journal of Network and Systems Management*, vol. 28, pp. 553-575, 2020.
- [32] K. S. Sahoo, M. Tiwary, B. Sahoo, B. K. Mishra, S. RamaSubbaReddy, and A. K. Luhach, "RTSM: Response time optimisation during switch migration in software-defined wide area network," *IET wireless sensor systems*, vol. 10, no. 3, pp. 105-111, 2020.
- [33] G. S. Begam, M. Sangeetha, and N. Shanker, "Load balancing in dcn servers through sdn machine learning algorithm," *Arabian Journal for Science and Engineering*, pp. 1-12, 2022.
- [34] K. A. Jadhav, M. M. Mulla, and D. Narayan, "An efficient load balancing mechanism in software defined networks," in 2020 12th international conference on computational intelligence and communication networks (CICN), 2020: IEEE, pp. 116-122.
- [35] M. Forghani, M. Soltanaghaei, and F. Z. Boroujeni, "Dynamic optimization scheme for load balancing and energy efficiency in software-defined networks utilizing the krill herd meta-heuristic algorithm," *Computers and Electrical Engineering*, vol. 114, p. 109057, 2024.
- [36] G. Li, X. Wang, and Z. J. I. A. Zhang, "SDN-based load balancing scheme for multi-controller deployment," vol. 7, pp. 39612-39622, 2019.
- [37] Z. Li and E. Peng, "Software-Defined Optimal Computation Task Scheduling in Vehicular Edge Networking," (in eng), *Sensors (Basel)*, vol. 21, no. 3, Feb 1 2021, doi: 10.3390/s21030955.
- [38] T. Han and N. Ansari, "A traffic load balancing framework for software-defined radio access networks powered by hybrid energy sources," *IEEE/ACM Transactions on Networking*, vol. 24, no. 2, pp. 1038-1051, 2015.
- Workflow Scheduling," *Journal of Information and Communication Technology*, vol. 57, no. 57, p. 62, 2023.
- [22] H. Xue, K. T. Kim, and H. Y. Youn, "Dynamic Load Balancing of Software-Defined Networking Based on Genetic-Ant Colony Optimization," (in eng), *Sensors (Basel)*, vol. 19, no. 2, Jan 14 2019, doi: 10.3390/s19020311.
- [23] X. Xu, W. K. Jia, Y. Wu, and X. Wang, "On the Optimal Lawful Intercept Access Points Placement Problem in Hybrid Software-Defined Networks," (in eng), *Sensors (Basel)*, vol. 21, no. 2, Jan 9 2021, doi: 10.3390/s21020428.
- [24] Z. Kabiri, B. Barekatin, and A. Avokh, "GOP-SDN: an enhanced load balancing method based on genetic and optimized particle swarm optimization algorithm in distributed SDNs," *Wireless Networks*, vol. 28, no. 6, pp. 2533-2552, 2022.
- [25] R. Sharma, I. Sharma, and A. Sharma, "Load Balancing and Resource Utilization Approach in Cloud Computing Using Honey Bee-Inspired Algorithm," in International Conference on Mobile Computing and Sustainable Informatics: ICMCSI 2020, 2021: Springer, pp. 811-820.
- [26] S. Ejaz, Z. Iqbal, P. A. Shah, B. H. Bukhari, A. Ali, and F. Aadil, "Traffic load balancing using software defined networking (SDN) controller as virtualized network function," *IEEE Access*, vol. 7, pp. 46646-46658, 2019.
- [27] O. Adekoya, A. Aneiba, and M. Patwary, "An improved switch migration decision algorithm for SDN load balancing," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1602-1613, 2020.
- [28] Y. Zhao, X. Wang, Q. He, C. Zhang, and M. Huang, "PLOFR: An online flow route framework for power saving and load balance in SDN," *IEEE Systems Journal*, vol. 15, no. 1, pp. 526-537, 2020.
- [29] X. Shi et al., "An openflow-based load balancing strategy in SDN," *Comput. Mater. Contin.*, vol. 62, no. 1, pp. 385-398, 2020.
- [30] H. Babbar, S. Rani, D. Gupta, H. M. Aljahdali, A. Singh, and F. Al-Turjman, "Load balancing algorithm on the immense scale of

Improving Code Coverage Metrics for Discovering Vulnerabilities in Stateful Network Protocols using Hybrid Fuzzing

Hamid Rezaei Rahvard¹, Mehdi Salkhordeh Haghighi²

¹ Faculty of Computer Engineering and IT, Sadjad University, Mashhad, Iran

² Faculty of Computer Engineering and IT, Sadjad University, Mashhad, Iran

Received: 15 October 2023, Revised: 05 January 2024, Accepted: 02 February 2024

Paper type: Research

Abstract

Fuzzing software is a method for finding security vulnerabilities in applications. In this method, by sending random data to the program, attempts are made to find cases that lead to undesirable behaviors and errors such as memory corruption or unauthorized access. One of the proposed methods for improving and enhancing fuzzing is the use of symbolic analysis and dynamic-symbolic execution. In this method, in addition to generating random data, logical analysis of the program and its symbolic execution are used to generate data that can cover new paths in program execution. In this research, we have shown that the dynamic-symbolic execution method can be used for fuzzing network protocols and also improve this process. For this purpose, the first framework for hybrid fuzzing of network protocols has been designed and implemented. The results on two services dcmtk and dnsmasq show that hybrid fuzzing performs better in terms of code coverage compared to traditional fuzzing. Branch coverage in the dcmtk service improved by 2.71 percent compared to AFLNet, which was able to make the negative performance of NyxNet compared to AFLNet positive. Also, branch coverage in the dnsmasq service improved by 37.72 percent compared to AFLNet and by 11.82 percent compared to NyxNet.

Keywords: Fuzz Testing, Network Protocol Testing, Vulnerabilities, Symbolic Execution, Concolic Execution.

* Corresponding Author's email: haghighi@sadjad.ac.ir

بهبود معیار پوشش کد برای کشف آسیب‌پذیری در پروتکل‌های شبکه دارای حالت توسط فازینگ ترکیبی

حمید رضایی رهورد^۱، مهدی سالخورده حقیقی^{۲*}

^۱ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه سجاد، مشهد، ایران

^۲ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه سجاد، مشهد، ایران

تاریخ دریافت: ۱۴۰۲/۰۷/۲۳ تاریخ بازبینی: ۱۴۰۲/۱۰/۱۵ تاریخ پذیرش: ۱۴۰۲/۱۱/۱۳

نوع مقاله: پژوهشی

چکیده

فازینگ نرم‌افزار، روشی برای یافتن آسیب‌پذیری‌های امنیتی در برنامه‌های کاربردی است. در این روش با ارسال داده‌های تصادفی به برنامه، سعی می‌شود مواردی پیدا شود که منجر به رفتارهای نامطلوب و خطاهایی همچون خرابی حافظه یا دسترسی‌های غیرمجاز شود. یکی از روش‌های پیشنهادی برای بهبود و اثربخشی فازینگ، استفاده از تحلیل نمادین و اجرای پویا-نمادین است. در این روش علاوه بر تولید داده‌های تصادفی، از تحلیل منطقی برنامه و اجرای نمادین آن برای تولید داده‌هایی استفاده می‌شود که بتوانند مسیرهای جدیدی از اجرای برنامه را پوشش دهند. در این پژوهش نشان داده‌ایم که می‌توان از روش اجرای پویا-نمادین برای فازینگ پروتکل‌های شبکه استفاده نمود و همچنین این فرایند را بهبود بخشید. بدین منظور اولین چارچوب برای فازینگ ترکیبی پروتکل‌های شبکه طراحی و پیاده‌سازی شده است. نتایج بر روی دو سرویس dcmtk و dnsmasq نشان می‌دهد که فازینگ ترکیبی در معیار پوشش کد نسبت به فازینگ سنتی عملکرد بهتری دارد. پوشش شاخه در سرویس dcmtk مقدار ۲,۷۱ درصد نسبت به AFLNet بهبود داشته است که توانسته عملکرد منفی NyxNet نسبت به AFLNet را مثبت نماید. همچنین پوشش شاخه در سرویس dnsmasq نسبت به AFLNet مقدار ۳۷,۷۲ درصد و نسبت به NyxNet مقدار ۱۱,۸۲ درصد بهبود داشته است.

کلیدواژه‌گان: آزمون فازینگ، آزمون پروتکل‌های شبکه، آسیب‌پذیری، اجرای نمادین، اجرای پویا-نمادین.

* رایانامه نویسنده مسؤول: haghghi@sadjad.ac.ir

۱- مقدمه

موجود است.

۲- مفاهیم پایه

۲-۱- تست فایزینگ

در فایزینگ، داده‌های غیرمنتظره و تصادفی و گاه ساختارمند، به برنامه‌ها ارسال می‌شود. این داده‌ها می‌توانند شامل مقادیر نامعتبر، خارج از محدوده و یا الگوهای غیرمنتظره‌ای باشند که هدف از ارسال آن‌ها، بررسی واکنش برنامه در برابر ورودی‌های غیراستاندارد است. اگر برنامه به درستی کار نکند به عنوان مثال خراب شود یا به اشتباه عمل کند، نشان می‌دهد مشکلی^۳ وجود دارد. فایزینگ برای پیدا کردن مشکلات امنیتی مانند اجرای کد دلخواه، مشکلات مرتبط با حافظه و انکار سرویس مفید است.

۲-۲- اجرای نمادین^۴

اجرای نمادین تلاش می‌کند به طور سیستماتیک با بررسی یک برنامه به صورت نمادین، مسیرهای مختلف اجرای برنامه را شناسایی کند. این روش به طور کلی با تغییر مقادیر شرطی، حالت‌های مختلف اجرا را بررسی می‌کند و برای هر حالت اجرا، مسیر مجزایی ایجاد می‌شود. سپس حل‌کننده‌های SMT^۵ برای تعیین امکان‌پذیری مسیرها و انتخاب راه‌حل‌ها استفاده می‌شوند. در صورت امکان پذیر بودن یک مسیر، خروجی SMT منجر به تولید ورودی‌های متفاوت خواهد شد. یکی از مهم‌ترین مشکلات روش اجرای نمادین، انفجار مسیرها^۶ است که منجر به کاهش کارایی الگوریتم برای استفاده روی برنامه‌های بزرگ و واقعی می‌شود. افزایش تصاعدی تعداد مسیرهای اجرا با افزایش شاخه‌ها^۷ در کد، پیچیدگی فضای جستجو را به شدت افزایش داده و مانع از به کارگیری این روش بر روی برنامه‌های پیچیده واقعی می‌گردد.

۲-۳- اجرای پویا-نمادین^۸

در اجرای پویا-نمادین، برنامه تحت آزمون به طور همزمان هم با داده‌های واقعی و هم با داده‌های نمادین اجرا می‌شود و سعی بر مرتفع نمودن مشکلات اجرای نمادین دارد. داده‌های واقعی برای اجرای مسیر خاصی از برنامه و داده‌های نمادین برای ثبت رفتار برنامه استفاده می‌شوند. مراحل اجرای روش پیشنهادی پویا-نمادین به شرح زیر است که بخشی از چگونگی پیاده‌سازی این پژوهش را

فایزینگ به عنوان یکی از مهمترین تکنیک‌های آزمون امنیت نرم‌افزار، نقش بسزایی در شناسایی آسیب‌پذیری‌های امنیتی دارد. با توجه به بررسی‌های انجام شده طی ده سال گذشته [1]، روش فایزینگ به طور قابل توجهی نسبت به روش‌های دیگر همچون تحلیل ایستا و یا استفاده از روش تحلیل نمادین به تنهایی، مورد توجه قرار گرفته است. همچنین ترکیب روش فایزینگ با سایر روش‌ها، خروجی‌های بسیار بهبودیافته و قابل توجهی داشته است که می‌توان گفت استفاده از چند تکنیک، ضعف‌های سایر تکنیک‌ها را پوشش می‌دهد.

از آنجا که پروتکل‌های شبکه پایه و اساس ارتباطات شبکه‌ای هستند، فایزینگ آنها از اهمیت ویژه‌ای برخوردار است. با این حال، فایزینگ پروتکل‌های شبکه به دلیل ماهیت پیچیده آنها، همواره یک چالش بزرگ محسوب می‌شده است. تاکنون پژوهش‌های اندکی در زمینه فایزینگ پروتکل‌های شبکه صورت گرفته و عمدتاً محدود به بهبود حفظ وضعیت پروتکل و نیز نحوه ایجاد داده تست به صورت ساختارمند بوده‌اند. از طرفی، تحقیقات انجام شده در زمینه فایزینگ ترکیبی که یکی از روش‌های افزایش معیار پوشش کد می‌باشد، بیشتر معطوف به بهبود فایزینگ تجزیه‌گر فایل^۱ بوده و کمتر به پروتکل‌های شبکه پرداخته‌اند. فایزینگ تجزیه‌گر فایل‌ها یکی از انواع فایزینگ است که در آن فایل‌های ورودی برنامه‌های کاربردی مورد هدف تست قرار می‌گیرند. این فایل‌ها معمولاً شامل فایل‌های پیکربندی، فایل‌های ورودی XML، فایل‌های ورودی کاربر و فایل‌های داده‌ای هستند. با توجه به اهمیت فایزینگ پروتکل‌های شبکه و کمبود تحقیقات در این حوزه، در این مقاله سعی شده است با بررسی راهکارهای موجود و نقاط ضعف آنها، یک معماری فایزینگ ترکیبی متناسب با پروتکل‌های شبکه دارای حالت ارائه شود و گامی در جهت بهبود فرآیند فایزینگ برداشته شود. مهمترین چالش‌های بررسی شده در فرآیند فایزینگ این پروتکل‌ها شامل حفظ وضعیت برنامه‌های چندنخی^۲ در تحلیل پویا-نمادین و همچنین تحلیل نمادین از نقطه دلخواه در طی فرآیند اجرا و پردازش بسته‌ها می‌باشد. همچنین برخی از راهکارها در جهت افزایش سرعت اجرا و همچنین کاهش بار پردازشی در بخش‌های مختلف خواهد بود. پس از معرفی معماری پیشنهادی، نتایج حاصل از پیاده‌سازی و ارزیابی آن ارائه می‌شود که بیانگر برتری روش پیشنهادی نسبت به فازهای

⁵ Satisfiability Modulo Theories

⁶ Path Explosion

⁷ Branch

⁸ Concolic Execution

¹ File Parser

² Multi-thread

³ Bug

⁴ Symbolic Execution

نیز تشریح می‌نماید:
 ۱. یک داده ورودی واقعی برای بررسی مسیر اجرایی برنامه تحت آزمون انتخاب می‌شود.

۲. برنامه با این داده‌های ورودی واقعی اجرا می‌شود و همزمان داده‌ها و عملیات‌های مرتبط با داده ورودی به صورت نمادین ثبت می‌شوند.
۳. یک مدل نمادین از رفتار برنامه با استفاده از اطلاعات ثبت شده ساخته می‌شود.
۴. با استفاده از تحلیل نمادین، شرایطی که منجر به رسیدن به مسیرهای اجرایی جدید می‌شوند، استخراج می‌گردد.
۵. داده‌های ورودی واقعی جدیدی که این شرایط را برآورده می‌کنند، تولید می‌شوند.

۴-۲- فازینگ ترکیبی^۱

فازینگ ترکیبی با ترکیب دو تکنیک فازینگ و اجرای پویا-نمادین تلاش می‌کند تا پوشش کد^۲ و کارایی فرآیند فازینگ را افزایش دهد. شکل ۱ معماری یک سیستم آزمون فازینگ ترکیبی را نشان می‌دهد. این سیستم شامل سه بخش اصلی است:

۱. فازر
۲. موتور اجرای پویا-نمادین
۳. هماهنگ‌کننده

بخش هماهنگ‌کننده یک نرم‌افزار میانی است که تکنیک‌های فازینگ و اجرای پویا-نمادین را کنترل می‌کند و سه وظیفه اصلی دارد. ابتدا، فازر را نظارت می‌کند تا بتواند زمان اجرای پویا-نمادین را تشخیص دهد. دوم، محیط در حال اجرا را برای اجرای پویا-

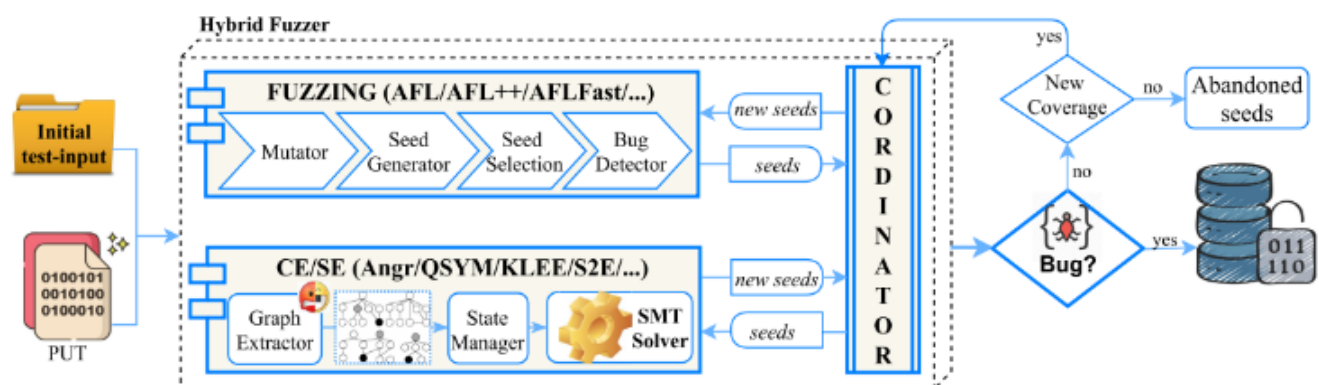
۲-۵- پروتکل‌های شبکه

پروتکل‌های شبکه به دو دسته دارای حالت^۳ و بدون حالت^۴ تقسیم می‌شوند. در پروتکل‌های بدون حالت مانند HTTP، اطلاعات مربوط به حالت ارتباط ذخیره نمی‌شود و هر بسته به طور مستقل ارسال و پردازش می‌گردد. اما در پروتکل‌های دارای حالت مانند TLS، اطلاعاتی در مورد حالت اتصال ذخیره می‌شود که هر بسته می‌تواند به بسته‌های قبلی وابستگی داشته باشد و نحوه پردازش هر بسته ارسالی متفاوت باشد.

۳- بررسی مطالعات پیشین

بزرگترین چالش در فازینگ پروتکل‌های شبکه دارای حالت، سرعت اجرا و حفظ حالت براساس ماشین حالت پروتکل است. اخیراً راه‌حل‌های متنوعی مطرح شده‌اند که تا حد زیادی چالش‌های اولیه در فازینگ این نوع محصولات نرم‌افزاری را برطرف کرده‌اند. حال گام بعدی در تکامل و پیشرفت این حوزه استفاده از تکنیک‌های بهبوددهنده و مکمل مانند اجرای پویا-نمادین در طول فرآیند آزمون فازینگ برای تولید داده‌های آزمون با کیفیت‌تر است.

این پژوهش، اولین پیاده‌سازی اجرای پویا-نمادین بر روی اهداف سرویس‌های شبکه دارای حالت، بر پایه تحقیقات و کارهای انجام شده تاکنون خواهد بود که دسته‌بندی آن‌ها بر روی سه موضوع اصلی زیر انجام شده است:



شکل ۱. ساختار کلی یک فازر ترکیبی [3]

³ Statefull
⁴ Stateless

¹ Hybrid Fuzzing
² Code Coverage

۳-۱- فازینگ پروتکل‌های شبکه دارای حالت

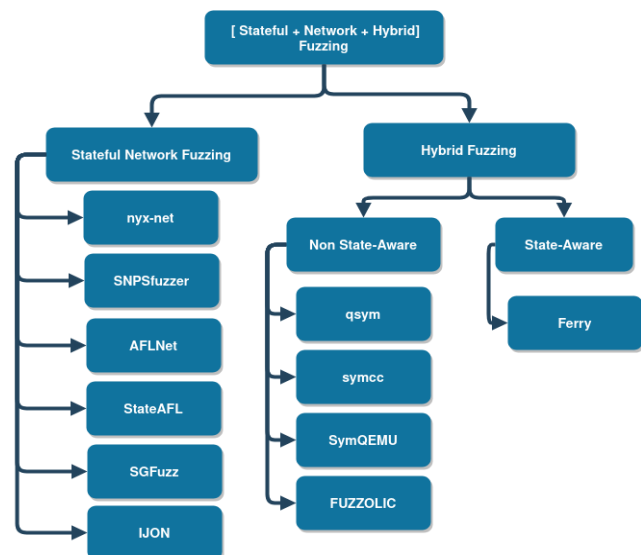
پس از موفقیت فازر AFL، برخی از محققین سعی کرده‌اند تا آن را برای سرویس‌های شبکه نیز به کار گیرند. بدین منظور سعی شده اهداف تحت آزمون را تغییر دهند تا به جای دریافت ورودی از سوکت شبکه، از فایل‌های ورودی دریافت کنند. اما این تغییرات ذات شبکه‌ای بودن اهداف را نقض کرده و برخی مشکلات را از دید فازر پنهان می‌کند. فازر AFLNet [5] سعی کرده است روش را معکوس کند؛ یعنی به جای تغییر اهداف تحت آزمون، خود فازر AFL را تغییر دهد تا متناسب با اهداف شبکه‌ای باشد. در بررسی کارایی طرح پیشنهادی این پژوهش AFLNet به عنوان State-of-the-art و پایه مقایسه‌ها در نظر گرفته می‌شود.

AFLNet از ترافیک ضبط‌شده بین سرویس‌دهنده و سرویس‌گیرنده به عنوان داده‌های اولیه استفاده می‌کند. به عنوان سرویس‌گیرنده عمل کرده و درخواست‌های جهش‌یافته را ارسال می‌کند. از کدهای پاسخ سرویس‌دهنده برای استنتاج ماشین حالت پروتکل و هدایت فازر استفاده می‌کند. در مقایسه با AFL برای پروتکل‌های پیچیده و دارای حالت موثرتر است. اما همچنان معایبی مانند نیاز به ترافیک اولیه، محدود بودن به پروتکل‌های خاص، عدم حفظ وضعیت و نیازمند اسکریپت cleanup و اجرای قبل از انجام هر تست، عملکرد ضعیف و همچنین نیاز به تجزیه‌گرهای سفارشی دارد.

فازر NYX [6] با استفاده از QEMU و KVM با ماشین مجازی یکپارچه شده و حالت آن را به یک پشتیبان‌گیری مقطعی^۱ تنظیم می‌کند. این فازر از یک عامل نماینده^۲ در ماشین مجازی برای کنترل چرخه فازینگ استفاده می‌کند. تکنیک پشتیبان‌گیری مقطعی در این فازر بر پایه Dirty-Page است که از امکانات سخت‌افزاری پردازنده‌های مدرن برای ردیابی صفحات تغییر یافته استفاده می‌کند. نسخه توسعه‌یافته این فازر به نام NyxNet [7] است که از قابلیت‌های آن برای فازینگ پروتکل‌های شبکه استفاده می‌کند. معماری این فازر در شکل ۳ قابل مشاهده است. پشتیبان‌گیری مقطعی افزایشی^۳ منجر می‌شود تا تنها صفحات تغییر یافته ذخیره و بازیابی شوند و سرعت را افزایش دهد. یکی دیگر از قابلیت‌های بسیار مفید فازر NyxNet هماهنگی کامل با زیرساخت و ساختار فازر AFL می‌باشد. همچنین اهدافی که برای این نوع فازر کامپایل شده‌اند، نیز به راحتی توسط NyxNet قابل استفاده خواهند بود و پیچیدگی و ابهاماتی در یادگیری و تسلط بر روی این فازر به

- فازینگ پروتکل‌های شبکه دارای حالت: فازرهای مرتبط لیست شده‌اند تا بهترین تحقیقات انجام شده به عنوان فازر پایه انتخاب شود.
- فازینگ ترکیبی بر روی اهداف دارای حالت: مقالاتی که فازینگ ترکیبی را بر روی تمام جریان پردازش و ارتباط بین حالت‌های تعریف شده درون برنامه انجام می‌دهند.
- فازینگ ترکیبی بر روی اهداف بدون حالت: مقالاتی که فازرهای ترکیبی از آن‌ها استفاده کرده‌اند و یا به عنوان بخشی از فازرهای ترکیبی می‌باشند که بر روی اهداف بدون حالت و یا با در نظر نگرفتن حالت‌های گذشته عمل می‌کنند.

با بهره‌گیری از دانش و تجربیات به دست آمده از تحقیقات پیشین، هدف انتخاب درست تکنیک‌ها و ابزارهای مورد استفاده در این پژوهش است. دسته‌بندی مقالات مطالعه شده در شکل ۲ قابل مشاهده است.



شکل ۲. دسته‌بندی کارهای پیشین

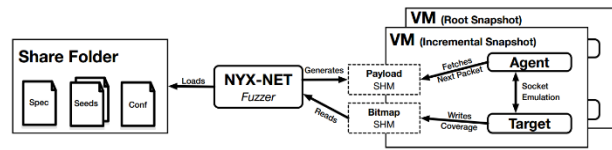
در بخش فازینگ ترکیبی آگاه به حالت بر روی اهداف دارای حالت تنها یک مقاله [4] وجود دارد که تحقیقات آن بر روی تجزیه‌گر فایل‌ها انجام شده است که نشان‌دهنده جدید بودن این موضوع و عدم تحقیقات کافی در این زمینه است. برای پیاده‌سازی فازینگ ترکیبی آگاه به حالت بر روی پروتکل‌های شبکه دارای حالت، ابتدا باید زیرساخت لازم برای فازینگ ترکیبی فراهم شود. بنابراین پیاده‌سازی فازینگ ترکیبی آگاه به حالت بر روی اهدافی که دارای حالت هستند، خارج از حوزه تحقیقاتی این پژوهش می‌باشد.

³ Incremental Snapshot

¹ Snapshot

² Agent

حداقل‌ترین حالت خود خواهد رسید.



شکل ۳. معماری NyxNet [7]

و عدم نیاز به دانش قبلی در مورد حالت‌های برنامه اشاره کرد. اما معایبی همچون کاهش سرعت اجرا، خطا در شناسایی متغیرها و عدم پوشش فضای حالت‌های ضمنی مانند تغییرات پایگاه داده‌ها، برای این روش وجود دارد. همچنین برای برنامه‌های با فضای حالتی بسیار بزرگ، کارایی کمتری خواهد داشت.

مقاله IJON [11] رویکرد پیشنهادی با الهام از مدل Human-in-the-loop ارائه می‌نماید که امکان دخالت کاربر در خروجی یک فرایند یا رویداد را فراهم می‌سازد. نتایج تحقیق نشان می‌دهد یک تحلیلگر انسانی می‌تواند فازر را در فضای حالتی برنامه هدایت کرده و بخش‌هایی از فضای حالتی را که نیاز به بررسی کامل دارند، مشخص نماید. با استفاده از قابلیت نشانه‌گذاری، می‌توان اطلاعات بیشتری در مورد وضعیت متغیرها و حالت‌های برنامه در اختیار فازر قرار داد که منجر به افزایش پوشش کد و تولید داده‌های آزمون باکیفیت‌تر می‌شود. اما مهم‌ترین محدودیت این روش، نیاز به نیروی انسانی متخصص و آشنا با کد منبع هدف است که نیازمند صرف زمان زیادی برای آشنایی با پیچیدگی‌های برنامه و معماری پیاده‌سازی آن است.

۳-۲- فازینگ ترکیبی

ابزار Angr [12] یک موتور اجرای نمادین کلاسیک است که از زبان میانی VEX استفاده می‌کند که باعث کاهش سرعت آن می‌شود، اما این کار باعث ساده شدن پیاده‌سازی آن شده است. معماری آن در شکل ۴ قابل مشاهده است. با توجه به اینکه Angr بر پایه VEX است، بنابراین می‌تواند تمام معماری‌هایی که VEX پشتیبانی می‌کند را به ارث برده و پشتیبانی نماید. از آنجایی که کد Angr در پایتون نوشته شده است، سرعت آن نسبتاً کند است.

ابزار S2E [13] یک سیستم‌عامل کامل را درون QEMU شبیه‌سازی می‌کند و آن را به موتور اجرای نمادین KLEE متصل می‌کند. این کار باعث می‌شود S2E بتواند تجزیه و تحلیل جامعی از کد در تمامی لایه‌های سیستم‌عامل ارائه کند. اما پیچیدگی زیاد آن باعث شده استفاده از آن دشوار باشد.

در دو راهکار S2E و Angr به عنوان state-of-the-art، استفاده از یک زبان میانی در جهت کمتر شدن پیچیدگی پیاده‌سازی انتخاب شده است اما با کم شدن هزینه پیاده‌سازی، عملکرد و کارایی به شدت کاهش یافته است.

فازر SNPSFuzzer [8] راهکاری مبتنی بر پشتیبان‌گیری مقطعی برای تسریع فازینگ پروتکل‌های شبکه پیشنهاد می‌کند. این فازر با ذخیره و بازیابی اطلاعات مربوط به هر حالت، از تکرار ارسال پیشوندهای طولانی جلوگیری می‌کند. همچنین با استفاده از الگوریتم‌های تحلیل نقطه مقطع و زنجیره پیام، پوشش حالت‌ها را افزایش می‌دهد. اما پیچیدگی پیاده‌سازی، نیاز به تنظیم الگوریتم‌ها و کندی نسبی از محدودیت‌های این روش به شمار می‌روند. این فازر برای مدیریت پشتیبان‌گیری‌های مقطعی از CRIU استفاده می‌کند. کارایی CRIU را می‌توان با قابلیت مشابه در Qemu مقایسه کرد که جنبه‌های عملکردی مهم در فازینگ مانند کارایی و سرعت در آن مورد توجه نبوده است.

فازر StateAFL [9] یک فازر جعبه خاکستری برای سرورهای شبکه‌ای است که بدون نیاز به سفارشی‌سازی دستی، صرفاً با تجزیه و تحلیل ساده برنامه تحت آزمون کار می‌کند. این فازر کدهایی را درون برنامه هدف تزریق می‌کند تا اطلاعات حافظه و شبکه جمع‌آوری کند و پس از هر تعامل درخواست-پاسخ، یک پشتیبان‌گیری مقطعی از حافظه بلندمدت گرفته می‌شود. سپس هر پشتیبان‌گیری به یک شناسه حالت یکتا نگاشت می‌شود. این روش نیازی به تجزیه‌گرهای سفارشی پروتکل ندارد. نتایج نشان می‌دهد StateAFL بدون سفارشی‌سازی، می‌تواند پوشش و خرابی‌های قابل مقایسه و حتی بهتر از فازرهای سفارشی قبلی به دست آورد. اما محدودیت‌هایی مانند نیاز به دسترسی به کد منبع و افزایش زمان اجرا وجود دارد.

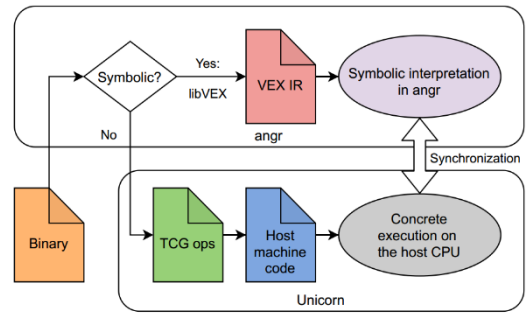
مقاله SGFUZZ [10]، روشی نوین بر پایه Libfuzzer برای فازینگ برنامه‌های دارای حالت همچون پروتکل‌های ارتباطی ارائه می‌دهد. راه حل پیشنهادی، شناسایی متغیرهای حالت برنامه با استفاده از الگوییابی و پیگیری مقادیر آنها حین اجرای برنامه است.

الگوی استفاده شده بر این فرض استوار است که اغلب حالت‌های برنامه به صورت شمارشی ذخیره^۱ و در دستورات شرطی بررسی می‌شوند. از مزایای این روش می‌توان به کشف خودکار فضای حالتی

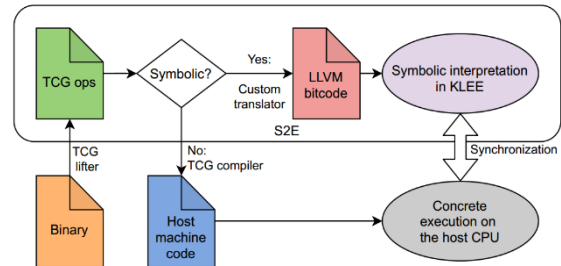
^۱ Enum

فرآیند کامپایل منجر به شکست گردد و همچنین دسترسی به کد منبع اهداف تحت آزمون به عنوان یک الزام مطرح می‌شود.

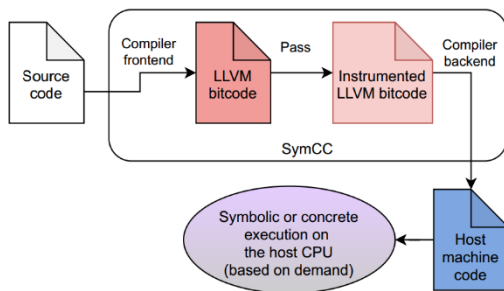
ابزار SymQEMU [16] تلاش نموده است تا تمامی مزایا موجود در تحقیقات انجام شده تاکنون را در کنار هم قرار دهد. از سرعت و کارایی symcc و همچنین از Qemu و TCG که زبان میانی آن است استفاده نموده تا قابلیت پشتیبانی معماری مختلف را بدست آورد. این ابزار به عنوان شبیه‌سازی سطح کاربر می‌باشد و عملکردی مشابه به Angr و SymCC/QSym خواهد داشت. همچنین راه حلی را پیشنهاد می‌دهد تا بتوانند ایده اصلی SymCC که برای مراحل و فرآیند کامپایل کد منبع بوده است را بر روی باینری اعمال کنند. ابزار SymQemu فقط بر روی معماری x86 پیاده‌سازی شده است. همچنین این پیاده‌سازی بر روی فایزینگ فایل تمرکز داشته است. نمی‌توان بصورت سفارشی از آن برای تجزیه و تحلیل در سطح توابع استفاده کرد و یا به صورت سفارشی شده بخش‌های تعیین شده‌ای در برنامه تحت آزمون مورد بررسی و اجرای نمادین قرار گیرد.



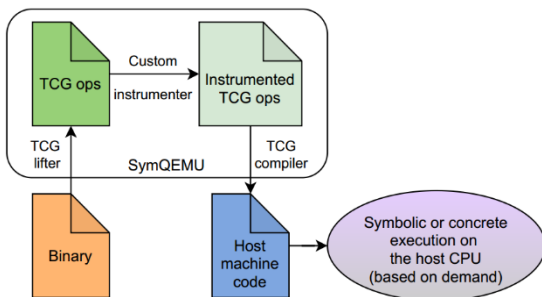
شکل ۴. معماری Angr [16]



شکل ۵. معماری S2E [16]

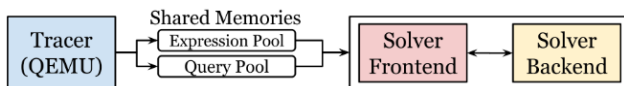


شکل ۷. معماری SymCC [16]



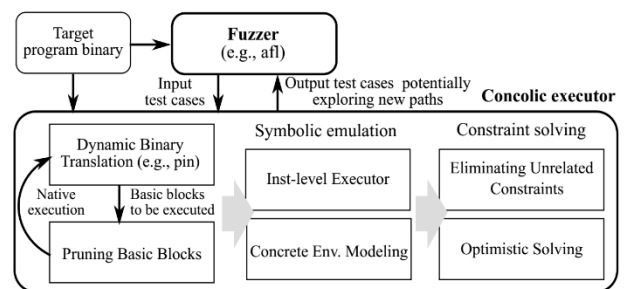
شکل ۸. معماری SymQemu [16]

ابزار Fuzzolic [17] همانند SymQEMU از زبان میانی TCG استفاده می‌کند، اما با جدا کردن بخش حل‌کننده، توانسته است عملکرد بهتری داشته باشد. Fuzzolic سه حالت مختلف تجزیه و تحلیل پویا دارد که باعث بهبود دقت آن شده است.



شکل ۹. معماری Fuzzolic [17]

ابزار QSym [14] بدون استفاده از زبان میانی و با Instrument کردن فایل باینری برنامه تحت آزمون در زمان اجرا، توانسته سرعت زیادی کسب کند. اما این روش محدود به معماری x86 است و پیاده‌سازی آن برای سایر معماری‌ها بسیار دشوار خواهد بود. ذات QSym برای فایزینگ فایل می‌باشد و هنگام تجزیه و تحلیل یک فایل ۲۱۸ بیتی عکس png در برنامه libpng برای اجرای تمام شاخه‌ها به بیش از دو ساعت زمان نیاز دارد. بزرگترین عیب این روش عدم سفارشی‌سازی و محدود نمودن شاخه‌های مورد بررسی می‌باشد.



شکل ۶. معماری QSym [14]

ابزار SymCC [15] به عنوان یک لفاف بر روی کامپایلر برپایه QSym ایجاد شده است که با افزودن کد مورد نیاز هنگام کامپایل، توانسته است سرعت اجرای خود را تا هزار برابر نسبت به سایر راهکارها افزایش دهد. اما نیازمند یک کامپایلر خاص است و قابلیت کار با باینری‌های موجود را ندارد. براساس مشاهدات تجربی ممکن است این کامپایلر خاص در برخی اهداف قابل استفاده نباشد و

چالش ۳، برخی ابزارهای بررسی شده بخش 3 در سطح باینری کار کرده و فرایند اجرا را در کتابخانه‌هایی مانند glibc نیز درگیر می‌کنند. این موضوع می‌تواند بر سرعت تجزیه و تحلیل و اجرای نمادین تأثیر منفی بگذارد. استفاده از توابع نمادین^۱ می‌تواند با کاهش حجم محاسبات، سرعت تحلیل و اجرای این ابزارها را افزایش دهد.

با توجه به معایب مطرح شده در بخش ۳ و همچنین سه چالش اصلی قبلاً ذکر شده در این بخش، در این پژوهش سعی بر آن است تا راهکاری جامع و منعطف برای اهداف سرویس‌های شبکه و پروتکل‌های دارای حالت ارائه شود.

به منظور مقابله با انفجار مسیر، بهینه‌سازی‌ها و روش‌هایی مانند محدود کردن عمق تجزیه و تحلیل، کاهش شرایط پیچیده و استفاده از ترکیب‌های تحلیل نمادین و آزمون‌های واقعی می‌توانند مورد استفاده قرار گیرند که به عبارت دیگر اجرای Concrete به نوعی مکمل تحلیل نمادین است به توان عیوب آن را مرتفع و از فواید آن استفاده نماییم. بدین منظور در این پژوهش با استفاده از آزمون‌های واقعی به همراه تحلیل نمادین که به آن تحلیل پویا-نمادین می‌گوییم، سعی شده است تا بتوان با استفاده از مزایای هر دو روش بهبود در عملکرد ایجاد داده ای آزمون جدید برای فازر NyxNet انجام شود.

بسیاری از پروتکل‌ها شامل مفاهیم سرویس دهنده-سرویس گیرنده هستند که در آن سرویس گیرنده اتصال را آغاز می‌کند و سرویس دهنده پس از درخواست سرویس گیرنده پاسخ می‌دهد. در مرحله آزمون سرویس دهنده، عملکرد به نسبت آسان است. نرم‌افزار هدف در انتظار دریافت ارتباطات ورودی قرار دارد و فازر تنها نیاز دارد که به سرویس دهنده هدف متصل شود و داده‌های آزمون را ارسال کند. اما آزمون سرویس گیرنده به طور معمول پیچیده‌تر است. در این حالت، فازر باید مانند سرویس دهنده عمل کند و منتظر درخواست‌های ورودی از سوی سرویس گیرنده باشد. هر زمان که سرویس گیرنده هدف به فازر متصل می‌شود، فازر با ارسال یک داده آزمون به درخواست سرویس گیرنده پاسخ می‌دهد. به طور معمول، نیاز است که در سرویس گیرنده تغییراتی ایجاد شود تا به صورت مکرر به فازر متصل شود و بتواند دریافت داده‌های آزمون را ادامه دهد. چنین مواردی با ایجاد تغییر در چند خط کد منبع سرویس گیرنده با استفاده از وصله‌ها^۲ مرتفع شده است.

با توجه به اهداف تحت آزمون، اعداد تصادفی برای مواردی مانند

با توجه به پیاده‌سازی مشابه Fuzzolic و SymQemu هر دو دارای معایب مشترکی می‌باشند.

اخیراً در کاربردهای مختلفی این نوع آزمون‌ها مورد استفاده بوده‌اند از جمله مرجع [۲۰]. همچنین ابزارها و روش‌های مختلفی بر مبنای آن توسعه یافته‌اند [۲۱] و [۲۲].

همچنین، روش‌هایی نیز برای توسعه نرم‌افزار مبتنی بر آزمون معرفی شده‌اند که روش پیشنهادی در این تحقیق نیز می‌تواند در آنها مفید باشد [۲۳].

۴- طرح پیشنهادی و پیاده‌سازی

با توجه به بررسی‌های انجام شده در بخش 3.1، ابزار NyxNet به عنوان بهترین گزینه برای استفاده به عنوان فازر پایه در جهت توسعه و بهبود ابزار پیشنهادی انتخاب می‌شود. این ابزار علاوه بر سرعت بالای اجرا و حفظ حالت در پروتکل‌های شبکه‌ای حالت‌دار، از برنامه‌های تحت تست باینری و نیز دارای کد منبع باز می‌باشد. علاوه بر معایب هر کدام از روش‌های مورد بررسی در بخش ۳، در ادامه ۳ چالش مطرح می‌شود.

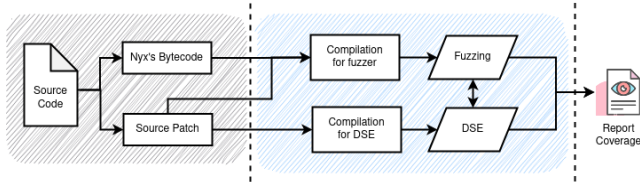
چالش ۱، همان‌طور که در بخش 3.2 بررسی شد، روش‌های موجود اجرای برنامه‌ها و توابع چندنخی را بصورت نمادین پشتیبانی نمی‌کنند. علت این محدودیت این است که اهداف این روش‌ها، تجزیه گره‌های فایل بوده و به طور ضمنی تنها برنامه‌های تک‌نخی را پشتیبانی می‌کنند. در حالی که اکثر برنامه‌ها و سرویس‌های شبکه‌ای به صورت چندنخی پیاده‌سازی می‌شوند تا بتوانند با سرعت بیشتری به درخواست‌ها پاسخ دهند. این محدودیت باعث می‌شود نتوان از نتایج کارهای گذشته در حوزه فازینگ ترکیبی برای اهداف شبکه‌ای چندنخی استفاده کرد. عدم پشتیبانی برنامه‌های چندنخی یکی از مهم‌ترین نقاط ضعف مشترک تمام این روش‌ها به‌شمار می‌رود.

چالش ۲، برنامه‌ها و سرویس‌های شبکه‌ای در طول اجرای خود و پردازش داده‌های یک ارتباط، چندین بسته دریافت کرده و هر بسته بر حالت برنامه تأثیر می‌گذارد. بنابراین در فازینگ ترکیبی نیاز است تا بتوان بخش‌های مختلف برنامه تحت تست را به صورت سفارشی و نمادین اجرا و بررسی کرد. به عنوان مثال، اگر داده تست شامل هفت بسته باشد، بخواهیم فقط روند اجرای بسته پنجم را به صورت نمادین بررسی کنیم. این امکان توسط روش‌های گذشته فراهم نبوده است.

² Patch

¹ Semantic Functions

سرویس^۲، اما بایت کدهای مختص به NYX و همچنین مفسرهای این بایت کدها باید توسط کاربر فراهم شده باشد تا بتوانند پکت‌های دریافت شده در طی فازینگ را تفسیر و به داخل برنامه تزریق نمایند. برای فرآیند فازینگ دو بخش بایت کدها و وصله‌ها نیاز است اما برای فرآیند اجرای پویا-نمادین تنها وصله‌ها لازم خواهد بود تا تغییرات لازمه را در فرآیند اجرای برنامه ایجاد نماید.



شکل ۱۰. طرح کلی پیاده‌سازی

اهداف تحت آزمون آماده شده برای اجرای فازینگ در هنگام کامپایل با استفاده از کامپایلر پیش‌فرض afl-clang-fast توسط Sanitizerها تغییر خواهند کرد تا به عنوان Oracle بتوانند مشکلات و خطاهای حافظه را تشخیص دهند. در صورت نیاز می‌توان پارامترهای کامپایل را تغییر داد و Sanitizerهای دیگری را به عنوان Oracle انتخاب نمود تا انواع دیگری از باگ‌ها و آسیب‌پذیری‌ها را با استفاده از فازینگ کشف نماییم. اما با توجه به اینکه Sanitizerها کدهایی را به درون برنامه تزریق می‌نمایند، در جهت عدم پردازش این کدها در تحلیل نمادین نیاز خواهیم داشت تا اهداف تحت آزمون را بدون چنین ویژگی‌هایی کامپایل نماییم. بنابراین دو روش کامپایل خواهیم داشت و فایل باینری مورد استفاده توسط فازر و تحلیل نمادین متفاوت خواهند بود. در ادامه دو فرآیند فازینگ و تحلیل پویا-نمادین همزمان اجرا شده و داده‌های خود را با یکدیگر به اشتراک خواهند گذاشت.

در ادامه شبه کد و منطق پیاده‌سازی انجام شده آورده شده است:

```
def coordinator():
    set_corpuses_notify()
    list_corpuses()
    while True:
        py_script = reproduce_corpus(
            get_one_corpus()
        )
        packets = load_packets(py_script)
        new_seeds = explore(packets)
        feed_to_nyxnet(new_seeds)

def explore():
    run_CE()
    handle_instructions()
    return solve_path_constraints()
```

رمزنگاری نقش بسیار مهمی را ایفا می‌کند. گاهی تولید اعداد تصادفی می‌تواند برنامه تحت آزمون را در وضعیتی نامعین قرار دهد و همچنین خطاهای برنامه با استفاده از داده‌های آزمون، می‌بایست قابل بازتولید باشند. بسیاری از محصولات نرم‌افزاری که بر روی بستر شبکه کار می‌کنند، با توجه به تأخیرات در انتقال داده‌ها از توابعی مانند sleep(), usleep() استفاده می‌کنند تا وقفه‌های مصنوعی ایجاد کنند. اما این وقفه‌ها می‌توانند عملکرد فازر را به شدت کاهش دهند. برای به حداقل رساندن تاثیر وقفه‌ها و تأخیرها، همچنین مواردی اعداد تصادفی که منجر به ایجاد مسیرهای متفاوت در هر بار اجرای برنامه تحت آزمون می‌شود، سعی شده است که از قابلیت‌های DBI^۱ به طور کامل بهره‌برداری شود تا تمامی فرآیندها و تغییرات مورد نیاز به صورت خودکار انجام شود.

ابزارهای DBI به وسیله ترجمه باینری در زمان اجرا، توانایی اعمال تغییرات در اجرای یک برنامه را داراست. به عنوان مثال، این قابلیت را دارد که بدون دسترسی به کد منبع، کدهای نظارتی را در نقاط مختلف درون باینری تزریق کند تا اطلاعاتی در مورد جریان اجرا جمع‌آوری کند. همچنین قادر است خطاها را شناسایی و اصلاح نماید یا فرآیند اجرا را با تغییرات مورد نیاز سازگار کند. با این حال، یکی از مشکلات اصلی این روش، سربار اجرایی بالای آن است. زیرا ترجمه باینری در زمان اجرا هزینه‌های زیادی دارد و باعث کندی در اجرای برنامه می‌شود. به همین دلیل، برای مواردی مانند آزمون فازینگ که نیاز به اجرای سریع و مکرر برنامه دارند، این روش چندان مناسب نمی‌باشد. خوشبختانه، Dynamorio [18] با استفاده از مکانیزم‌های مختلفی که پیاده‌سازی نموده است، توانسته است سربار اجرایی را به حداقل ممکن، به حدی که تنها ۱/۱ برابر سرعت اجرای اصلی باشد، کاهش دهد.

طرح کلی و تغییراتی که این پژوهش در فرآیند فازینگ NyxNet ایجاد خواهد کرد در شکل ۱۰ قابل مشاهده است. با توجه به شکل ۱۰ برای کامپایل و آماده‌سازی اهداف آزمون دو روش کامپایل نیاز است.

استفاده از Patch اجباری نمی‌باشد و می‌تواند هدف آزمون بدون تغییرات در کد منبع مناسب برای فازینگ باشد. گاهی نیاز است که برخی تغییرات بسیار کمی در کد منبع ایجاد شود تا وضعیت پایداری در فرآیند اجرا داشته باشیم. برخی از این تغییرات گاهی توسط خود اهداف آزمون به صورت پارامترهای اجرایی در اختیار کاربر قرار می‌گیرند، به عنوان مثال اجرا نشدن برنامه تحت آزمون به صورت

² Daemon

¹ Dynamic Binary Instrumentation

عبارت دیگر، پکت‌ها توسط SE انجام خواهد شد و در صورت نیاز به دریافت پکت، CE می‌تواند آن را از طریق حافظه مشترک از SE درخواست و دریافت نماید.

در پیاده‌سازی فعلی تحلیل نمادین تنها بر روی آخرین پکت انجام می‌شود. در صورت دریافت آخرین بسته، حافظه مورد استفاده بصورت Taint شده خواهد بود. بدین ترتیب چالش ۲ مطرح شده در این بخش مرتفع خواهد شد و در صورت نیاز می‌توان برای هر بسته دریافتی شرایط و پردازش خاصی را مدنظر داشته باشیم که این عمل می‌تواند برای داشتن زیرساخت تحلیل نمادین آگاه به حالت^۳ بسیار کارآمد باشد.

در تابع `process_instruction` بررسی می‌شود اگر فراخوانی تابعی بر روی یکی از توابع نمادین شده به عنوان مثال `stcmp` باشد به جای پردازش دستورالعمل‌ها توابع نمادین اجرا خواهد شد و همچنین بخش CE ارسال دستورالعمل‌ها را تا خاتمه یافتن تابع نمادین شده برای SE متوقف خواهد کرد. بدین ترتیب چالش ۳ مطرح شده در این بخش مرتفع می‌گردد و بار پردازشی در تحلیل نمادین کاهش و سرعت اجرای CE افزایش خواهد یافت.

پس از پردازش کامل آخرین پکت توسط برنامه تحت آزمون، عملیات تحلیل نمادین و حل نمودن `Path Constraints` انجام می‌شود تا با حل نمودن محدودیت‌های موجود در طی مسیر پردازش آخرین پکت، داده‌های آزمون جدیدی ایجاد شود و به فرآیند فازینگ تزریق گردد تا بتوان پوشش کد را بهبود بخشید.

کدمنبع بخش `Coordinator` و همچنین SE به زبان پایتون نوشته شده است. اما بخش ارسال داده آزمون جدید به `NyxNet` به زبان `Rust` برنامه‌نویسی شده است تا بتوان براساس ساختارهای تعریف شده در این فازر داده‌های جدید را به فرآیند فازینگ تزریق نماییم. همچنین بخش CE به زبان `C/C++` نوشته شده است.

یکی از نکات بسیار مهم که در طول تحقیقات در نظر گرفته شده، این است که اهداف آزمون پروتکل‌های شبکه وضعیت دار هستند. بنابراین، نمی‌توان به همان روشی که در آزمون‌های رایج برای تجزیه گره‌های فایل‌ها استفاده می‌شود، اقدام به درون‌ریزی داده از ابتدا به داخل برنامه کرد و سپس تحلیل را انجام داد. بهمین دلیل با استفاده از قابلیت‌های DBI توابعی مهم که مرتبط با ارتباطات شبکه می‌باشند را `Hook` کرده و متناسب با نیاز تغییرات در فرآیند اجرا در آن ایجاد می‌شود. به عنوان مثال هنگام خواندن اطلاعات از

```
def handle_instructions():
    done = False
    while(not done):
        action = read_shm()
        if action == get_packet:
            packet = get_next_packet()
            if packet.is_last_one():
                taint memory(packet)
                send(packet)
            elif action == set_reg_mem:
                change_reg_mem_value()
            elif action == instruction:
                done = process_instruction()
```

بخش `Coordinator` وظیفه جمع‌آوری فایل‌های `Corpus` ایجاد شده توسط `NyxNet` را بر عهده دارد. سپس این فایل‌ها را بر اساس اندازه آنها مرتب می‌کند، به گونه‌ای که فایل‌های کوچکتر در اولویت بررسی قرار می‌گیرند. این اقدام به منظور بهینه‌سازی مصرف حافظه در فرآیند تحلیل نمادین انجام می‌شود. وظیفه دیگر این بخش دریافت اطلاعات پکت‌ها است تا بتواند آنها را به عنوان ورودی به پروسه اجرای پویا-نمادین ارسال کند.

در بخش SE^۱ همان تابع `explore`، ابتدا پروسه CE^۲ اجرا می‌شود و سپس شروع به ارسال/دریافت داده می‌کند. بخش CE با استفاده از ابزار `DynamoRio` که یک `DBI` می‌باشد، پیاده‌سازی شده است. CE وظیفه اجرا برنامه تحت آزمون و تغییرات لازم در طول فرآیند اجرا مانند کاهش تاثیر وقفه‌ها و اعداد تصادفی را اعمال می‌نماید تا اجرا آسان و سریع شود و به علاوه، همچنان بتواند همگرایی و یکپارچگی برنامه تحت آزمون که یک برنامه شبکه‌ای است، را حفظ کند.

ارتباط میان CE/SE با استفاده از حافظه مشترک `Shared Memory` برقرار خواهد بود و برای رفع مشکلات و تداخلات همزمانی در ارسال و دریافت داده‌ها و همچنین پشتیبانی از برنامه‌های تحت آزمون چندنخی، از `Semaphore` و `Mutex` در پیاده‌سازی CE/SE استفاده شده است. در طی پردازش دستورالعمل‌های دریافت شده توسط SE عمل `Context-Switching` برای برنامه‌های چندنخی با استفاده از اطلاعات دریافت شده از CE انجام خواهد شد، بنابراین چالش ۱ مطرح شده در این بخش بدین صورت مرتفع می‌گردد. اگر پردازش دستورالعمل‌های دریافت شده به داده خاصی از حافظه و یا برخی از رجیسترهای خاص مانند `XMM` و `YMM` نیاز داشته باشد، می‌تواند آن را از CE درخواست و دریافت نماید. همچنین مدیریت ورودی‌ها یا به

³ Satte-Aware

¹ Symbolic Executor

² Concrete Executor

همانطور که اعداد جدول ۱ برای هدف آزمون dcmk نشان می‌دهند، عملکرد فازر NyxNet در برابر AFLNet نسبتاً ضعیف بوده است. دلیل این امر می‌تواند پیچیدگی ساختار برنامه تحت آزمون و همچنین وجود شرایط خاص ورودی در dcmk باشد که منجر به کاهش کارایی فازر شده است و به دلیل تصادفی بودن ایجاد داده آزمون‌ها شرایط قابل پاس شدن نبوده‌اند. با این حال، فایزینگ ترکیبی توانسته است این عملکرد نسبتاً ضعیف را بهبود بخشد و درصد تفاوت در پوشش را مثبت کند. بنابراین می‌توان نتیجه گرفت که فایزینگ ترکیبی تأثیر مثبت خود را در طی فرایند ایفا نموده است.

اما برای هدف آزمون dnsmasq، عملکرد و بهبود هر دو فازر نسبت به AFLNet کاملاً مشهود است. فایزینگ ترکیبی همچنان موجب افزایش پوشش در این هدف آزمون شده است و این بهبود نزدیک به ۳۸ درصد می‌باشد.

شاخص بعدی ارزیابی پوشش کد می‌باشد که بر مبنای تعداد خط کد ارزیابی انجام می‌شود. این مقادیر با ابزارهایی مانند gcov قابل بدست آوردن است. جدول ۲ مقادیر عددی برای این شاخص ارزیابی را نشان می‌دهد.

جدول ۲. مقایسه عددی پوشش خط کد

NyxNet-Hybrid		NyxNet-Balanced		AFLNet	
4526	+1.95	4407	-0.7	4439	dcmk
1753	+28.89	1606	+18	1361	dnsmasq

با توجه به شاخص پوشش شاخه که یکی از معیارهای مهم در ارزیابی کیفیت فرایند آزمون نرم‌افزار می‌باشد، از نظر مثبت و منفی بودن خروجی، مفهوم کلی به دست آمده از جدول ۲ و جدول ۱ شبیه به یکدیگر است، اما از نظر عددی مقادیر دو جدول تفاوت‌هایی دارند. همچنان عملکرد نسبتاً ضعیف فازرها در هدف آزمون dcmk با وجود پیچیدگی‌های خاص این برنامه قابل مشاهده است. با این وجود، فایزینگ ترکیبی توانسته است تعداد ۸۷ خط کد بیشتری را در طی زمان ۳۰ دقیقه پوشش دهد که خود نشان‌دهنده بهبود نسبی در فرایند آزمون فایزینگ ترکیبی می‌باشد. اما این اعداد در هدف آزمون dnsmasq به مراتب بیشتر بوده و فایزینگ ترکیبی توانسته است تا نزدیک به ۴۰۰ خط کد بیشتری را نسبت به سایر فازرها پوشش دهد در شکل ۱۱ و ۱۲ نیز می‌توان نمودار پوشش خط کد هر دو سرویس تحت آزمون را مشاهده کرد.

Socket آن را به عنوان دریافت کننده ورودی تشخیص داده و درخواست دریافت پکت را به SE ارسال خواهد کرد بدین ترتیب اطلاعات با سرعت بیشتری نسبت به سوکت دریافت خواهد شد و بهبود در سرعت اجرا خواهیم داشت.

همچنین در جهت پیشگیری از احتمال انفجار مسیرها ورودی‌های با سایز کوچکتر ابتدا پردازش خواهند شد و همچنین بازه زمانی حداکثر دو دقیقه برای پردازش SE بر روی هر داده تست در نظر گرفته شده است تا در صورت پردازش ورودی‌هایی با سایز بزرگتر مشکلی در فرآیندها ایجاد نگردد.

۵- آزمایش‌ها و ارزیابی نتایج

شاخص ارزیابی و عملکرد میان ابزارها و تکنیک‌های فایزینگ میزان افزایش پوشش کد در اهداف تحت تست می‌باشد. در این پژوهش نیز به منظور حفظ اصول و همچنین ارزیابی منصفانه در برابر تحقیقات انجام شده، شاخص بهبود بر پایه مقدار پوشش کد در نظر گرفته شده است. تمامی بهبودها در فایزینگ سعی بر آن دارند تا به پوشش کد صد درصد دست یابند و هر بهبود کمک در جهت رسیدن به این هدف می‌باشد. بررسی و ارزیابی‌ها بر روی یک دستگاه با معماری ۱۲ هسته اینتل Intel Xeon(R) با فعال بودن hyper-threading (مجموعاً ۲۴ هسته منطقی) و ۳۲ گیگابایت حافظه RAM در حال اجرای بر روی سیستم عامل Ubuntu 22.04 LTS x64 انجام شده است. برای اطمینان و قابلیت استناد به خروجی‌های ارزیابی، اجراها به صورت ۲۰ مرتبه با مدت زمان ۳۰ دقیقه انجام شدند. این تعداد اجراها و زمان مورد نیاز برای هر بار اجرا، به ما امکان می‌دهد از داده‌های قابل اعتماد و قابل استناد برای تحلیل و ارزیابی استفاده کنیم.

داده‌های مرتبط با پیاده‌سازی این پژوهش با نام NyxNet-Hybrid در مقایسه‌ها و تحلیل‌ها قابل مشاهده است. شاخص بهبود عملکرد NyxNet-Hybrid در مقایسه با فازرهای AFLNet و NyxNet می‌باشد. فازر NyxNet سه روش در گرفتن پشتیبان‌گیری مقطعی را ارائه می‌دهد که با توجه به ارزیابی‌های درون مقاله بهترین عملکرد آن بصورت Balanced بوده است و از همین رو این مکانیزم را برای مقایسه انتخاب نموده ایم و همچنین پیاده‌سازی این تحقیق نیز با Nyx-Balanced اجرا شده است.

جدول ۱. مقایسه عددی پوشش شاخه

NyxNet-Hybrid		NyxNet-Balanced		AFLNet	
2948	+2.71	2869	-0.03	2870	dcmk
1139	+37.72	1042	+25.9	827	dnsmasq

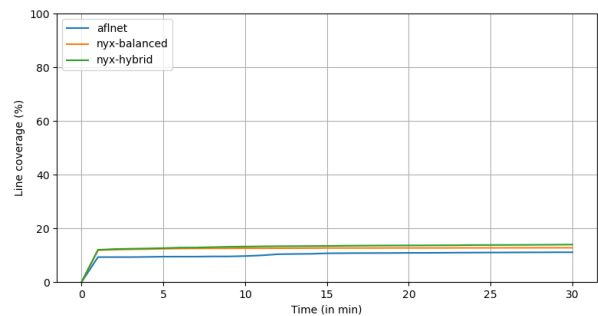
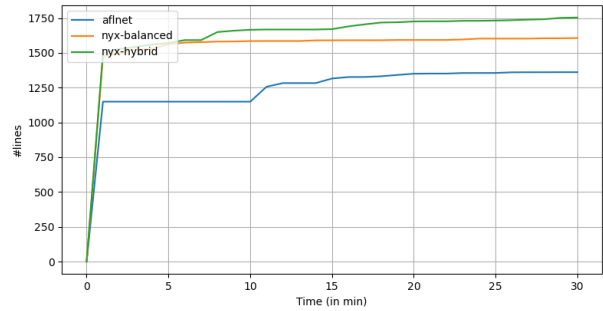
برنامه پیچیده با ویژگی‌های خاص خود می‌باشد، احتمالاً پیاده‌سازی فازینگ ترکیبی و اجرای نمادین برای آن دشوارتر بوده است. اما در آزمون‌های انجام شده روی برنامه dnsmasq شاهد بهبود قابل توجهی در معیارهای ارزیابی بودیم که نشان دهنده تاثیر مثبت روش پیشنهادی می‌باشد. بنابراین می‌توان نتیجه گرفت که میزان بهبود حاصل از فازینگ ترکیبی بستگی به ویژگی‌ها و پیچیدگی برنامه مورد نظر دارد.

علیرغم چالش‌های موجود در پیاده‌سازی این روش، با توجه به نتایج به دست آمده، به نظر می‌رسد از لحاظ صرفه جویی در زمان نتایج رضایت بخشی حاصل شده است. همچنین با بهینه‌سازی بیشتر در پیاده‌سازی، امکان ارتقای نتایج وجود دارد. در مجموع این پژوهش نشان داد که فازینگ ترکیبی می‌تواند گزینه مناسبی برای بهبود فرآیند فازینگ پروتکل‌های شبکه باشد و مسیر مناسبی را برای پژوهش‌های آتی در این حوزه هموار نماید.

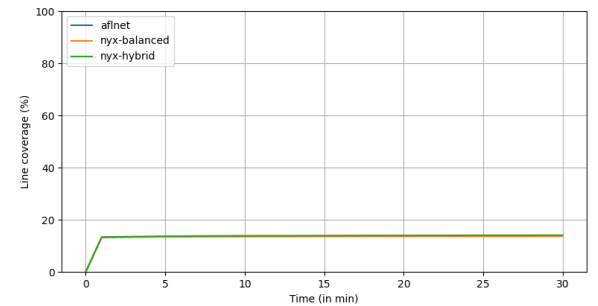
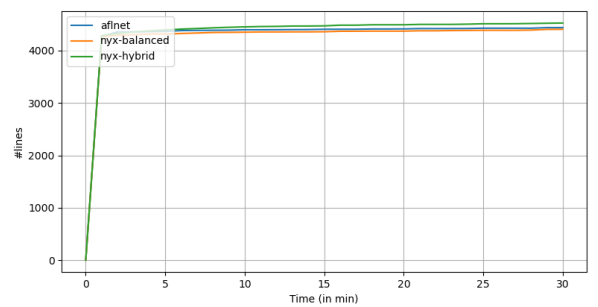
۶- چالش‌های فرارو و پیشنهادها

در این پژوهش سعی شده است تا آخرین و جدیدترین مقالات مرتبط با فازینگ ترکیبی پروتکل‌های شبکه مطالعه شود و بهترین راهکارهای موجود نیز مورد بررسی قرار گیرند. از آنجایی که تاکنون چنین پیاده‌سازی جامعی صورت نگرفته است، لذا باید بهترین روش از میان راهکارهای موجود انتخاب شود تا بتوان از آن برای رسیدن به اهداف مورد نظر بهره برد. البته پیاده‌سازی‌های موجود تمرکز بیشتری بر روی فرآیند فازینگ فایل‌ها داشته‌اند و کمتر با دید شبکه‌ای صورت گرفته‌اند. همان‌طور که در ارزیابی‌های انجام شده مشخص است، آنها در کنار فاکتورهای مربوط به فازینگ فایل، تنها بر اساس همان معیارها ارزیابی و پیش‌رفته‌اند. با این حال، کدمنع تمامی آنها بررسی و در برخی موارد اصلاح شد تا امکان‌سنجی استفاده از آنها در این پژوهش فراهم شود.

اما این پژوهش تلاش کرده است مسیر جدیدی را طی کند و با تمرکز بر پروتکل‌های شبکه، پیاده‌سازی اختصاصی متناسب با نیازهای فازینگ شبکه انجام دهد. دو ابزار DynamoRIO و Triton که بهترین در حوزه کاری خود هستند، مورد استفاده قرار گرفتند تا بهترین عملکرد و خروجی حاصل شود که خوشبختانه چنین هم شد. پس از پیاده‌سازی، ارزیابی‌ها نشان می‌دهند که انتخاب زیرساخت مناسب و همچنین تکنیک فازینگ ترکیبی، همچنان فرآیند فازینگ پروتکل‌های شبکه را بهبود می‌بخشد. بنابراین می‌توان از این تکنیک برای فازینگ پروتکل‌های شبکه بهره برد.



شکل ۱۱. پوشش خط کد سرویس dnsmasq



شکل ۱۲. پوشش خط کد سرویس dcmtk

با توجه به نتایج حاصل از ارزیابی‌های انجام شده، می‌توان نتیجه گرفت که روش پیشنهادی فازینگ ترکیبی با استفاده از ابزارهای Triton و DynamoRIO [19] توانسته است منجر به بهبود فرآیند فازینگ شود. البته این بهبود در تمامی موارد یکسان نبوده و بستگی به شرایط و ویژگی‌های برنامه مورد آزمایش داشته است. در آزمون‌های انجام شده روی برنامه dcmtk، اگرچه شاهد بهبود چشمگیری در معیارهای ارزیابی مانند پوشش کد و پوشش شاخه نبوده ایم، اما با این حال نتایج حاکی از آن است که فازینگ ترکیبی توانسته به نحوی شرایط را بهبود بخشد. از آنجایی که dcmtk یک

- [9] R. Natella, "Statefl: Greybox fuzzing for stateful network servers," *Empirical Software Engineering*, vol. 27, no. 7, p. 191, 2022.
- [10] J. Ba, M. Böhme, Z. Mirzamomen, and A. Roychoudhury, "Stateful greybox fuzzing," in 31st USENIX Security Symposium (USENIX Security 22), pp. 3255–3272, 2022.
- [11] C. Aschermann, S. Schumilo, A. Abbasi, and T. Holz, "Ijon: Exploring deep state spaces via fuzzing," in 2020 IEEE Symposium on Security and Privacy (SP), pp. 1597–1612, IEEE, 2020.
- [12] Y. Shoshitaishvili, R. Wang, C. Salls, N. Stephens, M. Polino, A. Dutcher, J. Grosen, S. Feng, C. Hauser, C. Kruegel, and G. Vigna, "SoK: (State of) The Art of War: Offensive Techniques in Binary Analysis," in IEEE Symposium on Security and Privacy, 2016.
- [13] V. Chipounov, V. Kuznetsov, and G. Candea, "S2e: A platform for in-vivo multipath analysis of software systems," *Acm Sigplan Notices*, vol. 46, no. 3, pp. 265–278, 2011.
- [14] I. Yun, S. Lee, M. Xu, Y. Jang, and T. Kim, "{QSYM}: A practical concolic execution engine tailored for hybrid fuzzing," in 27th USENIX Security Symposium (USENIX Security 18), pp. 745–761, 2018.
- [15] S. Poeplau and A. Francillon, "Symbolic execution with {SymCC}: Don't interpret, compile!" in 29th USENIX Security Symposium (USENIX Security 20), pp. 181–198, 2020.
- [16] S. Poeplau and A. Francillon, "Symqemu: Compilation-based symbolic execution for binaries," in NDSS 2021, Network and Distributed System Security Symposium, Internet Society, 2021.
- [17] L. Borzacchiello, E. Coppa, and C. Demetrescu, "Fuzzolic: Mixing fuzzing and concolic execution," *Computers & Security*, vol. 108, p. 102368, 2021.
- [18] D. Bruening and T. Garnett, "Building dynamic instrumentation tools with dynamorio," in Proc. Int. Conf. IEEE/ACM Code Generation and Optimization (CGO), Shen Zhen, China, 2013.
- [19] F. Soudel and J. Salwan, "Triton: A dynamic symbolic execution framework," in Symposium sur la sécurité des technologies de l'information et des communications, SSTIC, France, Rennes, pp. 31–54, 2015.
- [20] Leonardo Xompero, Tahir Mohammad, Jouni Isoaho, Jürgen Grossi, "Fuzz Testing Approach for Embedded Avionic Software", The 15th International Conference on Ambient Systems, Networks and Technologies (ANT), 2024.
- [21] Chi Zang et al. "TR-Fuzz: A syntax valid tool for fuzzing C compilers", *Science of Computer Programming*, vol. 238, 2024.
- [22] Haorn Xu et al. "Fuzzing JavaScript engines with a syntax-aware neural program model", *Computers & Security*, vol. 144, 2024.
- [۲۳] زهره مافی، سیدحسن میریان حسین‌آبادی. "انتخاب موارد آزمون مبتنی بر روش تولید آزمون رانه"، دو فصلنامه فناوری اطلاعات و ارتباطات ایران، دوره ۱۶ شماره ۵۹، ۱۴۰۳.
- از جمله چالش‌ها و محدودیت‌های پیش‌روی فازینگ ترکیبی پروتکل‌های شبکه می‌توان به موارد زیر اشاره کرد:
- پیاده‌سازی توابع بیشتر به صورت نمادین در SE که می‌تواند دقت بیشتر و همچنین سربار محاسباتی کمتری را به همراه داشته باشد.
 - استفاده از تکنیک‌های یادگیری ماشین برای اولویت‌بندی و انتخاب مناسب‌ترین داده‌های ورودی و پرهیز از انفجار مسیرها در بخش SE
 - به نظر می‌رسد با پیگیری این موارد در تحقیقات آتی، بتوان چالش‌های فازینگ ترکیبی پروتکل‌های شبکه را به طور مؤثری مدیریت کرد.

مراجع

- [1] Cui, Lei, Jiancong Cui, Zhiyu Hao, Lun Li, Zhenquan Ding, and Yongji Liu. "An empirical study of vulnerability discovery methods over the past ten years." *Computers & Security*, 2022.
- [2] M. Zalewski, "American fuzzy lop - a security-oriented fuzzer".
- [3] F. Rustamov, J. Kim, J. Yu, and J. Yun, "Exploratory review of hybrid fuzzing for automated vulnerability detection," *IEEE Access*, vol. 9, pp. 131166–131190, 2021.
- [4] S. Zhou, Z. Yang, D. Qiao, P. Liu, M. Yang, Z. Wang, and C. Wu, "Ferry: {StateAware} symbolic execution for exploring {State-Dependent} program paths," in 31st USENIX Security Symposium (USENIX Security 22), pp. 4365–4382, 2022.
- [5] V.-T. Pham, M. Böhme, and A. Roychoudhury, "Aflnet: a greybox fuzzer for network protocols," in 2020 IEEE 13th International Conference on Software Testing, Validation and Verification (ICST), pp. 460–465, IEEE, 2020.
- [6] S. Schumilo, C. Aschermann, A. Abbasi, S. Wörner, and T. Holz, "Nyx: Greybox hypervisor fuzzing using fast snapshots and affine types," in 30th USENIX Security Symposium (USENIX Security 21), pp. 2597–2614, 2021.
- [7] S. Schumilo, C. Aschermann, A. Jemmett, A. Abbasi, and T. Holz, "Nyx-net: network fuzzing with incremental snapshots," in Proceedings of the Seventeenth European Conference on Computer Systems, pp. 166–180, 2022.
- [8] J. Li, S. Li, G. Sun, T. Chen, and H. Yu, "Snpsfuzzer: A fast greybox fuzzer for stateful network protocols using snapshots," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2673–2687, 2022.

Presenting a Method for Agile Enterprise Architecture Modeling

Ali Razi¹, Reza Rezaei^{2*}, Ahmad Ali Yazdanpanah³

¹ Department of IT Management, Science and Research branch, Islamic Azad University, Tehran, Iran

² Department of Computer Engineering, Saveh Branch, Islamic Azad University, Saveh, Iran

³ Faculty Member of Statistic Researches & Information Technology Group,
Institute for Research & Planning in Higher Education (IRPHE),
Ministry of Science, Research and Technology, Tehran, Iran

Received: 06 May 2023, Revised: 03 March 2024, Accepted: 16 March 2024

Paper type: Research

Abstract

One of the important issues and concerns in agile enterprise architecture is the documenting and modeling of requirements and artifacts. Architectural models and documents should be produced and updated as necessary and while meeting the requirements and concerns of the stakeholders, they should be presented in the shortest time and at the lowest cost. On the other hand, methods and tools should be used that are applicable and simple, and while using agile practices and modeling standards, have the necessary comprehensiveness. Researches has been done regarding the modeling and documenting of agile enterprise architecture. Since there is no comprehensive, standard and minimal method for modeling and documenting agile enterprise architecture in related researches, therefore, in this paper, a method for modeling agile enterprise architecture is presented and evaluated by the combined (qualitative + quantitative) method. Qualitative evaluation is performed through a case study. For quantitative evaluation, the AHP method is used to weighting and ranking the criteria. Evaluation criteria are extracted based on library studies and experts' opinions. In order to quantitatively evaluate the proposed method, six criteria have been proposed in this paper, which are: production of required and necessary artifacts, reduction of modeling time, reduction of cost, improvement of stakeholders satisfaction, improvement of applicability and increase of simplicity, and criteria of production of required and necessary artifacts, has won the highest rank.

Keywords: Agile Enterprise Architecture Modeling, Agile Enterprise Architecture Documenting, Agile Enterprise Architecture Method, Minimum Viable Enterprise Architecture, Agile Enterprise Requirement Modeling, Evaluating Agile Modeling.

* Corresponding Author's email: rezarezaei@iau-saveh.ac.ir

ارائه روشی برای مدل‌سازی معماری سازمانی چابک

علی راضی^۱، رضا رضایی^{۲*}، احمدعلی یزدان پناه^۳

^۱ دانشجوی رشته مدیریت فناوری اطلاعات، گروه مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

^۲ گروه کامپیوتر، دانشکده فنی و مهندسی، واحد ساوه، دانشگاه آزاد اسلامی، ساوه، ایران

^۳ استادیار موسسه پژوهش و برنامه‌ریزی آموزش عالی، موسسه پژوهش و برنامه‌ریزی آموزش عالی، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۲/۱۶ تاریخ بازبینی: ۱۴۰۲/۱۲/۱۳ تاریخ پذیرش: ۱۴۰۲/۱۲/۲۶

نوع مقاله: پژوهشی

چکیده

یکی از مسائل و دغدغه‌های مهم در معماری سازمانی چابک، مستند سازی و مدل‌سازی نیازمندی‌ها و فرآورده‌ها است. مدل‌ها و مستندات معماری باید به اندازه و در حد ضرورت تولید و بروزرسانی شوند و ضمن رفع نیازمندی‌ها و دغدغه‌های ذینفعان، در کمترین زمان و با کمترین هزینه ارائه گردند. از طرف دیگر باید از روش‌ها و ابزارهایی استفاده نمود که کاربردپذیر و ساده بوده و ضمن بکارگیری روش‌های عملی چابک و استانداردهای مدل‌سازی جامعیت لازم را داشته باشند. تحقیقاتی در خصوص مدل‌سازی و مستندسازی معماری سازمانی چابک انجام شده است. از آنجایی که در تحقیقات مرتبط روشی جامع، استاندارد و کمینه برای مدل‌سازی و مستندسازی معماری سازمانی چابک ارائه نشده، لذا در این مقاله روشی برای مدل‌سازی معماری سازمانی چابک ارائه گردیده و با روش ترکیبی (کیفی + کمی) ارزیابی می‌گردد. ارزیابی کیفی از طریق مطالعه موردی انجام می‌پذیرد. برای ارزیابی کمی از روش AHP برای وزن‌دهی و رتبه‌بندی معیارها استفاده می‌شود. معیارهای ارزیابی بر اساس مطالعات کتابخانه‌ای و نظرات خبرگان استخراج می‌گردند. برای ارزیابی کمی روش پیشنهادی در این مقاله شش معیار مطرح شده‌اند که عبارتند از: تولید فرآورده‌های مورد نیاز و ضروری، کاهش زمان مدل‌سازی، کاهش هزینه، بهبود رضایت ذینفعان، بهبود کاربردپذیری و افزایش سادگی که معیار تولید فرآورده‌های مورد نیاز و ضروری رتبه اول را کسب کرده است.

کلیدواژگان: مدل‌سازی معماری سازمانی چابک، مستندسازی معماری سازمانی چابک، روش معماری سازمانی چابک، حداقل معماری سازمانی قابل دوام، مدل‌سازی نیازمندی‌های سازمانی چابک، ارزیابی مدل‌سازی چابک.

* رایانامه نویسنده مسؤل: rezarezaei@iau-saveh.ac.ir

۱- مقدمه

سازمانی چابک تاکید کرده‌اند [۱۱] تا [۱۳]. بسیاری از تحقیقات انجام شده فاقد جامعیت بوده و برای مدل‌سازی همه لایه‌های معماری سازمانی راه‌حل مشخصی ارائه نکرده‌اند [۳۱] تا [۳۳]. همچنین در برخی از تحقیق‌های انجام شده مراحل و گام‌های اجرایی جهت تولید مدل‌ها و فرآورده‌ها مشخص نشده‌اند [۸] و [۹]. طبق بیانیه چابک و اصول حاکم بر معماری سازمانی چابک، ضروری است مدل‌های معماری در کمترین زمان ممکن و با کمترین هزینه تولید و بروزرسانی شوند [۱] و [۳] و [۴]. همچنین سادگی و کاربردپذیری روش مدل‌سازی و ابزارهای مورد استفاده و بهبود رضایت ذینفعان معماری از مهمترین شاخص‌ها در مدل‌سازی معماری سازمانی چابک هستند [۲] و [۲۳]. چارچوب‌ها و روش‌های چابک معماری سازمانی بر تمرکز بر مدل‌سازی وضعیت مطلوب معماری تاکید کرده‌اند [۳۴] تا [۳۶]، با این حال اکثر روش‌ها و راهکارهای ارائه شده در این خصوص سکوت کرده‌اند. اما سوال اول اینکه آیا روشی برای مدل‌سازی معماری سازمانی چابک با ویژگی‌های زیر وجود دارد؟

- جامعیت (مدل‌سازی همه لایه‌های معماری سازمانی)
- تاکید و تمرکز بر مدل‌سازی معماری وضعیت مطلوب و گلوگاه‌های سازمان
- تعیین مراحل انجام مدل‌سازی و گام‌های هر مرحله
- تاکید بر مدل‌سازی استاندارد
- کمینه کردن نمادها و عناصر استانداردهای مدل‌سازی
- تولید فرآورده‌های ضروری و لازم و حذف فرآورده‌های غیر ضروری و زاید
- بکارگیری روش‌های عملی چابک

و سوال دوم اینکه در صورت وجود روشی با ویژگی‌های فوق، چگونه ارزیابی می‌شود؟

با توجه به اینکه تحقیقات انجام شده در خصوص روش‌ها و چارچوب‌های مدل‌سازی معماری سازمانی چابک، جامعیت لازم را برای پاسخگویی به سوالات مطرح شده ندارند لذا در این مقاله به آن پرداخته شده و یک روش برای مدل‌سازی معماری سازمانی چابک با ویژگی‌های اشاره شده ارائه و به روش ترکیبی^{۱۱} ارزیابی می‌گردد. در ادامه در بخش دوم پیشینه و زمینه تحقیق بررسی

معماری سازمانی چابک از بیانیه چابک^۱ و اصول حاکم بر آن متأثر است [۱]. در معماری سازمانی چابک، ضروری است که مدل‌ها و فرآورده‌ها به اندازه و به جا تولید و بروزرسانی شوند [۲]. مدل‌سازی معماری سازمانی چابک، مجموعه‌ای از اصول، ارزش‌ها و تجارب عملی چابک است که به ذینفعان معماری سازمانی کمک می‌کند که فرآورده‌ها و مدل‌های معماری سازمانی را به شکل تکاملی، تدریجی و ساده تولید و بروزرسانی نمایند [۳] و [۴]. نزدیک شدن به معماری تک صفحه‌ای یکی از اهداف مهم معماران سازمانی چابک است [۵]. تحقیقاتی در خصوص ارائه روش‌ها، راهکارها و چارچوب‌ها در حوزه معماری سازمانی چابک و مدل‌سازی معماری سازمانی چابک انجام شده‌اند [۶] تا [۱۳]. برخی از تحقیق‌ها بر مدل‌سازی استاندارد و ترکیب و بکارگیری استانداردهای مدل‌سازی نظیر: [۱۴] ArchiMate^۳، [۱۵] UML^۳، [۱۶] BPMN^۴، [۱۷] FAML^۵، [۱۸] SoaML^۶، [۱۹] BMM^۷، [۲۰] DMN^۸، [۲۱] CMMN^۹ و [۲۲] SysML^{۱۰} در معماری سازمانی تاکید کرده‌اند [۲۳] و [۲۴]. با این حال با افزایش تعداد استانداردهای مدل‌سازی و به تبع آن افزایش تعداد عناصر و نمادهای آنها، این نگرانی و دغدغه ایجاد می‌شود که زمان تولید مدل‌ها و مستندات و حجم آنها افزایش یابد [۲۵] و [۲۶]. ضمن اینکه ممکن است در یادگیری و آموزش استانداردها و پشتیبانی ابزارها پیچیدگی ایجاد شده و هزینه‌ها نیز افزایش یابند که این بر خلاف پارادایم حاکم در معماری سازمانی چابک است [۲]. برخی از معماران چابک با در نظر گرفتن اصول و قواعد مدل‌سازی و مستندسازی چابک و با توجه به وجود دید معمارانه و کلان‌نگر در معماری سازمانی از تعدد بکارگیری استانداردهای مختلف اجتناب کرده و سعی می‌کنند به ساده‌ترین روش و کمترین تعداد استاندارد، مدل‌های معماری را تولید نمایند [۲۷] تا [۳۰]. از طرف دیگر برخی از معماران چابک نیز به دلیل تنوع نیازمندی‌های معماری و ایجاد زبان مشترک بین ذینفعان با تخصص‌ها و فرهنگ‌های مختلف بر بکارگیری و ترکیب حداکثری استانداردهای مدل‌سازی تاکید دارند [۷]. برخی دیگر از معماران چابک نیز بر کمینه کردن عناصر استانداردهای مدل‌سازی تمرکز کرده‌اند [۲۵] و [۲۶]. برخی از تحقیق‌های انجام شده نیز بر بکارگیری اصول و تجارب عملی چابک در مدل‌سازی معماری

6 Service Oriented Architecture Modeling Language
7 Business Motivation Model
8 Decision Model and Notation
9 Case Management Model and Notation
10 The Systems Modeling Language
11 Mixed Method

1 Agile Manifesto
2 Architecture-Animate
3 Unified Modeling Language
4 Business Process Model and Notation
5 FAME [Framework for Agent-Oriented Method Engineering] Language

سازمانی ارائه کرده‌اند [۶]. در این روش بر تعریف اسپرینت‌ها، تحویل مداوم، تعاملات کوتاه و زیاد با مشتریان و بهبود سیستماتیک و بکارگیری استانداردهای ArchiMate و BPMN تاکید شده است. با این حال روشی برای کمینه کردن فرآورده‌ها و عناصر مدل‌سازی ارائه نشده است.

شیرازی و همکارانش چارچوبی برای معماری سازمانی چابک ارائه کرده‌اند [۸]. علی‌رغم تاکید بر روش‌های عملی چابک و چابک شدن مدیریت فرآیند معماری سازمانی روشی برای تولید مدل‌های چابک ارائه نشده است.

امیری در تحقیق خود چالش‌ها و نقاط ضعف روش‌های چابک را برای بکارگیری در معماری سازمانی بررسی کرده است [۹]. با این حال در خصوص تولید فرآورده‌ها و مدل‌های چابک راهکاری ارائه نشده است.

راهکارها و روش‌های دیگری هم در حوزه معماری سازمانی چابک ارائه شده‌اند که علی‌رغم تاکید بر روش‌های عملی چابک و تاکید بر چابک شدن مدیریت فرآیند معماری سازمانی، روشی را برای تولید مدل‌های چابک ارائه نکرده‌اند [۷] و [۱۰] تا [۱۲].

با توجه به تحقیق‌های انجام شده، نواقصی شامل عدم جامعیت برای مدل‌سازی همه لایه‌های معماری سازمانی، عدم کمینه کردن عناصر استانداردهای مدل‌سازی و فرآورده‌های معماری و عدم تعیین شفاف مراحل و گام‌های اجرایی برای مدل‌سازی، مشاهده می‌شود.

Werewka و همکارانش روشی را برای معماری سازمانی چابک مبتنی بر اسکرام ارائه کرده‌اند [۱۳]. برای مدل‌سازی، استاندارد ArchiMate به صورت تکرارپذیر پیشنهاد شده است. با این حال روشی برای کمینه کردن عناصر و فرآورده‌های معماری و بکارگیری تجارب عملی چابک ارائه نشده است.

Gill راهکاری مبتنی بر ترکیب شش استاندارد مدل‌سازی در معماری سازمانی ارائه کرده و با مطالعه موردی کاربردی و یکپارچگی آن را ارزیابی کرده است [۲۳]. در راهکار ارائه شده توسط Gill، مراحل و گام‌های انجام مدل‌سازی دقیقاً مشخص نشده‌اند و در خصوص چابکی روند مدل‌سازی نیز راه‌حل مشخصی ارائه نشده است. همچنین تحقیق ایشان در خصوص مدل‌سازی قوانین و تصمیم‌های کسب‌وکار نیز قابل بهبود است.

Buchalceva در تحقیق خود راهکاری برای خلاصه‌سازی و ساده‌سازی استاندارد ArchiMate برای مدل‌سازی سازمان‌های

می‌گردد. در بخش سوم روش پیشنهادی تشریح می‌گردد. در بخش چهارم روش ارزیابی ارائه می‌گردد. در بخش پنجم ارزیابی روش پیشنهادی و تجزیه و تحلیل داده‌ها به روش ترکیبی انجام می‌شود. در بخش ششم روند انجام کار و نتایج حاصله بررسی گردیده و بحث و تحلیل انجام می‌پذیرد. در نهایت نتیجه‌گیری انجام شده و کارهای آینده معرفی می‌شوند.

۲- پیشینه و زمینه تحقیق

در معماری سازمانی چابک، برخلاف روش‌های سنتی و آبشاری بر تولید مدل‌ها و مستندات به اندازه و به شکل تکاملی و تدریجی تأکید شده است [۲]. تولید مدل‌ها در کمترین زمان، با کمترین هزینه، به ساده‌ترین روش ممکن و دستیابی به معماری تک‌صفحه‌ای از مهمترین اهداف معماران چابک است [۳] تا [۵]. روش‌ها، راهکارها و چارچوب‌هایی برای مدل‌سازی معماری سازمانی چابک ارائه شده‌اند. برخی از تحقیقات انجام شده در این حوزه به ارائه راهکارهایی مبتنی بر بکارگیری تجارب عملی چابک^۱ نظیر: تعریف تکرارها و اسپرینت‌ها، تحویل مداوم مدل‌ها و تعاملات نزدیک با مشتریان و ذینفعان پرداخته‌اند. با این حال در خصوص ارائه مراحل و گام‌های مدل‌سازی و نحوه تولید مدل‌ها و فرآورده‌ها راهکار خاصی ارائه نکرده‌اند.

برخی دیگر از محققان و صاحب‌نظران بر بکارگیری مدل‌سازی استاندارد و ترکیب استانداردهای خوش‌تعریف و شناخته شده برای مدل‌سازی معماری سازمانی چابک تاکید کرده‌اند. آنها اعتقاد دارند بکارگیری استانداردهای مدل‌سازی معروف در مدل‌سازی معماری سازمانی چابک سبب ایجاد درک و فهم مشترک بین ذینفعان معماری می‌گردد. تحقیق‌هایی در خصوص بکارگیری استاندارد ArchiMate در مدل‌سازی معماری سازمانی انجام شده است. برخی از تحقیقات نیز به ترکیب استاندارد ArchiMate با سایر استانداردهای مدل‌سازی نظیر UML، BPMN، FAML، SoaML و BMM، DMN، CMMN و SysML پرداخته‌اند.

برخی از معماران چابک دغدغه دارند که با افزایش تعداد استانداردها و عناصر و نمادهای آنها در مدل‌سازی پیچیدگی ایجاد شده و زمان مستندسازی و هزینه‌ها افزایش یافته و یادگیری و پشتیبانی مدل‌ها نیز با چالش مواجه گردد. بر این اساس این دسته از معماران و مدل‌سازان، راهکارهایی را برای کمینه کردن عناصر استانداردهای مدل‌سازی ارائه کرده‌اند.

Medeiros و همکارانش یک رویکرد چابک برای مدل‌سازی معماری

^۱ Agile Practices

مشخصی وجود ندارد.

Sadovykh و همکارانش در تحقیق خود روشی را برای مدل‌سازی معماری با ترکیب سه استاندارد مدل‌سازی ارائه کرده‌اند [۳۸]. در این تحقیق بر مدل‌سازی معماری سرویس‌گرا و همراستایی آن با حوزه کسب‌وکار و استراتژی تأکید شده است. با این حال در خصوص بکارگیری روش‌های عملی چابک و تعیین مراحل و گام‌های اجرا راهکاری ارائه نشده است.

Ding و همکارانش روشی را برای بومی‌سازی و مختصرسازی استاندارد ArchiMate برای مدل‌سازی لایه‌های نرم‌افزارهای کاربردی و زیرساخت فناوری ارائه کرده‌اند [۴۲]. با این حال در خصوص تعیین مراحل و گام‌ها برای مدل‌سازی همه لایه‌های معماری راهکار مشخصی ارائه نشده ضمن اینکه در خصوص تجارب عملی چابک نیز روشی مطرح نشده است.

Kirikova در مقاله خود روشی را برای تولید مدل‌ها و فرآورده‌های سازمانی ارائه کرده است [۴۳]. در خصوص بکارگیری استانداردهای مدل‌سازی و کمینه کردن فرآورده‌ها و عناصر استانداردهای مدل‌سازی روش خاصی ارائه نشده است.

Vernadat در مقاله خود کارهای تحقیقاتی انجام شده در مدل‌سازی سازمانی در طول چهار دهه گذشته را بررسی کرده است. وی اصول مدل‌سازی، ساختارها، زبان‌ها، چارچوب‌ها و استانداردهای مدل‌سازی را معرفی و تشریح نموده است. سپس در مورد تحولات آینده مدل‌سازی سازمانی در زمینه تولید هوشمند یا صنعت ۴٫۰ بحث نموده است [۴۴]. با این حال در خصوص مدل‌سازی معماری سازمانی چابک، راه‌حل مشخصی ارائه نشده است.

چارچوب‌های چابک معماری سازمانی بر تمرکز بر مدل‌سازی وضعیت مطلوب تأکید کرده‌اند [۳۴] و [۳۵] و [۳۶]، با این حال تقریباً هیچ کدام از تحقیق‌های انجام شده برای این موضوع راه‌حل مشخصی ارائه نکرده‌اند. همچنین تحقیق‌های انجام شده برای ارزیابی راهکارها و روش‌ها از مطالعه موردی برای نشان دادن مدل‌های تولید شده استفاده کرده‌اند، اما ارزیابی کمی مبتنی بر شاخص‌های مدل‌سازی معماری سازمانی چابک و نظرات خبرگان، انجام نشده است. این در حالی است که ارزیابی کمی در کنار ارزیابی کیفی در پژوهش‌های مرتبط با معماری سازمانی حائز اهمیت است. برای ارزیابی کمی راه‌حل‌های ارائه شده در معماری سازمانی و از جمله مدل‌سازی معماری سازمانی چابک، روش‌های مختلفی وجود دارد. بکارگیری روش‌های تصمیم‌گیری چند معیاره^۱ از قبیل AHP^۲

کوچک ارائه کرده است. علی‌رغم ارائه یک روش با مراحل و گام‌های مشخص، راه‌حل مشخصی برای بکارگیری تجارب عملی چابک ارائه نشده است [۲۶].

Zrnc و همکارانش و Armour و همکارانش در تحقیق‌های خود روشی برای بکارگیری عناصر استاندارد UML در مدل‌سازی سازمانی ارائه کرده‌اند اما در خصوص بکارگیری تجارب عملی چابک و تعیین مراحل و گام‌های مدل‌سازی راهکار مشخصی ارائه نشده است. ضمن اینکه عناصر استاندارد UML برای مدل‌سازی معماری نرم‌افزار طراحی شده‌اند و کمتر مقبولیت سازمانی دارند [۲۷] و [۲۸].

Yamamoto و همکارانش در مقاله خود با استفاده از روش تحلیل جنبه‌گرا، راهکاری را برای ساده کردن بکارگیری استاندارد ArchiMate مطرح کرده و به روش مطالعه موردی ارزیابی نموده‌اند [۲۵]. Yamamoto در مقاله دیگر راهکاری برای مدل‌سازی سیستم سرویس هوش مصنوعی ارائه کرده است [۲۹]. وی در مقاله دیگری راهکاری برای محدود کردن تعداد عناصر استاندارد ArchiMate برای مدل‌سازی یک شهر هوشمند ارائه کرده و از طریق مطالعه موردی آن را ارزیابی نموده است [۳۰]. با این حال در خصوص بکارگیری تجارب عملی چابک راهکار مشخصی ارائه نشده است.

Aldea و همکارانش و Kitsios و همکارانش در تحقیق‌های خود روشی برای مدل‌سازی لایه استراتژی با عناصر استاندارد ArchiMate مطرح کرده‌اند با این حال در خصوص مدل‌سازی سایر لایه‌های معماری و بکارگیری تجارب عملی چابک راهکار مشخصی ارائه نکرده‌اند [۳۱] و [۳۲].

Bhattacharya در تحقیق خود روشی برای مدل‌سازی لایه استراتژیک معماری با ترکیب دو استاندارد مدل‌سازی ارائه کرده است [۳۳]. با این حال در خصوص مدل‌سازی سایر لایه‌های معماری و بکارگیری تجارب عملی چابک راهکار مشخصی ارائه نشده است.

Pankowska در تحقیق خود راهکاری برای مدل‌سازی معماری کسب‌وکار با ترکیب سه استاندارد مدل‌سازی ارائه کرده ولی راه‌حل چابکی برای بکارگیری آن مطرح نکرده است [۳۹]. ضمن اینکه راهکاری برای مدل‌سازی سایر لایه‌های معماری سازمانی نیز ارائه نشده است. Pankowska در تحقیق دیگر نیز روشی برای مدل‌سازی چابک نیازمندی‌های سازمانی با ترکیب دو استاندارد مدل‌سازی ارائه و با یک مطالعه موردی برای فروشگاه آنلاین آن را ارزیابی کرده است [۳۷]. با این حال برای مدل‌سازی همه لایه‌های معماری راه‌حل

2 Analytical Hierarchy Process

1 Multiple Attribute Decision-making (MADM)

ANP¹ و ... برای ارزیابی راهکارهای ارائه شده در معماری سازمانی و از جمله مدل‌سازی معماری سازمانی چابک مورد پذیرش هستند [۴۵] و [۴۶].

بکارگیری شبکه‌های پتری^۲ [۴۷] و نظریه صف^۳ [۴۸]، روش‌های دیگری برای ارزیابی کمی کارایی و قابلیت اطمینان راه‌حل‌های ارائه شده هستند. شبکه‌های پتری برای مدل‌سازی ریاضی فرآیندها و ارزیابی آن‌ها مفید هستند. نظریه صف نیز، برای مدل‌سازی ریاضی یک صف در انتظار، و تصمیم‌گیری در مورد منابع مورد نیاز آن کاربرد دارد. تحقیقاتی در خصوص بکارگیری شبکه‌های پتری در ارزیابی معماری و مدل‌سازی استاندارد انجام شده‌اند [۴۹] و [۵۰] و [۵۱] و [۵۲]. [۵۳] و [۵۴]. با این حال بهبود و تکمیل آن‌ها در آن کارهای آینده ضروری هستند.

مدل‌سازی معماری سازمانی چابک دارای ویژگی‌هایی است که عبارتند از: جامعیت مدل‌سازی، تاکید و تمرکز بر مدل‌سازی معماری وضعیت مطلوب، تاکید و تمرکز بر مدل‌سازی گلوگاه‌های سازمان، تعیین مراحل انجام مدل‌سازی و گام‌های هر مرحله، تاکید بر مدل‌سازی استاندارد، کمینه‌سازی، تولید فرآورده‌های ضروری و لازم و حذف فرآورده‌های غیر ضروری و زاید، تعیین نقش مدل‌ساز چابک، برنامه‌ریزی تکرار و انتشار، مدل‌سازی نیازمندی‌های چابک و اولویت‌بندی مدل‌سازی طبق نیازمندی‌ها و دغدغه‌های ذینفعان [۱] و [۲] و [۵] و [۲۳] و [۳۴] و [۳۶].

از آنجاییکه در پیشینه تحقیق برخی از ویژگی‌های مدل‌سازی معماری سازمانی چابک توسط روش‌های ارائه شده پوشش داده نشده و یا به شکل ناقص پوشش داده شده‌اند، لذا ارائه یک روش جامع و کامل برای مدل‌سازی معماری سازمانی چابک برای پوشش کامل همه ویژگی‌های مورد نظر لازم و ضروری به نظر می‌رسد.

۳- روش پیشنهادی

در این مقاله یک روش برای مدل‌سازی معماری سازمانی چابک با هدف پوشش نقایص پیشینه تحقیق ارائه می‌گردد. روش پیشنهادی بر اساس یک روش تحقیق تدوین و طراحی شده است. تحقیق انجام شده از نظر نتایج یا پیامد از نوع کاربردی، از نظر هدف از نوع توصیفی و به شکل مطالعه موردی و از نظر روش و اجراء، ترکیبی (کیفی + کمی) است. همچنین تحقیق انجام شده از نوع پیمایشی نیز می‌باشد زیرا از ابزارهایی مثل پرسشنامه AHP و مصاحبه با خبرگان برای جمع‌آوری داده‌ها استفاده شده است. برای جمع‌آوری

داده‌ها از روش‌هایی نظیر: مطالعات کتابخانه‌ای، فیش برداری اطلاعات از اینترنت و کتابخانه و افکارسنجی، مصاحبه و نظرسنجی از خبرگان به شکل تحقیق میدانی استفاده شده است. جامعه آماری تحقیق، محدود و شامل ۱۲ نفر از خبرگان مدل‌سازی معماری سازمانی چابک است. برای نمونه‌گیری از روش نمونه‌گیری غیر تصادفی و از نوع نمونه‌گیری گلوله‌برفی و نمونه‌گیری قضاوتی استفاده شده است. سازمانی که به منظور اجرای مطالعه موردی بر اساس روش پیشنهادی برگزیده شده، شرکت مخابرات ایران است. سازمان مورد نظر یک سازمان ایرانی است و به معماری سازمانی در راستای اهداف و برنامه‌های دولت الکترونیک نیاز دارد. همچنین به دلیل اینکه در معرض تغییرات زیاد و مکرر محیطی، کسب‌وکار و فناوری قرار گرفته، به معماری سازمانی چابک نیاز دارد. به دلیل وجود ذینفعان متنوع و متخصص در حوزه فناوری اطلاعات و ارتباطات در این شرکت و نیاز به تولید و بروزرسانی مستندات و مدل‌های معماری به شکل مختصر و مفید با کمترین هزینه و در کمترین زمان ممکن، به یک روش جامع، کامل، ساده و کاربردی‌پذیر برای مدل‌سازی معماری سازمانی چابک نیاز دارد.

روش پیشنهادی برای پوشش همه ویژگی‌های مدل‌سازی معماری سازمانی چابک تدوین گردیده و بر مبنای اصول و ویژگی‌های زیر استوار است:

- جامعیت: روش پیشنهادی جامع است یعنی همه لایه‌ها و مراحل معماری سازمانی را برای مدل‌سازی چابک نیازمندی‌ها و فرآورده‌ها در نظر گرفته است.
- تعیین مراحل انجام مدل‌سازی و گام‌های هر مرحله: در روش پیشنهادی مراحل انجام مدل‌سازی معماری سازمانی چابک مشخص گردیده است. برای هر مرحله تعدادی گام در نظر گرفته شده است. همچنین مشخص شده است که از هر مرحله به کدام مراحل دیگر می‌توان حرکت کرد.
- تاکید بر مدل‌سازی استاندارد: در روش پیشنهادی بر مدل‌سازی استاندارد یعنی بکارگیری استانداردهای مدل‌سازی مشهور، شناخته شده و خوش تعریف تاکید شده است. بکارگیری مدل‌سازی استاندارد سبب ایجاد وحدت نظر و درک مشترک بین ذینفعان معماری سازمانی شده و از مدل‌سازی به شکل سلیقه‌ای جلوگیری می‌کند.
- کمینه کردن نمادها و عناصر استانداردهای مدل‌سازی: در روش پیشنهادی از همه عناصر و المان‌های استانداردهای

- مدل‌سازی استفاده نمی‌شود بلکه از عناصر مشخص و محدود برای تولید فرآورده‌های مورد نیاز استفاده می‌گردد. بکارگیری همه عناصر و المان‌های استانداردهای مدل‌سازی خود سبب پیچیدگی مدل‌سازی شده و چابکی را کاهش می‌دهد. در روش پیشنهادی در صورتی از المان‌ها و عناصر بیشتر برای مدل‌سازی استفاده می‌شود که نیاز به تولید فرآورده‌های با جزئیات بیشتر توسط ذینفعان تشخیص داده شود.
- تولید فرآورده‌های ضروری و لازم و حذف فرآورده‌های غیر ضروری و زاید: در روش پیشنهادی همه فرآورده‌های معماری سازمانی تولید نمی‌شوند بلکه مطابق با نیازمندی‌ها و دغدغه‌های ذینفعان معماری مهمترین آنها تولید و بروز می‌شوند.
- بکارگیری روش‌های عملی چابک: در روش پیشنهادی بر بکارگیری روش‌های عملی چابک از قبیل: برنامه‌ریزی تکرار، بکارگیری اپیک و استوری برای جمع‌آوری نیازمندی‌ها، تمرکز بیشتر بر معماری وضعیت مطلوب و نقاط کلیدی و گلوگاه‌های سازمان تاکید شده است.

۳-۲- مرحله مدل‌سازی نیازمندی‌های معماری (وظیفه‌مندی و غیر وظیفه‌مندی)

در این مرحله نیازمندی‌های معماری جمع‌آوری و مدل می‌شوند. نیازمندی‌های معماری مطابق شکل ۲ در قالب موضوع^۲، ابتکار^۳، اپیک^۴ (مجموعه حماسی) و استوری^۵ (داستان کاربری) طبقه‌بندی و مدل‌سازی می‌شوند [۵۵]. یک موضوع یک محدوده بزرگی از معماری است. یک ابتکار مجموعه‌ای از اپیک‌های مرتبط با هم است که به یک هدف مشترک منتهی می‌شوند. یک اپیک یک قطعه نیازمندی بزرگ و سطح بالاست. یک استوری هم یک درخواست و یا یک نیازمندی کوتاه است که توسط ذینفعان معماری مطرح می‌شود. به ازای هر استوری لازم است موارد زیر مشخص شوند:

- تعیین اولویت و اهمیت توسط ذینفعان (کم، خیلی کم، متوسط، زیاد، خیلی زیاد)
- تعیین میزان جزئیات مورد نیاز برای مدل‌سازی توسط ذینفعان (کم، خیلی کم، متوسط، زیاد، خیلی زیاد)
- تعیین نوع جمع‌آوری (مصاحبه، مشاهده، جلسات گروهی، طوفان فکری، نمونه‌سازی اولیه، نظرسنجی)
- تعیین نوع مستند کردن (در ابزار تالیف/مدل‌سازی، در قالب فایل کامپیوتری، نوشتن/ترسیم روی تخته وایت‌برد و تهیه عکس، نوشتن/ترسیم روی کاغذ و تهیه عکس)

۳-۳- مرحله مدل‌سازی سطح بالای معماری

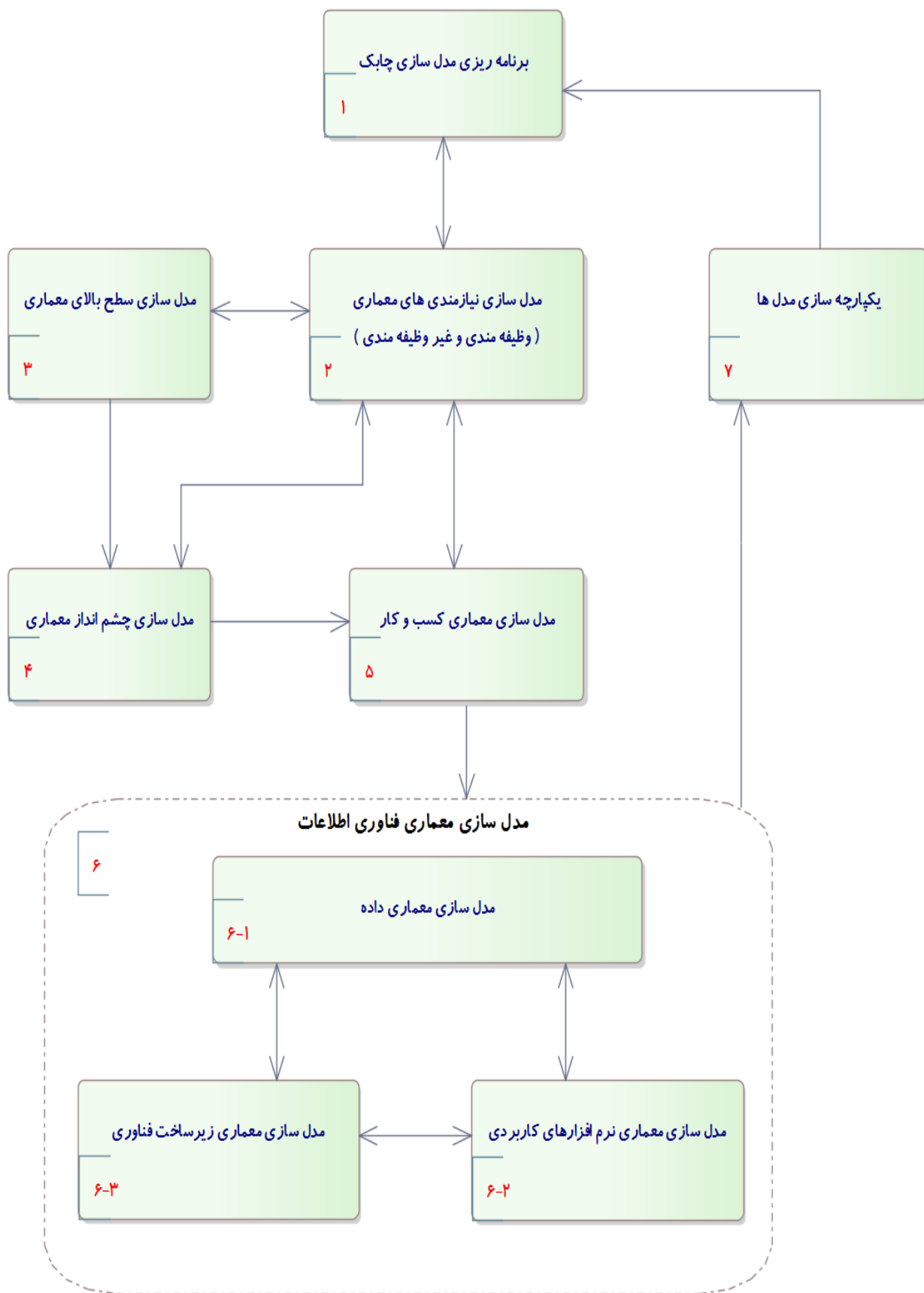
در این مرحله مدل‌سازی سطح بالای معماری با تعریف لایه‌ها و مراحل کلان معماری انجام می‌شود. لایه‌ها معماری شامل لایه چشم‌انداز و استراتژی، لایه کسب‌وکار، لایه اطلاعات و داده، لایه نرم‌افزارهای کاربردی، لایه فناوری و لایه مدیریت نیازمندی‌ها هستند.

- در شکل ۱ روش پیشنهادی و مراحل انجام آن نشان داده شده است. هر مرحله شامل اهداف، ورودی‌ها، فعالیت‌ها و خروجی‌هایی است. مراحل روش پیشنهادی و گام‌های هر مرحله به شرح زیر هستند:

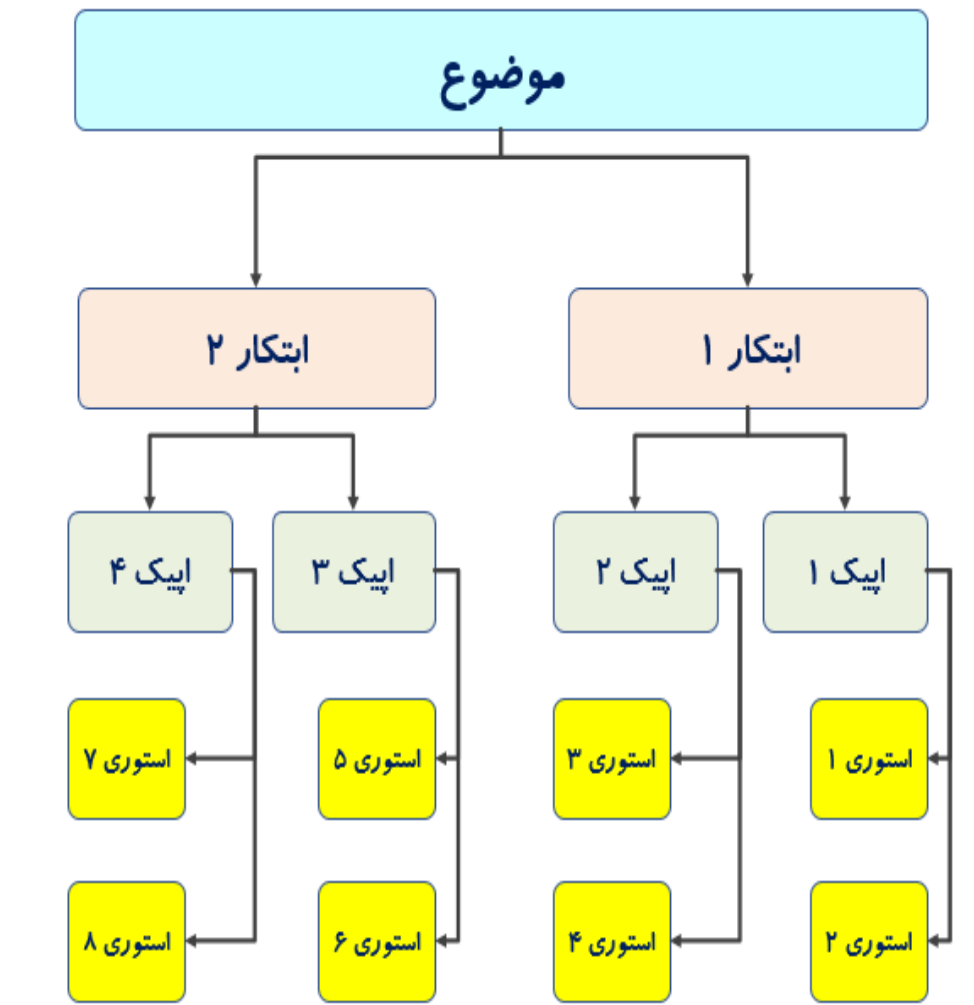
۳-۱- مرحله برنامه‌ریزی مدل‌سازی چابک

در این مرحله برنامه‌ریزی چابک انجام می‌شود و شامل گام‌هایی به شرح زیر است:

- تاکید و تمرکز بر معماری وضعیت مطلوب (مدل‌سازی وضعیت موجود در صورت نیاز و ضرورت انجام می‌شود).
- تعیین گلوگاه‌های^۱ سازمان و اولویت‌بندی آنها و تمرکز بر مدل‌سازی آنها متناسب با نظرات، نیازها و دغدغه‌های ذینفعان (گلوگاه‌های سازمان نقاط ضعیف، پرمخاطره و آسیب‌پذیر سازمان هستند).
- تعیین محدوده معماری
- برنامه‌ریزی تکرار و انتشار (تعریف تکرارها و فرسنگ شمارها)
- تعیین و انتخاب ابزارهای مناسب و ساده برای مدل‌سازی متناسب با نظرات، نیازها و دغدغه‌های ذینفعان
- تعیین استانداردهای مدل‌سازی و عناصر مورد نیاز آنها



شکل ۱. روش پیشنهادی و مراحل انجام آن



شکل ۲. ساختار مدل‌سازی نیازمندی‌ها

- مدل‌سازی اصول، محدودیت‌ها و ارزش‌ها
- مدل‌سازی دستاوردهای سازمانی
- مدل‌سازی عوامل تاثیرگذار بر روند معماری سازمانی
- مدل‌سازی منابع مورد نیاز

۳-۵- مرحله مدل‌سازی معماری کسب‌وکار

در این مرحله با توجه به نیازمندی‌های جمع‌آوری شده، تمرکز بر مدل‌سازی کنشگران، کارکردها، فرآیندها و سرویس‌های کسب‌وکار و ارتباطات و تعاملات بین آن‌ها است. در صورت ناقص بودن نیازمندی‌ها برای مدل‌سازی این مرحله می‌توان به مرحله مدل‌سازی نیازمندی‌های معماری برگشت، و اطلاعات و نیازمندی‌های تکمیلی را جمع‌آوری نمود.

- مدل‌کردن کنشگران کسب‌وکار و ساختار سازمانی
- مدل‌کردن کارکردهای کسب‌وکار و دسته‌بندی و ارتباطات آنها

۳-۴- مرحله مدل‌سازی چشم‌انداز معماری

در این مرحله با توجه به نیازمندی‌های جمع‌آوری شده، تمرکز بر مدل‌سازی اهداف و مدل‌سازی ذینفعان است. در صورت ناقص بودن نیازمندی‌ها برای مدل‌سازی این مرحله می‌توان به مرحله مدل‌سازی نیازمندی‌های معماری برگشت و اطلاعات و نیازمندی‌های تکمیلی را جمع‌آوری نمود.

- مدل‌کردن اهداف شامل: مدل بیانیه چشم‌انداز، مدل بیانیه ماموریت، مدل اهداف کوتاه مدت و بلند مدت و راهبردها و تاکتیک‌ها
- مدل‌سازی ذینفعان و مشخصات، نیازمندی‌ها و دغدغه‌های آنها
- در صورت نیاز ذینفعان به جزئیات بیشتر برای مدل‌سازی چشم‌انداز موارد دیگر نیز مدل می‌شوند، شامل:
- مدل‌سازی سیاست‌های سازمان
- مدل‌سازی پیشران‌های داخلی و خارجی

۳-۶-۲- مرحله مدل‌سازی معماری نرم‌افزارهای کاربردی

- مدل‌کردن کردن نرم‌افزارهای کاربردی و عملکرد و وقایع هر نرم‌افزار و تعاملات و ارتباطات بین آنها
- در صورت نیاز ذینفعان به مدل‌سازی جزئیات معماری نرم‌افزارهای کاربردی موارد دیگر نیز مدل می‌شوند، شامل:
- مدل‌کردن جزئیات نرم‌افزارهای شی‌گرا و تعیین کلاس‌ها، مولفه‌ها و اشیاء و تعیین ارتباطات و تعاملات و متدهای مربوطه
- مدل‌سازی جزئیات معماری نرم‌افزارهای سرویس‌گرا و تعیین سرویس‌ها و مولفه‌ها و ارتباطات و تعاملات و متدهای مربوطه
- مدل‌سازی جزئیات معماری نرم‌افزارهای عامل‌گرا و تعیین عامل‌ها و مولفه‌ها و ارتباطات و تعاملات و متدهای مربوطه

۳-۶-۳- مرحله مدل‌سازی معماری زیرساخت فناوری

- مدل‌سازی سرویس‌های زیرساختی و ارتباطات و تعاملات بین آنها
- مدل‌سازی فناوری‌ها و استانداردهای زیرساختی
- مدل‌سازی نودها، دستگاه‌ها و تجهیزات شبکه‌ای و ارتباطات و تعاملات بین آنها
- در صورت نیاز ذینفعان به مدل‌سازی جزئیات معماری نرم‌افزارهای کاربردی موارد دیگر نیز مدل می‌شوند، شامل:
- مدل‌کردن وقایع زیرساختی
- مدل‌سازی جزئیات مراکز داده‌ای و اتاق‌های سرور

۳-۷- مرحله یکپارچه‌سازی مدل‌ها

در این مرحله مدل‌های تهیه شده در مراحل قبل در قالب بلوک‌ها یا بسته‌های مرتبط با هم با قابلیت استفاده مجدد ادغام و یکپارچه می‌شوند. سپس مدل‌های یکپارچه شده در مخزن معماری درج و طبقه‌بندی می‌شوند. مدل‌های زاید و غیرقابل استفاده نیز از مخزن معماری حذف می‌شوند.

۴- روش ارزیابی

در این مقاله ارزیابی به روش ترکیبی انجام می‌شود. مرسوم است که راه‌حل‌های ارائه شده در حوزه معماری سازمانی به خصوص در حوزه مدل‌سازی معماری سازمانی از طریق ارائه مطالعات موردی، بررسی و ارزیابی شوند. انجام یک مطالعه موردی^۱ از

- مدل‌کردن فرآیندهای کسب‌وکار و دسته‌بندی و ارتباطات آنها
- مدل‌کردن سرویس‌های کسب‌وکار و دسته‌بندی و ارتباطات آنها
- مدل‌کردن ارتباطات بین کنشگران، کارکردها، فرآیندها و سرویس‌ها
- در صورت نیاز ذینفعان به جزئیات بیشتر برای مدل‌سازی معماری کسب‌وکار موارد دیگر نیز مدل می‌شوند، شامل:
- مدل‌سازی نقش‌های کسب‌وکار
- مدل‌سازی رابط‌های کسب‌وکار
- مدل‌سازی وقایع کسب‌وکار
- مدل‌سازی محصولات
- مدل‌سازی قراردادها
- مدل‌سازی جزئیات فرآیندهای کسب‌وکار و گردش کار و تعیین فعالیت‌ها، توالی، شروط، وقایع و نقش‌های مربوطه
- مدل‌سازی قوانین و تصمیم‌های کسب‌وکار مرتبط با هر فرآیند و تعیین سرویس‌های تصمیم‌گیری، جداول تصمیم، داده‌های ورودی، پایگاه دانش قوانین و تصمیم‌ها

۳-۶- مرحله مدل‌سازی معماری فناوری اطلاعات

در این مرحله با توجه به نیازمندی‌های جمع‌آوری شده، مدل‌سازی معماری فناوری اطلاعات انجام می‌شود. در صورت ناقص بودن نیازمندی‌ها برای مدل‌سازی این مرحله می‌توان به مرحله مدل‌سازی نیازمندی‌های معماری برگشت و اطلاعات و نیازمندی‌های تکمیلی را جمع‌آوری نمود. این مرحله خود شامل سه زیر مرحله است:

۳-۶-۱- مرحله مدل‌سازی معماری داده

- مدل‌کردن معماری داده از طریق کلاس‌ها و موجودیت‌ها و بسته‌بندی آنها و تعیین صفات و متدهای هر کلاس و ارتباطات بین کلاس‌ها
- مدل‌کردن ارتباط کلاس‌ها با فرآیندها و کارکردهای کسب‌وکار
- در صورت نیاز ذینفعان به مدل‌سازی جزئیات معماری نرم‌افزارهای کاربردی موارد دیگر نیز مدل می‌شوند، شامل: مدل‌سازی جداول پایگاه داده

مشاهده، مصاحبه حضوری، جلسات گروهی و نظرسنجی استفاده شد. مطابق شکل ۳ و با توجه به راه‌حل پیشنهادی، نیازمندی‌های حوزه بازاریابی و فروش غیرحضوری محصولات و خدمات مخابراتی استخراج و مدل شده‌اند. نیازمندی‌های وظیفه‌مندی و غیر وظیفه‌مندی تفکیک شده‌اند و به ازای هر کدام استوری‌های مربوطه استخراج و شماره گذاری شده‌اند. امنیت و دسترس پذیری نیازمندی‌های غیروظیفه‌مندی و فروش محصولات خانگی و تجاری مخابراتی به شکل غیرحضوری نیازمندی‌های وظیفه‌مندی استخراج شده در تکرارهای برنامه‌ریزی شده هستند. استوری‌های شماره ۲، ۸ و ۹ توسط ذینفعان با اولویت بالا شناخته شده و نیاز به جزئیات مدل‌سازی برای آنها اعلام شده است که در شکل ۳ مشخص شده‌اند.

• مرحله مدل‌سازی سطح بالای معماری

در این مرحله، لایه‌های کلان معماری شامل لایه چشم‌انداز و استراتژی، لایه کسب‌وکار، لایه داده، لایه نرم‌افزارهای کاربردی، لایه فناوری و لایه مدیریت نیازمندی‌ها مطابق شکل ۴ مدل گردیدند.

• مرحله مدل‌سازی چشم‌انداز معماری

در این مرحله، با توجه به راه‌حل پیشنهادی و بر اساس مستند استراتژیکی سازمان، مدل چشم‌انداز معماری شامل بیانیه چشم‌انداز، بیانیه مأموریت، هدف، مقصد، استراتژی‌ها و تاکتیک مطابق شکل ۵ تدوین گردید. برای مدل‌سازی بیانیه چشم‌انداز، بیانیه مأموریت، هدف و مقصد از عنصر Goal استاندارد ArchiMate و برای مدل‌سازی استراتژی‌ها و تاکتیک‌ها نیز از عنصر Outcome استاندارد ArchiMate استفاده شد. ضمن اینکه برای دسته‌بندی نیز از عناصر Vision، Mission، Goal، Objective، Strategy و Tactic استاندارد BMM استفاده شد. برای مدل‌سازی ذینفعان نیز از عنصر Stakeholder استاندارد ArchiMate و برای مدل کردن دغدغه‌های ذینفعان از عنصر Driver استاندارد ArchiMate استفاده شد. از بکارگیری سایر عناصر استانداردهای ArchiMate و BMM صرف نظر شد. متن توضیحات مدل چشم‌انداز معماری کوتاه در نظر گرفته شده و از بیان توضیحات زیاد صرف نظر گردید.

• مرحله مدل‌سازی معماری کسب‌وکار

در این مرحله، با توجه به روش پیشنهادی و نیازمندی‌های

طریق راه‌حل ارائه شده نشان خواهد داد که روش ارائه شده تا چه حد، می‌تواند در یک محک واقعی کاربردپذیری خود را بروز دهد. مطالعه موردی طبق روش مدل‌سازی معماری سازمانی چابک ارائه شده در بخش ۳ در یک سازمان ایرانی انجام می‌شود. از آنجاییکه روش کیفی از طریق ارائه مطالعه موردی لازم بوده ولی کافی نیست لذا با استفاده از روش فرآیند تحلیل سلسله مراتبی (AHP) که یکی از روش‌های تصمیم‌گیری چند معیاره است، ارزیابی کمی طبق نظر خبرگان انجام می‌شود [۵۶] و [۵۷].

مراحل قابل انجام با استفاده از روش AHP عبارتند از:

- تعیین تعدادی شاخص تاثیرگذار در مدل‌سازی معماری سازمانی چابک با استفاده از منابع کتابخانه‌ای و نظر خبرگان
- تعیین هدف تصمیم‌گیری
- انجام مقایسات زوجی شاخص‌ها توسط خبرگان
- انجام محاسبات وزنی برای تعیین اولویت شاخص‌ها
- ادغام وزن‌های نسبی به منظور رتبه‌بندی شاخص‌ها
- محاسبه نرخ ناسازگاری جهت تعیین قابل قبول بودن مقایسات

۵- ارزیابی و تجزیه و تحلیل داده‌ها

۵-۱- ارزیابی کیفی

مطالعه موردی بر اساس روش پیشنهادی در شرکت مخابرات ایران انجام شده که خلاصه گزارش آن به شرح زیر است:

• مرحله برنامه‌ریزی مدل‌سازی چابک

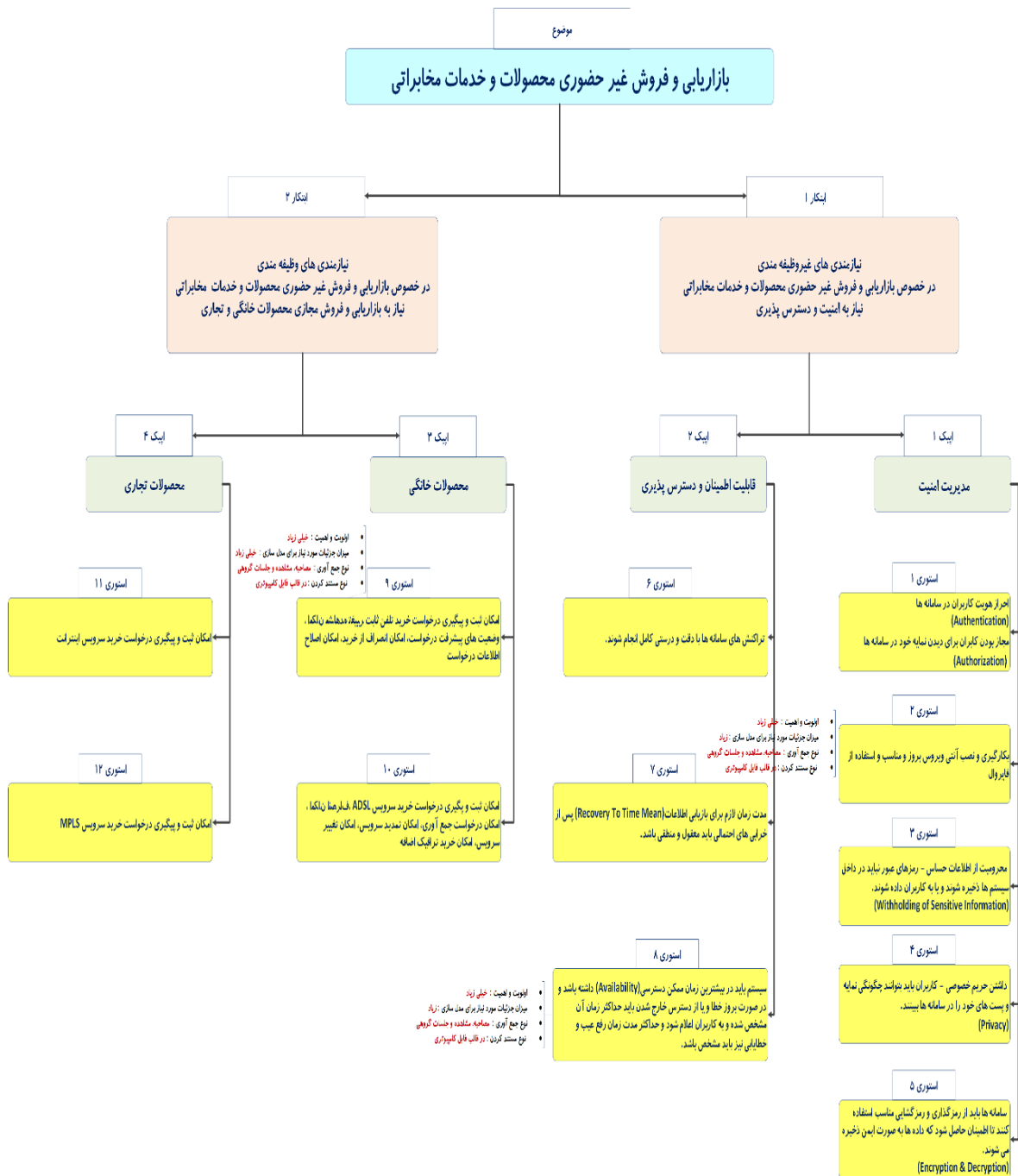
برای انجام معماری به وضعیت مطلوب^۱ آن توجه شده و از وضعیت موجود^۲ صرف نظر گردیده است. اداره تحول و توسعه پرتال‌های سازمانی که یکی از گلوگاه‌های سازمان است به عنوان نقطه شروع و محدوده اولیه معماری انتخاب گردید. سه تکرار پانزده روزه با هدف تولید و بروزرسانی مدل‌های معماری وضعیت مطلوب به صورت تکاملی و تدریجی تعریف شدند. طبق نظر و توافق ذینفعان برای تولید مدل‌ها از ترکیب دو نرم‌افزار MS Visio و Enterprise Architect استفاده گردید.

• مرحله مدل‌سازی نیازمندی‌های معماری (وظیفه‌مندی و غیروظیفه‌مندی)

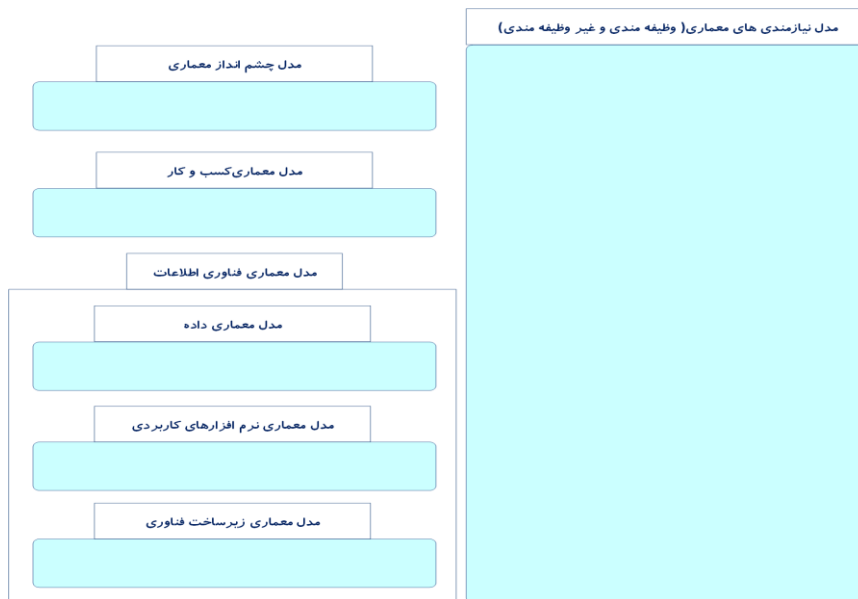
در این مرحله، برای جمع‌آوری نیازمندی‌های معماری اعم از وظیفه‌مندی و غیر وظیفه‌مندی از روش‌های نمونه‌سازی اولیه،

با توجه به نیاز و دغدغه‌های ذینفعان در خصوص مدل‌سازی جزئیات «تکمیل فرم درخواست» مطابق شکل ۷ مدل‌سازی جزئیات فرآیند کسب‌وکار آن با عناصر استاندارد BPMN شامل: Pool, Lane, Activity, Gateway, Event و DataStore انجام گردید. از بکارگیری سایر عناصر استاندارد BPMN صرف نظر گردید.

جمع‌آوری شده مدل معماری کسب‌وکار تدوین گردید. برای مدل‌سازی کسب کار از عناصر کسب‌وکار استاندارد ArchiMate شامل: Business Actor, Business Interface, Business Business, Business Process, Business Function, Event Role و Business Collaboration استفاده گردیده و از بکارگیری سایر عناصر این استاندارد صرف نظر گردید. مدل معماری کسب‌وکار در شکل ۶ نمایش داده شده است.



شکل ۳. مدل نیازمندی‌های معماری



شکل ۴. مدل سطح بالای معماری

بیانیه چشم انداز (Vision Statement)

تبدیل شدن به بهترین اپراتور کشور ایران در خصوص ارائه محصولات و خدمات مخابراتی غیر حضوری

بیانیه مأموریت (Mission Statement)

نهایت تلاشمان را می کنیم تا به مشتریان خود پایین ترین قیمت ممکن، بهترین حق انتخاب و حداکثر راحتی را ارائه دهیم

هدف (Goal)

ارائه تمام و کمال محصولات و خدمات مخابراتی تا یک سال آینده به شکل غیر حضوری

مقصد (Objective)

شناسایی و تحلیل کلیه محصولات و خدمات مخابراتی مورد نیاز مشتریان تا شش ماه آینده

استراتژی (Strategy)

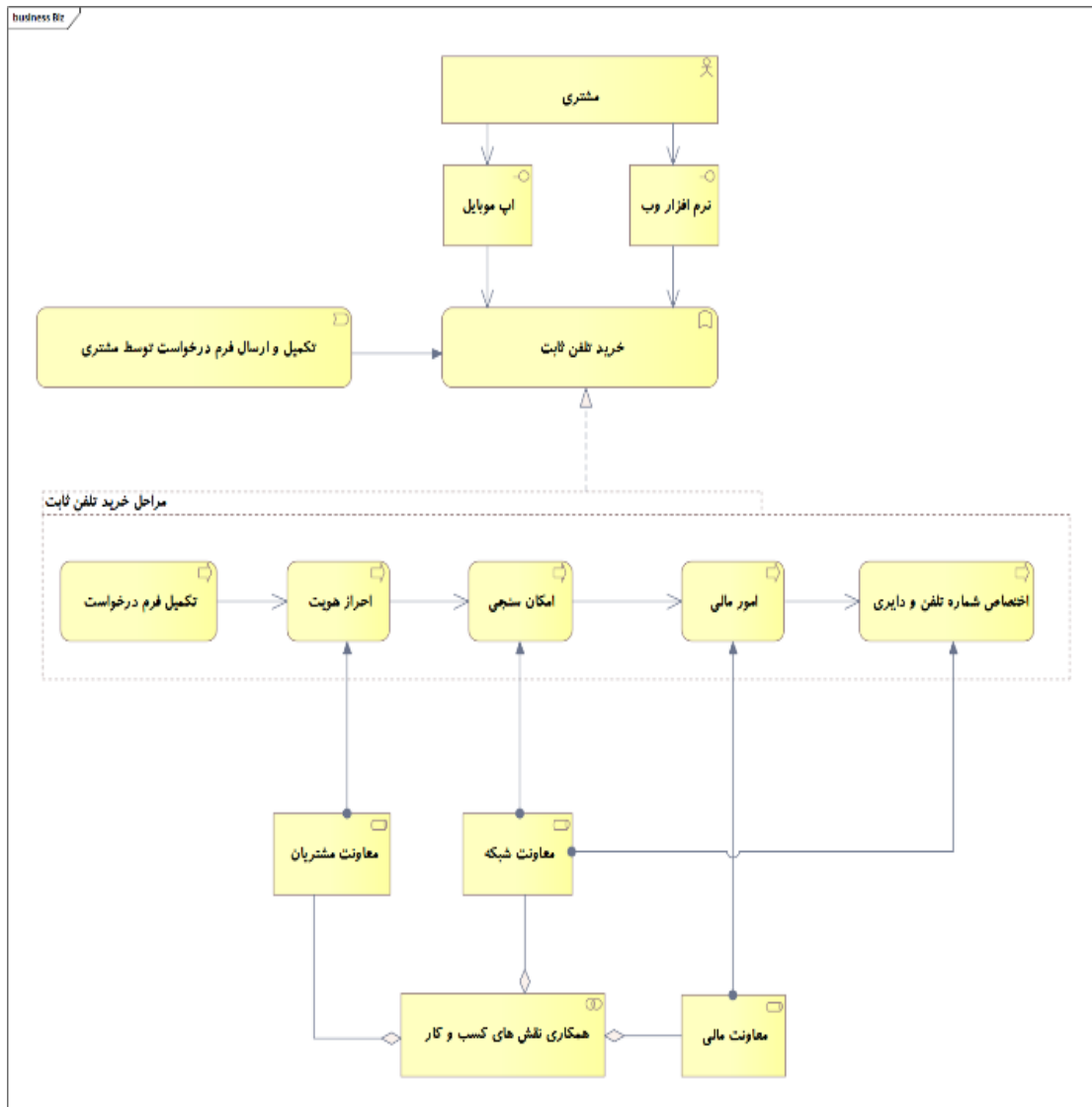
نقوذ در بازارهای نوظهور محصولات و خدمات مخابراتی
افزایش کیفیت ارائه خدمات و محصولات مخابراتی غیر حضوری با مناسب ترین قیمت برای مشتری

تاکتیک (Tactic)

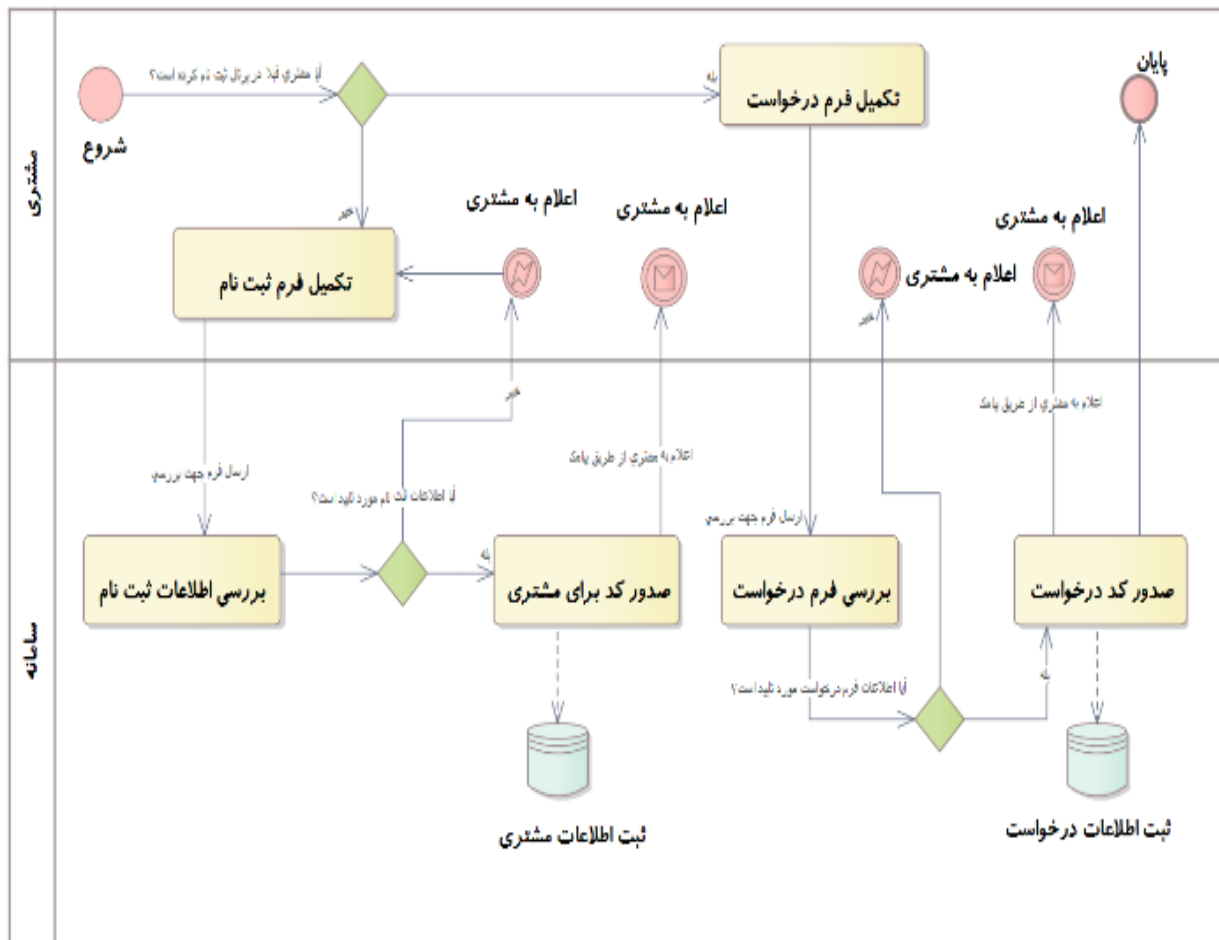
ارائه دسترسی آسان و کم هزینه به همه مشتریان مخابرات اعم از خانگی و تجاری
بکارگیری ابزارها و روش های نوین فناوری اطلاعات جهت ارائه محصولات و خدمات مخابراتی به شکل غیر حضوری



شکل ۵. مدل چشم انداز معماری



شکل ۶. مدل معماری کسب‌وکار

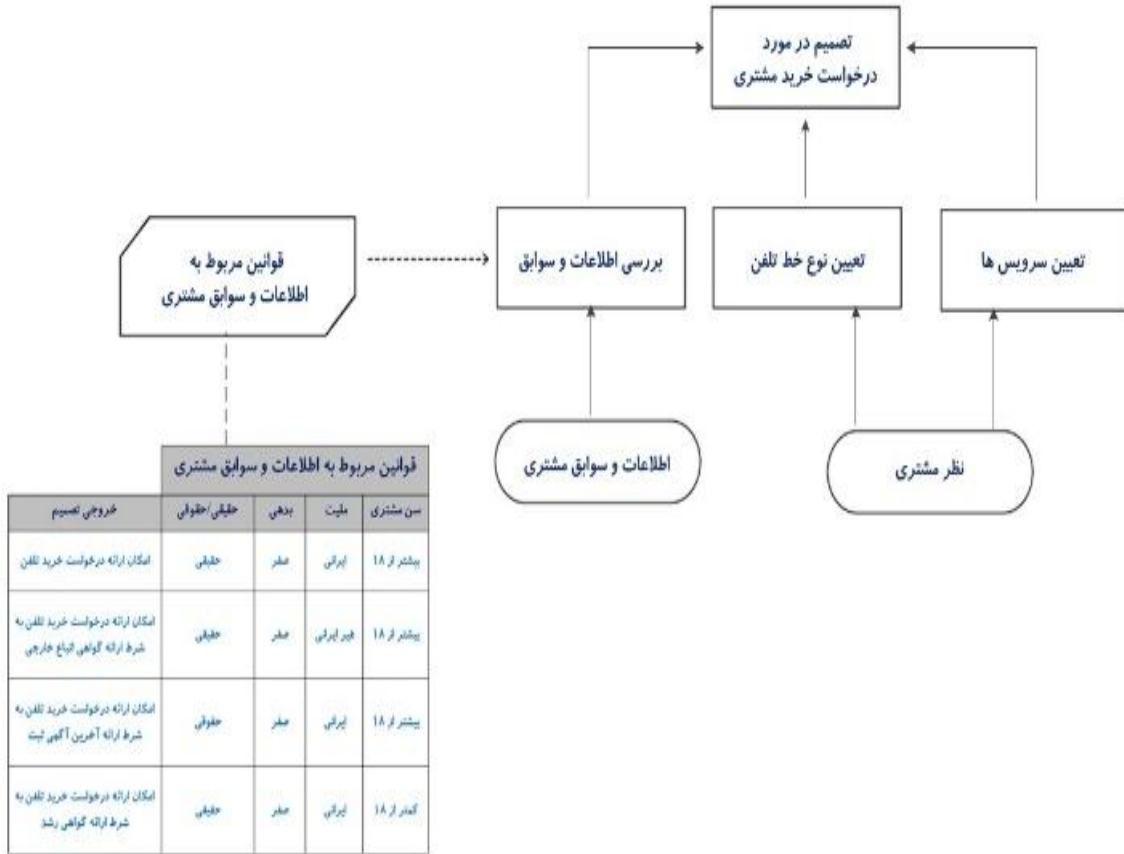


شکل ۷. مدل جزئیات کسب‌وکار فرآیند «تکمیل فرم درخواست» از خرید تلفن ثابت

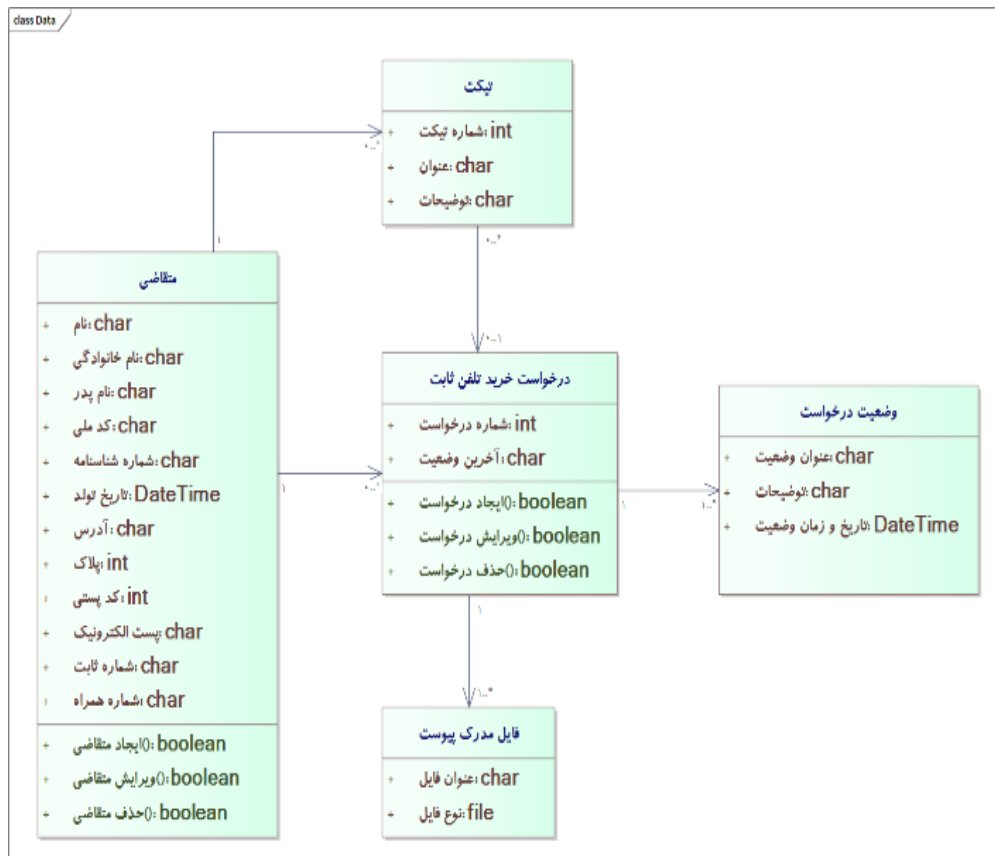
این مرحله یکی از مراحل مدل‌سازی معماری فناوری اطلاعات است. در این مرحله با توجه به روش پیشنهادی و نیازمندی‌های معماری با استفاده از عنصر Class استاندارد UML مدل معماری داده ترسیم گردید. عنصر Class استاندارد UML به پیمانکاران کمک می‌کند که به سرعت بتوانند مدل پایگاه داده‌ای و شیء‌گرا را ایجاد کرده و توسعه دهند. مدل معماری داده در شکل ۹ نشان داده شده است.

با توجه به شکل ۷ فعالیت «بررسی فرم درخواست» دارای قوانین و تصمیم‌هایی است که با توجه به نیاز و دغدغه‌های ذینفعان با جزئیات مدل گردید. برای مدل‌سازی آن از عناصر استاندارد DMN شامل: Input Data, Business Knowledge, Decision استفاده گردید. در شکل ۸ مدل جزئیات قوانین کسب‌وکار فعالیت «بررسی فرم درخواست» نشان داده شده است. از بکارگیری سایر عناصر استاندارد DMN صرف نظر گردید.

- مرحله مدل‌سازی معماری داده



شکل ۸. مدل قوانین و تصمیم‌های بررسی فرم درخواست از فرآیند خرید تلفن ثابت



شکل ۹. مدل معماری داده

این مرحله نیز یکی از مراحل مدل‌سازی معماری فناوری اطلاعات است. در این مرحله با توجه به روش پیشنهادی و نیازمندی‌های معماری مدل معماری زیرساخت فناوری تدوین گردید. برای مدل‌سازی این لایه از عناصر استاندارد ArchiMate شامل: Device، CommunicationNetwork، TechnologyInterface، SystemSoftware و Grouping استفاده گردید. از بکارگیری سایر عناصر استاندارد ArchiMate صرف نظر گردید. مدل معماری زیرساخت فناوری در شکل ۱۱ نمایش داده شده است.

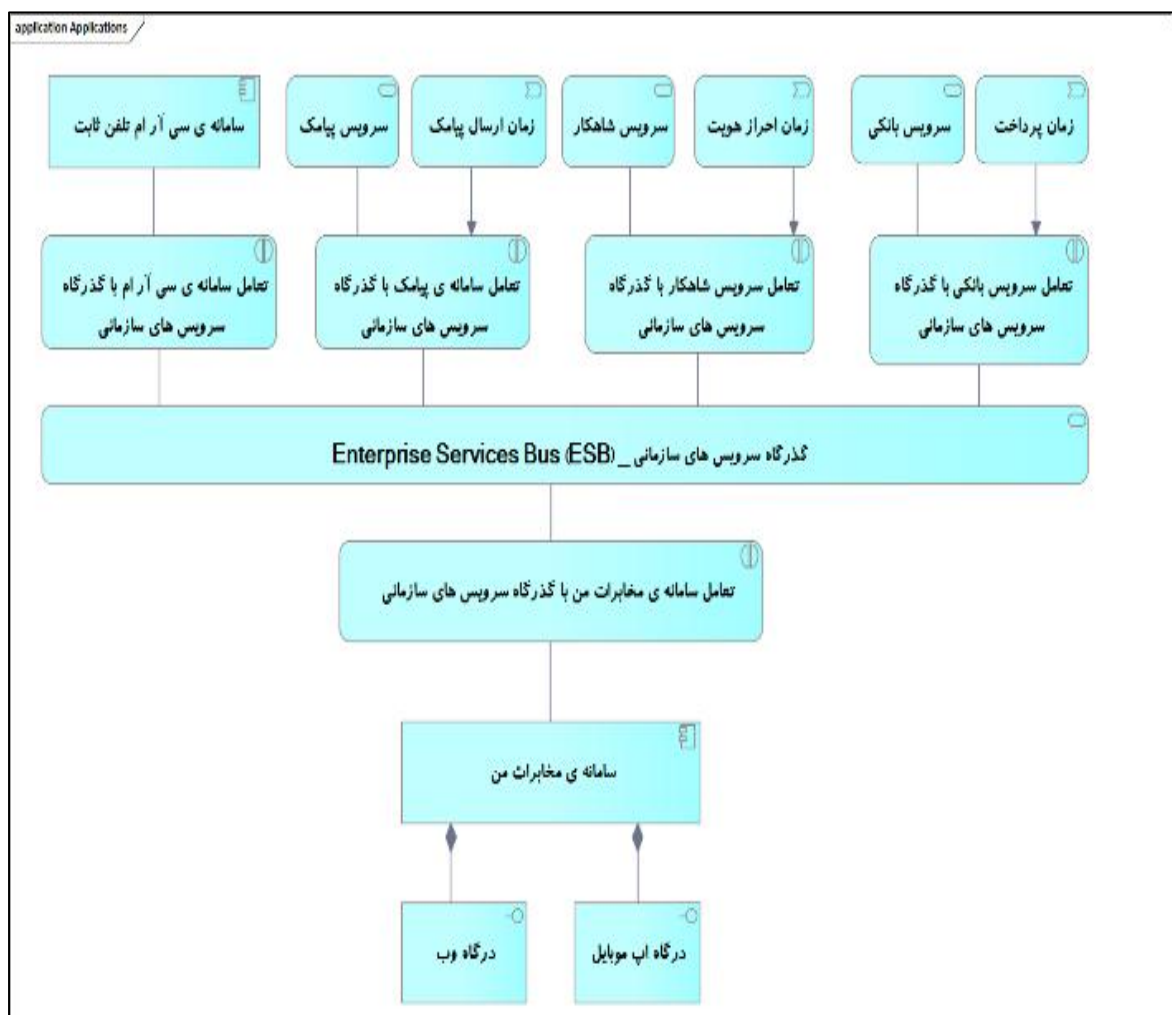
• مرحله یکپارچه سازی مدل‌ها

در این مرحله، مدل‌های تولید شده با هم در قالب شکل ۴ ادغام گردیدند.

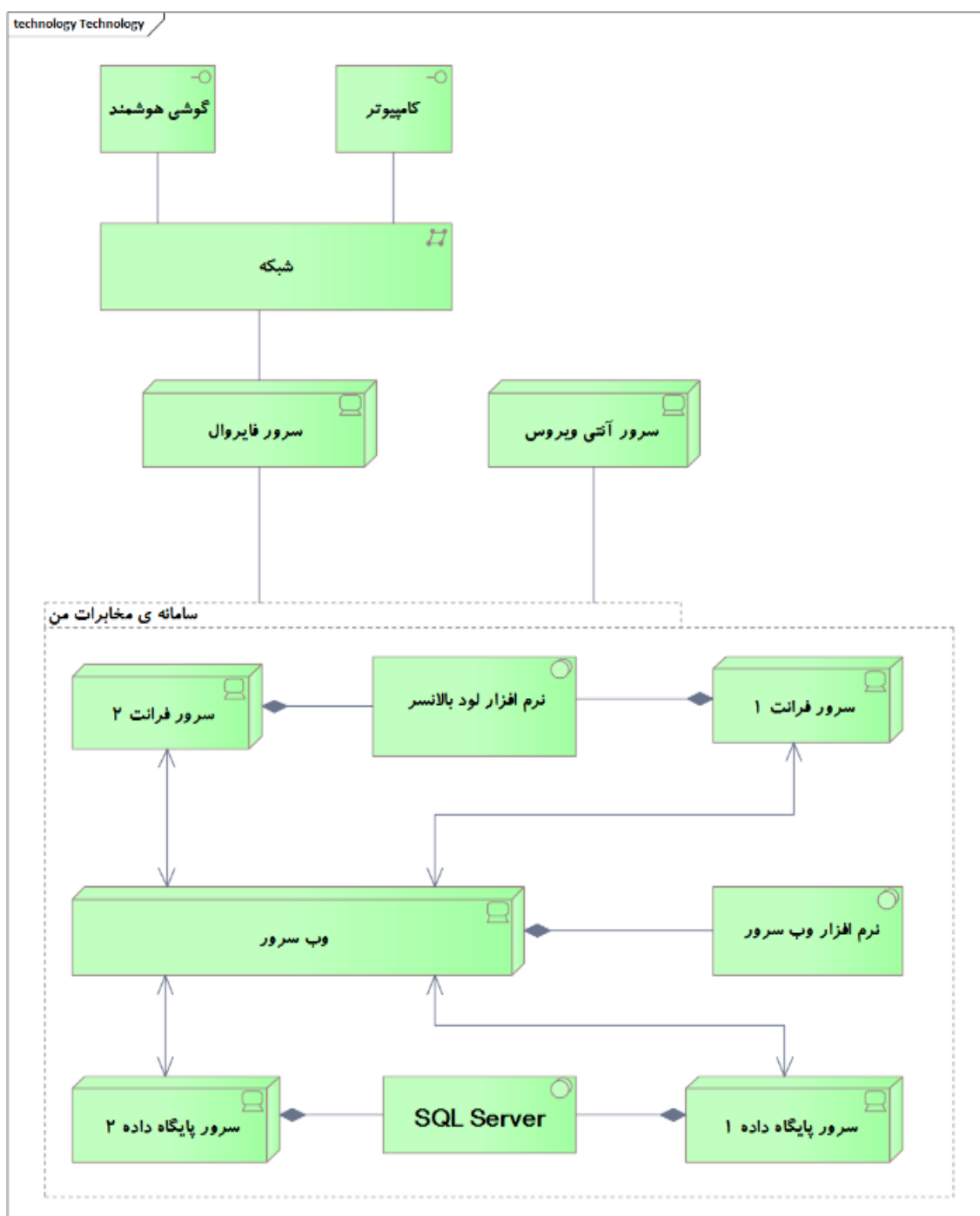
• مرحله مدل‌سازی معماری نرم‌افزارهای کاربردی

این مرحله نیز یکی از مراحل مدل‌سازی معماری فناوری اطلاعات است. در این مرحله با توجه به روش پیشنهادی و نیازمندی‌های معماری مدل معماری نرم‌افزارهای کاربردی تدوین گردید. برای مدل‌سازی این لایه از عناصر استاندارد ArchiMate شامل: ApplicationComponent، ApplicationService، ApplicationEvent، ApplicationInteraction و ApplicationInterface استفاده گردید. از بکارگیری سایر عناصر استاندارد ArchiMate صرف نظر گردید. مدل معماری نرم‌افزارهای کاربردی در شکل ۱۰ نمایش داده شده است.

• مرحله مدل‌سازی معماری زیرساخت فناوری



شکل ۱۰. مدل معماری نرم‌افزارهای کاربردی



شکل ۱۱. مدل معماری زیرساخت فناوری

۵-۲- ارزیابی کمی

جامعه آماری تحقیق محدود و شامل دوازده خبره در دسترس مدل‌سازی معماری سازمانی چابک است که به روش نمونه‌گیری گلوله‌برفی و روش نمونه‌گیری قضاوتی انتخاب گردیدند. ابزار جمع‌آوری داده‌ها، پرسشنامه و مصاحبه با خبرگان هستند. طبق منابع کتابخانه‌ای و تایید خبرگان، شش شاخص به شرح زیر استخراج گردیدند:

۱. تولید فرآورده‌های مورد نیاز و ضروری [۱] و [۴] و [۵]
 ۲. کاهش زمان مدل‌سازی [۱] و [۳]
 ۳. کاهش هزینه [۳]
 ۴. بهبود رضایت ذینفعان [۱] و [۳]
 ۵. بهبود کاربردپذیری [۲۳]
 ۶. افزایش سادگی [۱] و [۲]
- هدف این است که ببینیم از نظر خبرگان، راه‌حل پیشنهادی هر کدام از شاخص‌های ارزیابی را با چه وزن و رتبه‌ای محقق نموده

نسبت به هم نشان داده شده است.

در مرحله بعد مقایسات زوجی انجام شده توسط خبرگان با روش میانگین هندسی ادغام شدند که حاصل آن در جدول ۲ نمایش داده شده است.

در مرحله بعد نرمال‌سازی انجام شده است. برای نرمال‌سازی مقایسات زوجی هر درایه را بر مجموع درایه‌های ستونش تقسیم می‌کنیم. ماتریس نرمال شده در جدول ۳ نمایش داده شده است.

است. برای انجام محاسبات از نرم‌افزارهای اکسل^۱ و سوپردسیژن^۲ استفاده شد. ابتدا مقایسات زوجی شاخص‌های ارزیابی توسط خبرگان انجام گردید. این کار با انجام مقایسات دو به دو بین عناصر تصمیم (مقایسه زوجی) و از طریق تخصیص امتیازات عددی که نشان دهنده ارجحیت یا اهمیت بین دو عنصر تصمیم است، صورت می‌گیرد. برای انجام این کار معمولاً از مقایسه گزینه‌ها با شاخص‌های i ام نسبت به گزینه‌ها یا شاخص‌های j ام استفاده می‌شود که در جدول ۱ نحوه ارزش‌گذاری شاخص‌ها

جدول ۱. طیف ۹ تایی عبارات کلامی روش AHP

ارزش ترجیحی	وضعیت مقایسه i نسبت به j	توضیح
۱	اهمیت برابر	گزینه یا شاخص i نسبت به j اهمیت برابر دارند و یا ارجحیتی نسبت به هم ندارند.
۳	نسبتاً مهم‌تر	گزینه یا شاخص i نسبت به j کمی مهم‌تر است.
۵	مهم‌تر	گزینه یا شاخص i نسبت به j مهم‌تر است.
۷	خیلی مهم‌تر	گزینه یا شاخص i دارای ارجحیت خیلی بیشتری از j است.
۹	کاملاً مهم	گزینه یا شاخص i از j مهم‌تر و قابل مقایسه با j نیست.
۲ و ۴ و ۸		ارزش‌های میانی بین ارزش‌های ترجیحی را نشان می‌دهد.

جدول ۲. مقایسات زوجی شاخص‌ها

افزایش سادگی	بهبود کاربردپذیری	بهبود رضایت ذینفعان	کاهش هزینه	کاهش زمان مدل‌سازی	تولید فرآورده‌های مورد نیاز و ضروری
۱,۷۳۰	۱,۲۷۹	۱,۱۷۹	۱,۸۱۷	۱,۷۹۰	۱
۱,۰۱۵	۰,۷۹۸	۰,۹۸۵	۰,۹۱۳	۱	۰,۵۵۹
۰,۹۴۹	۰,۸۹۶	۱,۴۹۰	۱	۱,۰۹۶	۰,۵۵۰
۱,۲۰۷	۱,۰۱۵	۱	۰,۶۷۱	۱,۰۱۵	۰,۸۴۸
۱,۲۴۹	۱	۰,۹۸۵	۱,۱۱۶	۱,۲۵۳	۰,۷۸۲
۱	۰,۸۰۱	۰,۸۲۸	۱,۰۵۴	۰,۹۸۵	۰,۵۷۸

جدول ۳. مقایسات نرمال

افزایش سادگی	بهبود کاربردپذیری	بهبود رضایت ذینفعان	کاهش هزینه	کاهش زمان مدل‌سازی	تولید فرآورده‌های مورد نیاز و ضروری
۰,۲۴۲	۰,۲۲۱	۰,۱۸۲	۰,۲۷۷	۰,۲۵۱	۰,۲۳۲
۰,۱۴۲	۰,۱۳۸	۰,۱۵۲	۰,۱۳۹	۰,۱۴۰	۰,۱۲۹
۰,۱۳۳	۰,۱۵۵	۰,۲۳۰	۰,۱۵۲	۰,۱۵۴	۰,۱۲۷
۰,۱۶۹	۰,۱۷۵	۰,۱۵۵	۰,۱۰۲	۰,۱۴۲	۰,۱۹۷
۰,۱۷۵	۰,۱۷۳	۰,۱۵۲	۰,۱۷۰	۰,۱۷۶	۰,۱۸۱
۰,۱۴۰	۰,۱۳۸	۰,۱۲۸	۰,۱۶۰	۰,۱۳۸	۰,۱۳۴

تجربه نشان داده است که اگر نرخ ناسازگاری کمتر از ۰/۱۰ باشد سازگاری مقایسه‌ها قابل قبول بوده و در غیر این صورت مقایسه‌ها باید تجدید نظر شوند. مراحل زیر برای محاسبه نرخ ناسازگاری انجام گردیدند:

- ماتریس مقایسه‌های زوجی در بردار ستونی «وزن نسبی» ضرب گردیده و بردار جدیدی با عنوان بردار مجموع وزنی^۱ ایجاد گردید.
- عناصر بردار مجموع وزنی بر بردار اولویت نسبی تقسیم گردیده و بردار سازگاری^۲ ایجاد گردید.
- میانگین عناصر بردار سازگاری محاسبه و بر اساس آن پارامتر λ_{max} به دست آمد.
- شاخص سازگاری طبق رابطه ۱ محاسبه گردید. n تعداد شاخص‌ها بوده که برابر ۶ می‌باشد.

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (1)$$

- نسبت سازگاری از تقسیم شاخص سازگاری بر شاخص تصادفی^۳ طبق رابطه ۲ به دست آمد. شاخص تصادفی از جدول ۵ استخراج گردید.

$$CR = \frac{CI}{RI} \quad (2)$$

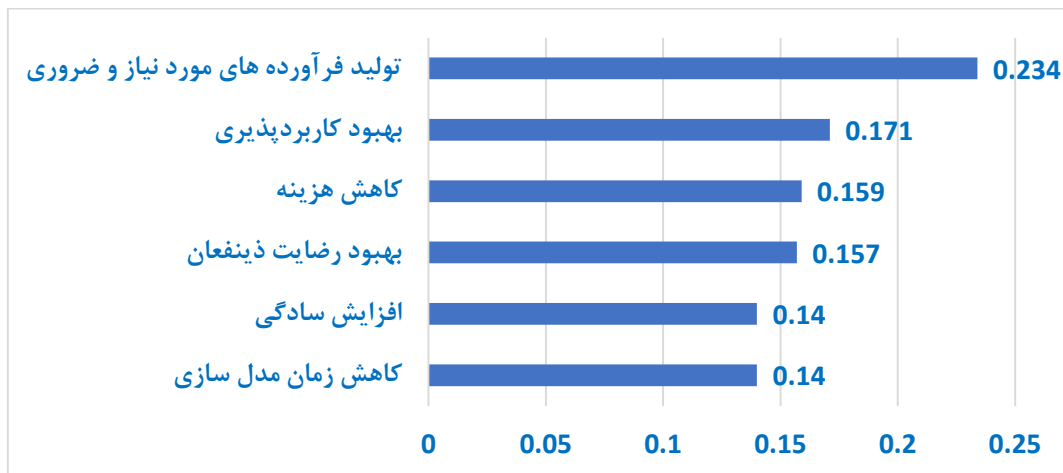
در مرحله بعد وزن شاخص‌ها محاسبه گردید. برای محاسبه وزن شاخص‌ها از درایه‌های جدول ۳ به صورت سطری میانگین حسابی گرفته شد. نتیجه در جدول ۴ نمایش داده شده است.

با توجه به جدول ۴، شاخص تولید فرآورده‌های مورد نیاز و ضروری با وزن ۰,۲۳۴ اولویت اول را کسب کرده است. شاخص‌های بهبود کاربردپذیری و کاهش هزینه به ترتیب با اوزان ۰,۱۷۱ و ۰,۱۵۹ اولویت‌های دوم و سوم را کسب کرده‌اند. در شکل ۱۲ وزن شاخص‌های ارزیابی نمایش داده شده است.

در مرحله آخر نرخ ناسازگاری مقایسه‌های زوجی جهت تعیین قابل قبول بودن مقایسه‌ها محاسبه گردید. وقتیکه تعداد مقایسه‌ها افزایش یابد اطمینان از سازگاری آن‌ها به راحتی میسر نبوده و باید با به کارگیری نرخ سازگاری به این اعتماد دست یافت.

جدول ۴. وزن شاخص‌ها

شاخص	وزن
تولید فرآورده‌های مورد نیاز و ضروری	۰,۲۳۴
کاهش زمان مدل‌سازی	۰,۱۴۰
کاهش هزینه	۰,۱۵۹
بهبود رضایت ذینفعان	۰,۱۵۷
بهبود کاربردپذیری	۰,۱۷۱
افزایش سادگی	۰,۱۴۰



شکل ۱۲. وزن شاخص‌های ارزیابی

جدول ۵. شاخص سازگاری تصادفی (RI)

n	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
RI	۰	۰,۵۸	۰,۹	۱,۱۲	۱,۲۴	۱,۳۲	۱,۴۱	۱,۴۵	۱,۴۹	۱,۵۱	۱,۴۸	۱,۵۶	۱,۵۷	۱,۵۹

3 Random Index = RI

1 Weighted Sum Vector = WSV

2 Consistency Index = CI

دینفعان مدل می‌گردند و در غیر این صورت از مدل‌سازی آنها صرف نظر می‌شود. در مرحله مدل‌سازی معماری داده، تاکید بر مدل‌سازی کلاس‌ها و جداول داده‌ای است تا به سرعت مورد بهره‌برداری پیمانکاران، طراحان و برنامه‌نویسان قرار بگیرد. در مرحله مدل‌سازی معماری نرم‌افزارهای کاربردی تمرکز بر مدل‌کردن نرم‌افزارهای کاربردی و عملکرد و وقایع هر نرم‌افزار و تعاملات و ارتباطات بین آنها است. در صورت نیاز دینفعان جزئیات معماری‌های شیء‌گرا، سرویس‌گرا و عامل‌گرا مدل می‌شوند و در غیر این صورت از مدل‌سازی جزئیات مربوطه صرف نظر می‌شود. در مرحله مدل‌سازی معماری زیر ساخت فناوری نیز تمرکز بر مدل‌سازی سرویس‌های زیرساختی، نودها، دستگاه‌ها و تجهیزات شبکه‌ای و ارتباطات و تعاملات بین آنها است و سایر موارد در صورت نیاز دینفعان مدل می‌شوند. در نهایت در مرحله آخر مدل‌های تولید شده یکپارچه می‌گردند. در بخش چهارم روش ترکیبی (کیفی + کمی) برای ارزیابی روش پیشنهادی مطرح گردید. روش کیفی از طریق ارائه یک مطالعه موردی در یک سازمان ایرانی انجام پذیرفت. مدل‌های تولید شده به ازای مراحل روش پیشنهادی در شکل‌های ۳ تا ۱۱ نشان داده شده‌اند. در ارزیابی کمی با استفاده از روش AHP و طبق نظر دوازده خبره در دسترس، شش شاخص ارزیابی وزن‌دهی و اولویت‌بندی شدند. پرسشنامه مقایسات زوجی مطابق جدول ۱ توسط خبرگان تکمیل گردید. برای انجام محاسبات ارزیابی از نرم‌افزارهای اکسل و سوپردسیژن استفاده شد. بر این اساس و طبق نظر خبرگان شاخص‌های تولید فرآورده‌های مورد نیاز و ضروری، بهبود کاربردپذیری، کاهش هزینه، بهبود رضایت دینفعان، کاهش زمان مدل‌سازی و افزایش سادگی به ترتیب توسط روش پیشنهادی در این مقاله محقق شده‌اند. این اولویت‌بندی در شکل ۱۲ نمایش داده شده است. روش پیشنهادی دارای جامعیت بوده زیرا همه لایه‌های معماری سازمانی را تحت پوشش قرار داده است. ضمن اینکه طبق شکل ۱ مراحل انجام نیز مشخص شده‌اند.

محاسبات نهایی نرخ ناسازگاری طبق مراحل قبل در جدول ۶ نمایش داده شده است. نتایج نشان می‌دهد که نرخ ناسازگاری (RI) برابر با ۰,۰۰۹ بوده (که کمتر از ۰,۱ است) و نشان دهنده سازگاری قابل قبول مقایسه‌های زوجی انجام شده است.

۶- بحث و تحلیل

در معماری سازمانی چابک مدل‌ها و مستندات به اندازه و به مقدار کافی تولید می‌شوند. روش‌ها و چارچوب‌هایی برای مدل‌سازی معماری سازمانی چابک ارائه شده‌اند که هر کدام دارای نواقصی هستند. در این مقاله روشی برای مدل‌سازی معماری سازمانی چابک جهت پوشش نواقص روش‌های موجود ارائه گردید. مطابق شکل ۱ روش پیشنهادی و مراحل انجام آن پیشنهاد گردید. در روش پیشنهادی مدل‌سازی همه لایه‌های معماری در نظر گرفته شده است. مدل‌سازی بر وضعیت مطلوب و گلوگاه‌های سازمان متمرکز گردیده است. وضعیت موجود و سایر محدوده‌ها در صورت نیاز و دغدغه دینفعان مدل می‌شوند. در روش پیشنهادی، مدل‌سازی به شکل تدریجی و از طریق تعریف تکرارها و با بکارگیری ابزارهای ساده صورت می‌گیرد. انتخاب ابزارها مطابق نظر دینفعان معماری انجام می‌شود. طبق روش پیشنهادی و مطابق شکل ۲ نیازمندی‌های معماری با روش‌های سریع از جمله مصاحبه و مشاهده جمع‌آوری و در قالب موضوع، ابتکار، اپیک و استوری مدل می‌شوند. نیازمندی‌ها توسط دینفعان اولویت‌بندی شده و میزان جزئیات لازم برای مدل‌سازی آنها تعیین می‌شوند. در مرحله مدل‌سازی سطح بالای معماری، لایه‌های کلان معماری مدل می‌شوند. در مرحله مدل‌سازی چشم‌انداز معماری، تاکید بر مدل‌سازی اهداف و مدل‌سازی دینفعان قرار گرفت و سایر موارد در صورت نیاز دینفعان مدل می‌گردند و در غیر این صورت از مدل‌سازی آنها صرف نظر می‌شود. در مرحله مدل‌سازی معماری کسب‌وکار نیز تمرکز بر مدل‌سازی کنشگران، کارکردها، فرآیندها و سرویس‌های کسب‌وکار و ارتباطات بین آنهاست و سایر موارد در صورت نیاز

جدول ۶. نرخ ناسازگاری مقایسه‌های زوجی

شاخص	λ	λ_{max}	I.I	R.I	R.I.I
تولید فرآورده‌های مورد نیاز و ضروری	۶,۰۵۹	۶,۰۵۴	۰,۰۱۱	۱,۲۴	۰,۰۰۹
کاهش زمان مدل‌سازی	۶,۰۵۴				
کاهش هزینه	۶,۰۵۷				
بهبود رضایت دینفعان	۶,۰۴۱				
بهبود کاربردپذیری	۶,۰۵۳				
افزایش سادگی	۶,۰۵۹				

اولویت بندی نیازمندی ها، برنامه ریزی تکرار، استفاده از اپیک و استوری برای جمع آور نیازمندی ها، بکارگیری ابزارهای ساده برای مدل سازی و استفاده از روش های سریع جمع آوری نیازمندی ها تاکید شده است.

این مقاله همه ابعاد معماری سازمانی چابک پوشش قرار نمی دهد و لذا لازم است با سایر روش های چابک ترکیب گردد. روش ارائه شده در این مقاله به شکل ترکیبی (کیفی + کمی) مورد ارزیابی قرار گرفت. در روش کیفی مدل های نمونه برای لایه های مختلف معماری ترسیم و نشان داده شدند. نیازمندی های معماری با روش های سریع جمع آوری شده و به شکل تکرار پذیر در قالب موضوع، ابتکار، اپیک و استوری مدل گردیدند. برای مدل سازی از نرم افزارهای Enterprise Architect و MS Visio استفاده شد.

در روش کمی با استفاده از روش تصمیم گیری چند معیاره AHP و نرم افزارهای اکسل و سوپردسیژن ارزیابی انجام پذیرفت. تعداد شش شاخص چابکی با بررسی منابع کتابخانه ای استخراج و به تایید دوازده نفر خبره در دسترس مدل سازی معماری سازمانی چابک رسیدند. هدف از ارزیابی این بود که بینیم از نظر خبرگان، راه حل پیشنهادی هر کدام از شاخص های ارزیابی را با چه وزن و رتبه ای محقق نموده است. که بر اساس نتایج حاصل شده شش شاخص، تولید فرآورده های مورد نیاز و ضروری، بهبود کاربردپذیری، کاهش هزینه، بهبود رضایت ذینفعان، کاهش زمان مدل سازی و افزایش سادگی به ترتیب وزن دهی و اولویت بندی گردیدند. در نهایت پیشنهادات برای پژوهش های آتی عبارتند از:

- برای مدل سازی نیازمندی های معماری روش های دیگر از جمله نمودار مورد کاربری (Use Case Diagram) در استاندارد UML و نمودار نیازمندی (Requirement Diagram) از استاندارد SysML مورد بررسی قرار گرفته و با روش ارائه شده در این مقاله مقایسه گردند.
- مدل سازی پیشامدها در معماری کسب و کار به روش پیشنهادی در این مقاله اضافه شود. عناصر استاندارد CMMN شامل: Case Plan Model, Task, Stage, Milestone, Event, Case File Item, Discretionary و ... جهت پوشش پیشامدها مورد بررسی قرار گیرند.
- برای مدل سازی اهداف (Goals Modeling) روش های دیگر مورد بررسی قرار گرفته و با روش ارائه شده در این مقاله مقایسه گردند.
- کارایی و قابلیت اطمینان روش ارائه شده با روش هایی نظیر شبکه های پتری و نظریه صف ارزیابی گردند.

ارزیابی کیفی نشان می دهد که ضمن استفاده از استانداردهای مدل سازی نظیر ArchiMate, UML, BPMN, DMN و BMM از همه عناصر آنها استفاده نشده بلکه نوعی کمینه سازی در استفاده از عناصر اعمال شده است. بر این اساس همه فرآورده های معماری نیز تولید نشده اند بلکه طبق روش پیشنهادی و نظر ذینفعان مهمترین آنها تولید شده اند. در روش پیشنهادی بر استفاده از روش های عملی چابک نظیر برنامه ریزی تکرار، اولویت بندی نیازمندی ها، استفاده از اپیک و استوری برای مدل سازی نیازمندی ها، برنامه ریز تکرار و استفاده از ابزارهای ساده تاکید شده است. در مطالعه موردی انجام پذیرفته طبق نظر ذینفعان از ابزارهای مدل سازی EA و Visio استفاده شد که سبب افزایش سرعت و بهبود رضایت ذینفعان گردید زیرا این دو ابزار مورد تایید آنان بوده است. همچنین تاکید و تمرکز بر مدل سازی وضعیت مطلوب و گلوگاه های سازمان سبب افزایش سرعت مدل سازی و کاهش حجم فرآورده ها و هزینه های تولید مدل ها گردید. ارزیابی کمی نیز نشان می دهد که طبق نظر خبرگان روش پیشنهادی در تولید فرآورده های مورد نیاز و ضروری موفق بوده است. همچنین در خصوص سایر شاخص های ارزیابی مدل سازی معماری سازمانی چابک یعنی بهبود کاربردپذیری، کاهش هزینه های تولید و بروزرسانی مدل ها، بهبود رضایت ذینفعان، کاهش زمان مدل سازی و افزایش سادگی، روش پیشنهادی موفق عمل نموده است. بدیهی است روش پیشنهادی در این مقاله تمام ابعاد و ویژگی های یک متدولوژی چابک را تبیین نمی کند و لذا نیاز است با روش های چابک دیگر ترکیب و بهره برداری گردد.

۷- نتیجه گیری

در این مقاله روشی برای مدل سازی معماری سازمانی چابک ارائه گردید و به روش ترکیبی مورد بررسی و ارزیابی قرار گرفت. روش ها و چارچوب های ارائه شده برای مدل سازی معماری سازمانی چابک دارای نواقص هستند. در روش پیشنهادی در این مقاله هدف بر این بود که نواقص موجود پوشش داده شوند. راه حل ارائه شده در این مقاله مدل سازی همه لایه های معماری سازمانی را در نظر گرفته و به شکل تدریجی در قالب تکرارها، فرآورده های ضروری را تولید می نماید. در روش پیشنهادی این مقاله بر مدل سازی استاندارد تاکید شده ولی از همه عناصر آنها استفاده نمی شود. در روش پیشنهاد شده تمرکز مدل سازی بر وضعیت مطلوب و گلوگاه های سازمان است. وضعیت موجود معماری و نقاط عادی در صورت نیاز و دغدغه ی ذینفعان مدل می شوند. در روش ارائه شده بر بکارگیری تجارب عملی چابک، از جمله

مراجع

- [19] OMG, Business Motivation Model 1.3(BMM), 2015, <<http://www.omg.org/spec/BMM/>>.
- [20] OMG, Decision Model and Notation 1.4 (DMN), 2022, <<http://www.omg.org/spec/DMN/>>.
- [21] OMG, Case Management Model and Notation 1.1(CMMN), 2016, <<http://www.omg.org/spec/CMMN/>>.
- [22] OMG, Systems Modeling Language 1.5(SYSML), 2017, <<http://www.omg.org/spec/SysML/>>.
- [23] A.Q. Gill, "agile enterprise architecture modelling: Evaluating the applicability and integration of six modelling standards," Information and Software Technology, vol. 67, pp. 196-206, November 2015.
- [24] راضی، علی، رضایی، رضا، یزدان پناه، احمد علی، "مدل سازی معماری سازمانی چابک: ارزیابی کاربردپذیری شش استاندارد مدل سازی بر مبنای چارچوب ملی معماری سازمانی ایران"، دو فصلنامه علمی فناوری اطلاعات و ارتباطات ایران، شماره های ۴۷ و ۴۸، ۱۰۵_۱۳۵، تهران، بهار و تابستان ۱۴۰۰.
- [25] S. Yamamoto, Q. Zhia, Z. Zhoua, "Aspect Analysis towards ArchiMate Diagrams," Procedia Computer Science., vol. 159, pp. 973-980, 2019.
- [26] A. Buchalceva, "Using ArchiMate to model ISO/IEC 29110 standard for very small entities," Computer Standards & Interfaces., vol. 65, pp. 103-121, 2019.
- [27] A. Zrnc, M. Bajec, M. Krisper, "Enterprise modelling with UML," Elektrotehni ski vestnik University of Ljubljana., vol. 68, pp. 109-114, 2001.
- [28] F. Armour, S. H. Kaisler, J. Getter, D. Pippin, "A UML-driven Enterprise Architecture Case Study," Proceedings of the 36th Annual Hawaii International Conference, February, 2003.
- [29] H. Takeuchi, S. Yamamoto, "AI Service System Development Using Enterprise Architecture Modeling," Procedia Computer Science., vol. 159, pp. 923-932, 2019.
- [30] S. Yamamoto, "Analysis of Smart City Reference Architecture by ArchiMate," Procedia Computer Science., vol. 207, pp. 514-521, 2022.
- [31] A. Aldea, M.E. Iacob, J.V. Hillegersberg, D. Quartel, L. Bodenstaff, H. Franken, "Modelling strategy with ArchiMate," Proceedings of the 30th Annual ACM Symposium on Applied Computing, Apr 13-17, 2015, Salamanca, Spain.
- [32] F. C. Kitsios, M. Kyriakopoulou, M. Kamariotou, "Exploring Business Strategy Modelling with ArchiMate: A Case Study Approach," Department of Applied Informatics, University of Macedonia, 2022.
- [33] P. Bhattacharya, "Modelling Strategic Alignment of Business and IT through Enterprise Architecture: Augmenting ArchiMate with BMM," Procedia Computer Science., vol. 121, pp. 80-88, 2017.
- [34] The TOGAF® Standard, 10th Edition, 2023, <<https://www.opengroup.org/togaf/10thedition>>.
- [35] Open Agile Architecture, A Standard of The Open Group, 2020, <<https://pubs.opengroup.org/architecture/o-aa-standard-single>>.
- [36] شمس علیی، فریدون، مهجوریان، امیر و همکاران. چارچوب و روش شناسی معماری سازمانی ایران، نسخه ۱، شورای اجرایی (عالی) فناوری اطلاعات کشور، کمیسیون توسعه دولت الکترونیکی، تهران، <https://www.ieaf.ir/>، ۱۳۹۵.
- [1] Manifesto for Agile Software Development, 2001, <<https://agilemanifesto.org/>>.
- [2] Scott W. Ambler. Agile Enterprise Architecture, 2021, <<http://agiledata.org/essays/enterpriseArchitecture.html>>.
- [3] Scott W. Ambler. Agile Modeling: Effective Practices for extreme Programming and the Unified Process, Published by John Wiley & Sons, Inc., New York, 2002, <<http://msoo.pbworks.com/f/Scott+W.+Ambler+-+Agile+Modeling.pdf>>.
- [4] Scott W. Ambler. The Agile Modeling, 2022, <<https://agilemodeling.com>>.
- [5] Enterprise Architecture on a Page, 2023, <<http://eaonapage.com/>>.
- [6] P. Medeiros, A. Santana, M. Lima, H. Moura, M.s, "An Agile Approach for Modeling Enterprise Architectures," 23rd International Conference on Enterprise Information Systems, April 26-28, 2021.
- [7] M. Hauder, S. Roth, C. Schulz, F. Matthes, "Agile Enterprise Architecture Management an Analysis on the Application of Agile Principles," The Fourth International Symposium on Business Modeling and Software Design, June 24-26, 2014.
- [8] H. M. Shirazi, B. D. Rouhani, M. M. Shirazi, "A Framework for Agile Enterprise Architecture," International Journal of Intelligent Information Technology Application., vol. 2, Issue. 4, pp. 182-186, 2009.
- [9] Z. A. Amiri, "CHALLENGES AND WEAKNESSES OF AGILE METHOD IN ENTERPRISE ARCHITECTURE," International Journal of Computer Science & Engineering Survey (IJCSES), vol. 3, No. 6, pp. 37-45, 2012.
- [10] S. Buckl, F. Matthes, I. Monahov, S. Roth, C. Schulz, C. M. Schweda, "Towards an Agile Design of the Enterprise Architecture Management Function," 15th IEEE International Enterprise Distributed Object Computing Conference Workshops, Aug. 29 - Sept. 2, 2011, Helsinki, Finland.
- [11] A. Rasnacic, S. Berzisa, "Method for Adaptation and Implementation of Agile Project Management Methodology," Procedia Computer Science., vol. 104, pp. 43-50, 2017.
- [12] T. Kaddoumi, M. Watfa, "A Proposed Agile Enterprise Architecture Framework," Sixth International Conference on Innovative Computing Technology (INTECH), Aug. 24-26, 2016, Dublin, Ireland.
- [13] J. Werewka, A. Spiechowicz, "Enterprise Architecture Approach to SCRUM Processes, Sprint Retrospective Example," Computer Science and Information Systems., vol. ۱۱, pp. ۱۲۲۱-۱۲۲۸, ۲۰۱۷.
- [14] OMG, ArchiMate 3.2, 2022, <<https://pubs.opengroup.org/architecture/archimate3-doc/>>.
- [15] OMG, Unified Modeling Language 2.5 (UML), 2015, <<http://www.omg.org/spec/UML/>>.
- [16] OMG, Business Process Model and Notation 2.0.2 (BPMN), 2013, <<http://www.omg.org/spec/BPMN/index.htm>>.
- [17] G. Beydoun, G. Low, B. Henderson-Sellers, H. Mouratidis, J.J. Gomez-Snaz, J. Pavon, C. Gonzalez-Perez, "FAML: a generic metamodel for MAS development," IEEE Trans. Softw. Eng., vol. 35, Issue 6, pp. 841-863, Nov-Dec 2009.
- [18] OMG, Service Oriented Architecture Modeling Language 1.0.1 (SoaML), 2012, <<http://www.omg.org/spec/SoaML/>>.

- [۴۷] اسماعیلی، حمیدرضا، آل محمد، متین، اصول و مبانی شبکه‌های پتری، تهران، انتشارات موسسه فرهنگی هنری دیباگران تهران، چاپ اول، ۱۳۹۶.
- [۴۸] مدرس، محمد، تیموری، ابراهیم، نظریه صف، تهران، انتشارات دانشگاه علم و صنعت ایران، چاپ ششم، ۱۳۹۵.
- [49] F. Basile, P. Chiacchio, D.D. Grosso, "A two-stage modelling architecture for distributed control of real-time industrial systems: Application of UML and Petri Net," *Computer Standards & Interfaces*, vol. 31, Issue 3, pp. 528-538, March 2009.
- [50] T. Bouabana-Tebibel, S. H. Rubin, "An interleaving semantics for UML 2 interactions using Petri nets," *Information Sciences*, vol. 232, pp. 276-293, May 2013.
- [51] N. Aoumeur, "Stepwise rigorous development of distributed agile information systems: from UML-diagrams to component-based Petri Nets," *Enterprise Information Systems*, vol. 2, No. 2, pp 125-160, May 2008.
- [52] A. Alhroob, N. Yousef, "Transforming UML State Machine Diagram to High Level Petri Net Using Genetic Algorithm," *Lecture Notes on Software Engineering*, vol. 2, No. 3, January 2014.
- [53] S. Toghyani, A. Harounabadi, "Validation of enterprise architecture through colored Petri nets," *Management Science Letters*, vol. 5, pp. 311-320, March 2015.
- [54] P. Szwed, "Verification of ArchiMate Behavioral Elements by Model Checking," *IFIP International Conference on Computer Information Systems and Industrial Management*, September, 2015.
- [55] Scott W. Ambler. *User Stories: An Agile Introduction*, 2022, <<https://agilemodeling.com/artifacts/userstory.htm>>.
- [۵۶] محمدی لرد، عبدالمحمود، فرآیندهای تحلیل شبکه‌ای (ANP) و سلسله مراتبی (AHP) به همراه معرفی نرم‌افزار Super Decision، تهران، انتشارات البرز فر دانش، ۱۳۸۸.
- [۵۷] عطائی، محمد، تصمیم‌گیری چند معیاره، شاهرود، انتشارات دانشگاه صنعتی شاهرود، چاپ اول، ۱۳۸۹.
- [37] M. Pankowska, "Business to System Requirements Agile Mapping," 17th International Conference on e-Business and Telecommunications, July 8 - 10, 2020.
- [38] A. Sadovykh, P. Desfray, B. Elvesæter, A. Berre, E. Landre, "Enterprise architecture modeling with SoaML using BMM and BPMN - MDA approach in practice," *Computer Science, 6th Central and Eastern European Software Engineering Conference*, Oct 13-15, 2010, Moscow, Russia.
- [39] M. Pankowska, "Business Models in CMMN, DMN and ArchiMate language," *Procedia Computer Science*, vol. 164, pp. 11-18, 2019.
- [40] M.D. Leoni, P. Felli, M. Montali, "Integrating BPMN and DMN: Modeling and Analysis," *Journal on Data Semantics*, vol. 10, pp. 165-188, June, 2021.
- [41] P. Desfray, G. Raymond, *Modeling Enterprise Architecture with TOGAF® A Practical Guide Using UML and BPMN*. Morgan Kaufmann; 1st edition, 2014.
- [42] B. Ding, T. Wu, Y. Yang, L. Dou, T. Jin, "ArchiMate Customization and Architecture Repository Management Practices: for a Technology-Intensive Enterprise," *Journal of Physics*, vol. 1187, Issue. 4, 2019.
- [43] M. Kirikova, "Variable Contents of Enterprise Models," *Procedia Computer Science*, vol. 104, pp. 89-96, 2017.
- [44] F. Vernadat, "Enterprise modelling: Research review and outlook," *Computers in Industry*, vol. 122, 103265, November 2020.
- [۴۵] شمس، فریدون، رضوی داوودی، مهسا، بدیع، کامبیز، "ارائه روشی جهت ارزیابی ویژگی‌های کیفی معماری سازمانی مبتنی بر Fuzzy AHP، نشریه مدیریت فناوری اطلاعات، دوره ۲، شماره ۴، از صفحه ۷۹ تا صفحه ۹۸، تهران، بهار و تابستان ۱۳۸۹.
- [46] J. Lakhrouit, K. Ba"ina, "Evaluating enterprise architecture complexity using fuzzy AHP approach: Application to university information system," *IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, November 17-20, 2015.

The Local Strategic Model of Smart Police of The Islamic Republic of Iran

Ahmad Dolatkhan¹, Mohamad Reza Movahedi Sefat^{2*}, Mehrab Ramak³

¹ PhD Student, Supreme National Defense University, Tehran, Iran

² Associate Professor, Supreme National Defense University, Tehran, Iran

³ Supreme National Defense University, Tehran, Iran

Received: 17 March 2024, Revised: 15 April 2024, Accepted: 17 April 2024

Paper type: Research

Abstract

The security of any country is the highest capital and the most favorable point for the system and people. For integrated management and command and control over situations and issuing appropriate orders, it is necessary for an organization like the police to intelligently monitor all issues and create an intelligent command and control center to complete missions efficiently. One of the components of increasing security is people's cooperation with the police in order to scientifically detect crimes, as well as benefiting from the latest technologies at the organization level. Despite the wide role of information technology in Faraja, in order for Faraja to carry out the assigned missions and establish order and security, the police organization must be transformed from a technology-oriented to a smart police, so the purpose of this research is to "provide a model It is a native strategy of the smart police of the Islamic Republic of Iran. This research is applied-developmental in terms of type and descriptive-analytical in terms of method. The statistical population of this research is targeted among the experts whose common feature is management experience in smart projects, with expertise in the fields of systems design or information technology, a strategic job position, which is estimated at 50 people. The analysis of the information and data obtained from the study of the sources has been done in a qualitative way and the information obtained from the questionnaire has been analyzed using the method of structural equations with the partial least squares approach and the use of SmartPLS software. The results of the research showed that the local strategic model of Iran's intelligent police consists of six dimensions: "people", "sovereignty", "data", "platform and system", "process" and "technology" and 49 components.

Keywords: Smart Police, Technology, Strategic Model, Data.

* Corresponding Author's email: movahedi@sndu.ac.ir

الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران

احمد دولتخواه^۱، محمدرضا موحدی صفت^{۲*}، مهراب رامک^۳

^۱ دانشجوی دکتری، دانشگاه عالی دانشگاه عالی دفاع ملی، تهران، ایران

^۲ دانشیار و عضو هیئت علمی دانشگاه عالی دفاع ملی، تهران، ایران

^۳ دکتری مدیریت راهبردی امنیت فضای سایبر، دانشکده امنیت ملی، دانشگاه عالی دفاع ملی، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۱۲/۲۷ تاریخ بازبینی: ۱۴۰۳/۰۱/۲۷ تاریخ پذیرش: ۱۴۰۳/۰۱/۲۹

نوع مقاله: پژوهشی

چکیده

امنیت هر کشور بالاترین سرمایه و مطلوب ترین نقطه برای نظام و مردم می‌باشد. برای مدیریت یکپارچه و فرماندهی و اشراف بر موقعیت‌ها و صدور دستورهای مناسب نیاز است تا پلیس به صورت هوشمند ابعاد مختلف جرم‌ها را رصد و با بکارگیری تجهیزات هوشمند، فرآیندی هوشمند و کارکنانی هوشمند بهترین نتیجه در یک مأموریت حاصل گردد. یکی از مولفه‌های ارزشمند و پایدار جهت افزایش امنیت، همراهی مردم با پلیس در راستای کشف علمی جرایم و همچنین بهره‌مندی پلیس از جدیدترین فناوری‌ها در سطح سازمان می‌باشد. با وجود نقش گسترده فناوری اطلاعات در فراجا برای آنکه فراجا بتواند به نحو مطلوب مأموریت‌های محوله را به انجام رساند و نظم و امنیت را برقرار کند، باید سازمان پلیس از فناوری محوری به پلیس هوشمند تبدیل شود، لذا هدف این تحقیق «ارائه الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران» می‌باشد. به منظور دستیابی به الگوی تحقیق در ابتدا با انجام مطالعات کتابخانه‌ای و بررسی نظری‌ها، ابعاد و مولفه‌های الگو به شکل اولیه احصاء شد، سپس با مراجعه به خبرگان، نظرات آنها دریافت و الگوی تحقیق طراحی گردید. این پژوهش از لحاظ نوع، کاربردی-توسعه‌ای و از نظر نوع روش، توصیفی-تحلیلی است. جامعه آماری این تحقیق بصورت هدفمند از جمله خبرگانی هستند که ویژگی مشترکشان سابقه مدیریت در طرح‌های هوشمندسازی، دارای تخصص در حوزه‌های طراحی سیستم‌ها و یا فناوری اطلاعات، جایگاه شغلی با ماهیت راهبردی است که تعداد ۵۰ نفر برآورد شده است. تجزیه و تحلیل اطلاعات و داده‌های حاصل از مطالعه منابع، به روش کیفی انجام شده و اطلاعات حاصل از پرسشنامه با استفاده از روش معادلات ساختاری با رویکرد حداقل مربعات جزئی و بهره‌گیری از نرم‌افزار اسمارت پی ال اس مورد تحلیل قرار گرفته است. نتایج تحقیق نشان داد الگوی راهبردی بومی پلیس هوشمند ج.ا.ایران مشتمل بر: شش بعد: «افراد»، «حاکمیت»، «داده»، «سکو و سامانه»، «فرایند» و «فناوری» و ۴۹ مولفه می‌باشد.

کلیدواژگان: پلیس هوشمند، فناوری، الگوی راهبردی، داده.

* رایانامه نویسنده مسؤول: movahedi@sndu.ac.ir

۱- مقدمه

گسترده‌تری حوزه فعالیت نیروی انتظامی و نیازها و ضرورت‌های جدید در ساماندهی به فعالیت‌ها و پیشبرد اهداف سبب شده است که این سازمان ملزم به هوشمندسازی خود و به خصوص کارکنان خود باشد. هدف آن است که پلیس فناوری محور به پلیس هوشمند که تلفیقی از رویکردهای پلیسی و تکامل یافته رویکردهای قبلی است، تغییر رویکرد داده تا بتواند بر مجرمین، که در دهه اخیر با بهره‌گیری از فناوری‌های نوظهور و فضای مجازی بر ارتکاب جرایم مبتنی بر فناوری افزوده‌اند، چیره شده و نظم و امنیت مطلوب جامعه اسلامی را برقرار سازد. علاوه بر این هوشمندی و هوشمند شدن یک گرایش جهانی است و هر سازمان که خودش را با این گرایش منطبق سازد از صحنه رقابت خارج می‌گردد. پلیس کشورهای پیشرفته به این نکته مهم پی برده‌اند که بدون دستیابی به هوشمند در پلیس انجام مأموریت‌های محوله و برقراری نظم و امنیت چندان موفق نخواهند بود و هر یک از این کشورها در مرحله خاصی از دستیابی به پلیس هوشمند قرار دارند. در این راستا، در حوزه هوشمندسازی در فراجا اقداماتی انجام شده لیکن لازم است با انجام پژوهش‌های متعدد و ارائه الگوی راهبردی و برنامه‌ریزی کلان قدم اساسی و کامل‌تری در این مسیر برداشته شود [۴].

تغییرات ماهیت جرم مثلا جرایم سایبری باعث ایجاد چالش‌های جدیدی در کشور و به خصوص برای پلیس شده است. بالندگی در چنین محیطی تنها با تغییر و سازگاری با پویایی‌ها امکانپذیر است. حجم انبوه داده‌هایی که به واسطه فعالیت‌ها و عملیات پلیس به شکل روزافزون تولید شده پلیس را با مساله انباشت داده مواجه ساخته است، مطالبات عمومی مردم برای برخط بودن پلیس نیز از دیگر علل لزوم هوشمندسازی پلیس است. مسائل روز جامعه از جمله شبکه‌های اجتماعی و فضای مجازی نیاز به هوشمند شدن پلیس را دوچندان کرده است. جریان دانش به هوشمندی سازمان و کارکنان آن مربوط می‌شود و تقویت جریان دانش، موجب توسعه هوشمندی سازمان و کارکنان آن و درنهایت توسعه پلیس می‌گردد. هوشمند شدن پلیس به معنی انتظام‌بخشی و امنیت‌آفرینی مطلوب و پایدار است، از این رو هوشمندسازی پلیس هم برای حفظ اقتدار کارکنان انتظامی و هم برای ارتقای میزان رضایتمندی مردم ضروری می‌باشد. سیستم‌های هوشمند در پلیس، ساختارها و فرآیندهای هوشمندی هستند که قادرند محیط را پایش کرده، سیگنال‌های اطلاعاتی را منتقل کرده، تصمیم‌گیری درست و اقدام سریع و به‌موقع را به‌منظور ایجاد و تأمین انتظارات از پلیس را فراهم آورند. در پلیس هوشمند از فناوری و تحلیل استفاده می‌شود تا به شیوه راهبردی از فعالیت‌های پلیسی پشتیبانی به عمل آید [۵].

در عصر دانایی، هوشمندی یکی از الزامات غیرقابل انکار توسعه سازمان‌ها و نمایانگر کل دانشی است که یک سازمان از محیطی که در آن رقابت می‌کند، در اختیار دارد و همچنین به عنوان یکی از توانایی‌های برتر سازمان‌های امروزی موجب می‌شود، داده‌های محیط پیرامون سازمان‌ها سریعتر و با دقت بیشتر تجزیه و تحلیل شده و در مواقع مقتضی در دسترس تصمیم‌گیرندگان قرار گیرد. مفهوم هوشمندی سازمانی، پیامد مستقیم به کارگیری و مدیریت اثربخش سرمایه‌های دانشی به مدد زیرساخت‌های کارآمد و نیروی انسانی خلاق و نوآور است که از طریق ترکیب اثربخش منابع و امکانات موجود درون سازمان و محیط بیرونی با تکیه بر قابلیت‌ها و توانمندی‌های منابع انسانی، سازمان‌ها را قادر می‌سازد تا به نحو مؤثری در راستای مأموریت‌ها و اهداف خود گام بردارند [۱].

سازمان هوشمند باید به عنوان یک سیستم عمل کند که در آن تصمیم‌های مشارکتی اتخاذ شود، همکاری کارکنان و تیم‌سازی در آن یک اصل اساسی تلقی شود، زمینه تمرکززدایی فراهم شود تا یادگیری سازمانی و انسجام فرایندها صورت گیرد و در نهایت، سازمان توانایی تولید دانش و استفاده از آن برای سازگاری با محیط را داشته باشد. براساس نظر برخی از صاحب‌نظران، در سازمان هوشمند به طور همزمان به جنبه‌های نرم (فرایندهای هوشمندی منابع انسانی) و سخت (فرایند هوشمند کردن سازمان از جنبه ساختاری) هوشمندی توجه می‌شود. بر اساس نظر آنها، هوشمندی ساختاری شامل دو شاخص تمرکز بر فناوری اطلاعات و تمرکز بر گروه‌های خودگردان می‌شود. همچنین هوشمندی انسانی بر اساس دو شاخص مهارت‌های فردی و مهارت‌های اجتماعی مورد سنجش قرار می‌گیرد [۲].

به طور معمول، تاکنون هوشمندسازی در سه سطح هوش مصنوعی، سیستم‌های خبره و داده کاوی مورد توجه بوده‌اند که در سازمان‌های هوشمند، از همه این سطوح و قابلیت هم‌افزایی آن‌ها بهره‌برداری می‌شود. چنین سازمان‌هایی قادرند با کمترین منابع، بالاترین راندمان را در اختیار قرار دهند و فرایندهای کنترل ساده و سریع را فراهم آورند. هوشمندسازی پلیس به منزله فرآیندی است که طبق آن با استفاده از سامانه‌های اطلاعاتی و ابزارهای الکترونیکی پلیس، داده‌ها از نقاط مختلف جمع آوری شده و ضمن یکپارچه‌سازی و پاک‌سازی داده‌ها، با استفاده از شیوه‌های هوش مصنوعی مورد تجزیه تحلیل قرار می‌گیرند و از نتایج تحلیل‌ها برای اخذ تصمیم در سطوح مدیریتی و مأموریت‌های فراجا استفاده می‌گردد [۳].

و نتایج تحقیق دربرگیرنده شاخص‌های آسیب‌شناسی هوشمندسازی مرزهای جمهوری اسلامی ایران در هفت محور (۱) موانع کارکنان، (۲) موانع مدیران، (۳) موانع فناوری، (۴) موانع اقتصادی، (۵) موانع محیطی، (۶) موانع فرایندی و (۷) موانع ساختاری ارائه گردید [۷]. در تحقیقی به بررسی هوشمندسازی و تاثیر فناوری دیجیتال در حوزه‌های مدیریت منابع انسانی پرداختند. نتایج تحقیق حاکی از آن است که فناوری به بخش مهمی از دنیای معاصر تبدیل شده است، و مدیریت منابع انسانی به شکل دیجیتال و با استفاده از رسانه‌های الکترونیکی، موبایل، تجزیه و تحلیل و فناوری اطلاعات برای کارآمدتر کردن مدیریت منابع انسانی، در جهان رو به رشد امروز، امری ضروری است. از آنجاییکه تاثیر فناوری بر مدیریت منابع انسانی می‌تواند عمیق باشد، و انقلاب اینترنت روش انجام تجارت در سراسر جهان را تغییر داده است، در نتیجه تغییرات سریع فناوری به ذی‌نفعان در رشد و کارایی سازمان‌ها در بلند مدت کمک می‌کند [۸]. در تحقیقی با عنوان تاثیر کیفیت خدمات پلیس هوشمند بر مشارکت شهروندان در پیشگیری از توزیع مواد مخدر با هدف تبیین تاثیر کیفیت خدمات پلیس هوشمند بر مشارکت شهروندان در پیشگیری از توزیع مواد مخدر انجام پذیرفت. نتیجه پژوهش نتایج نشان داد یکپارچگی بین ابعاد کیفیت خدمات پلیس هوشمند و سهولت دسترسی شهروندان به اطلاعات مورد نیاز از طریق سامانه اینترنت پلیس و اطلاع رسانی برخط زمینه‌های مشارکت در برنامه‌های پیشگیری از توزیع مواد مخدر را افزایش می‌دهد [۹].

پژوهشی با عنوان «پلیس هوشمند» انجام داد. برای تحقق حاکمیت هوشمند باید روشهای متداول را به سمت ایجاد روش‌های خلاقانه با تغییر در کیفیت خدمات با مفهوم پلیس هوشمند حرکت داد و جهت تحقق بخشیدن به پلیس هوشمند به آمادگی از جنبه‌های مختلف، مانند ساختار و زیرساخت‌ها نیاز است [۱۰]. پژوهشی با عنوان "تحقیق در مورد یکپارچه مدل پلیسی هوشمند" انجام و نتایج نشان داد که مدل پلیس هوشمند به عنوان ایده اصلی، افراد، سیستم‌های اطلاعاتی، زیرساخت‌های جامعه، استانداردهای خط مشی و غیره را به عنوان عناصر تشکیل دهنده در نظر می‌گیرد [۱۱].

۳- مدل سازمان هوشمند دیلویت

آمادگی استفاده از هوش مصنوعی در کشورها و سازمان‌ها را می‌توان با شاخص‌های مختلفی سنجید که یک مدل مطرح برای این کار که

بسیاری از سازمان‌ها استقرار و پیاده‌سازی سازمان هوشمند را در جهت پیشبرد و بهبود تصمیم‌گیری قرار می‌دهند، پیاده‌سازی سازمان هوشمند و سطح‌بندی و میزان کیفیت استقرار آن‌ها یک چالش برای مدیریت است. ظهور فناوری‌های پیشرفته موجب افزایش پیچیدگی در سازمان‌ها و مدیریت‌شده و هر قدر مخاطرات بیشتر باشد، ضروری است که برای مدیریت کارآمد و اتخاذ سیاست‌های مواجهه با خطرات، نگاهی فراتر از فناوری وجود داشته باشد؛ بنابراین در کنار توسعه فناوری، توسعه مدیریت و حکمرانی در سازمان هوشمند امری ضروری است.

در صورت انجام این تحقیق مسیر حرکت پلیس در هوشمندسازی شفاف می‌گردد و بخاطر مشخص بودن مسیر حرکت انگیزه مدیران سطوح عالی سازمان برای حمایت از طرح‌های هوشمندسازی بیشتر خواهد شد. همچنین این تحقیق باعث راهبری آسان و نظام‌مند طرح‌های هوشمندسازی پلیس ج.ا.ایران خواهد شد.

در صورت عدم انجام این تحقیق، بدلیل فقدان راهبردهای مشخص پلیس در هدایت هوشمند منابع محدود سازمانی اعم از عوامل انسانی و تجهیزاتی دچار مشکل خواهد شد و با ادامه شرایط فعلی شاخص‌های جرم در جامعه روز بروز افزایش خواهد کرد و احساس امنیت در بین شهروندان کاهش یافته و جایگاه اجتماعی پلیس بعنوان نهادی از حاکمیت دچار مشکل خواهد شد.

حال که الزام وجود الگوی راهبردی برای پلیس تشریح شد در بخش دوم پیشینه تحقیقات انجام شده بر روی موضوع مربوطه مرور خواهد شد. در بخش سوم الگوهای سازمان هوشمند ارائه و در بخش چهارم فناوری‌های نوین در راستای تحقق پلیس هوشمند بررسی می‌گردد. بخش پنجم الگوی راهبردی پلیس هوشمند ارائه و روش تحقیق انجام شده در بخش ششم مدل مفهومی تحقیق ارائه می‌گردد. روش انجام تحقیق در بخش هفتم، تحلیل و تفسیر نتایج در بخش هشتم ارائه خواهد شد. جمع‌بندی از دستاوردهای تحقیق در بخش نتیجه‌گیری ارائه شده است.

۲- پیشینه پژوهش

تحقیقی با عنوان الگوی پلیس هوشمند در فرماندهی انتظامی جمهوری اسلامی ایران انجام و نتایج نشان داد که تحقق پلیس هوشمند در گرو برنامه‌ریزی راهبردی برای چهار بُعد «کارکنان، فرایند، فناوری، ساختار» و ۱۷ مؤلفه و ۶۱ شاخص است و به‌کارگیری و هماهنگ‌سازی ابعاد و مؤلفه‌ها و شاخص‌های احصاشده نقش اساسی در تحقق پلیس هوشمند دارند [۶]. تحقیقی با عنوان آسیب‌شناسی هوشمندسازی مرزهای جمهوری اسلامی ایران انجام

مناسب جهت عملیاتی کردن مزایای هوش مصنوعی، شامل موارد مربوط به ارائه‌دهندگان، تعامل‌پذیری و محیط محاسباتی است. راه‌های مختلفی برای دنبال کردن هوش مصنوعی وجود دارد که بسته به شرایط زیرساخت‌ها و مالکیت فناوری (مثلاً اینکه درون سازمانی باشد یا شراکتی)، با هم تفاوت دارند. اما در همه موارد، هوش مصنوعی به یک رویکرد منسجم نیاز دارد تا تمامی نیازمندی‌های آینده خود را در مقیاس سازمانی و کاربردش در نظر بگیرد.

اخلاقیات: سازوکاری در نظر بگیرید تا بتوان تعصب و جانبداری هوش مصنوعی را درک کرد و از آن جلوگیری کرد، انصاف و شفافیت را ترویج دهید و مطمئن شوید که ارزش‌های درستی در ابتکارات هوش مصنوعی گنجانده شده است. در حالیکه گسترش هر فناوری باید اخلاقی باشد، هوش مصنوعی مواردی همچون شفافیت، حریم خصوصی و جانبداری را به مرکز توجه آورده است.

۴- فناوری‌های نوین در تحقق پلیس هوشمند

پلیس به عنوان یکی از نهادهای پیشرو در حوزه فناوری در اغلب کشورها هستند که همواره با بهره‌گیری از فناوری‌های نوین تلاش در هوشمندسازی سازمان خود و افزایش توان عملیاتی و تحلیلی دارند. در این بخش فناوری‌های مورد استفاده در حوزه هوشمندسازی پلیس بررسی شده‌اند.

پردازش تصویر: اکثر داده‌های دریافتی و مورد تحلیل در سازمان پلیس از نوع ویدئو و تصویر می‌باشد. برای کشف علمی جرایم نیاز است تا به درستی سیستم‌های هوشمند طراحی و آموزش داده شده تا با استفاده از پردازنده‌های قوی بتوان اطلاعات مهمی را از تصاویر استخراج نمود. تشخیص و تطبیق چهره، پلاک خوانی و تشخیص خودرو، تحلیل رفتار افراد و غیره تعدادی از خروجی‌های سیستم‌های بینایی ماشین می‌باشد [۱۳].

پرنده بدون سرنشین: طبق مطالعه سال ۲۰۲۰ توسط موسسه بارد، در مجموع ۱۵۷۸ آژانس امنیت عمومی ایالتی و محلی دارای پهپاد در ایالات متحده هستند. این مطالعه همچنین نشان می‌دهد که ۷۰ درصد از آژانس‌های امنیت عمومی با هواپیماهای بدون سرنشین در اجرای قانون کار می‌کنند. دپارتمان‌های پلیس از هواپیماهای بدون سرنشین برای کمک به واکنش سریع به حوادث، نقشه‌برداری از صحنه‌های جنایت، شناسایی مکان‌ها قبل از شروع عملیات، انجام ماموریت‌های نجات و اطمینان از ایمنی همه افراد درگیر در عملیات

توسط دیلویت^۱ ارائه شده است [۱۲]، شامل شش حوزه است که در زیر مرور شده‌اند. اگر یک سازمان تمایل به پیشرفت فراتر از نسخه‌های اولیه و آزمایشی دارد، در نظر گرفتن موارد زیر که در عین مجزا بودن به هم وابسته نیز هستند، می‌تواند به ارزیابی آمادگی جهت استفاده از هوش مصنوعی کمک نماید.

راهبرد: از آنجایی که هوش مصنوعی فناوری متحول‌کننده‌ای است، روی مسیر درست ماندن و سطح جاه‌طلبی بسیار مهم است. نیاز است چشم‌اندازی برای هوش مصنوعی مورد استفاده در نظر گرفته شود که هم راستا با اهداف سازمانی باشد. به این ترتیب می‌توان روشی جهت مدیریت قابلیت در دسترس خود در سراسر سازمان اتخاذ نمود در سازمان بزرگی مانند پلیس با تنوع ماموریت‌ها نیاز است راهبرد مربوط به هوش مصنوعی مطابق با اسناد بالادستی و نیازها استخراج گردد.

افراد: ممکن است سازمان‌ها با چالش‌هایی پیرامون دسترسی و جذب افراد با مهارت‌های فنی لازم و همچنین کمک به کارمندان فعلی در جهت توسعه و گسترش مهارت‌های مربوط به هوش مصنوعی، روبرو شوند. در این باره می‌توان، ادغام کار کارمندان با هوش مصنوعی، تجدید نظر در روش به کارگیری استعداد افراد و جلب مشارکت بیشتر سهامداران از طریق ارتباط موثر و تغییر مدیریت را در نظر داشت. در سازمان پلیس، کارکنان یکی از ارکان اصلی بوده که وجود مهارت می‌تواند موفقیت در ماموریت را تضمین نماید.

فرایندها: تعریف و طراحی فرایندها، کنترل‌گرها و فرماندهی سیستم برای پیاده‌سازی موفق هوش مصنوعی یک الزام است. در حالی که نسخه‌های اولیه و آزمایشی می‌تواند پتانسیل هوش مصنوعی را به اثبات برساند، اما ارزش واقعی آن زمانی مشخص می‌شود که به درستی در جریان کار و فرایندهای سازمان به کار گرفته شود. در پلیس نیاز است در ارائه خدمات هوشمند به مردم از ابزارهای هوش مصنوعی بهره برده و فرایندها را هوشمند نمود.

اطلاعات: هوش مصنوعی محدود به خوبی اطلاعاتی است که بر پایه آن ساخته شده است و تشنه دریافت اطلاعات بیشتر است. نظام اطلاعاتی طراحی شود که مهندسی و امنیت را شامل شود. این نظام باید قوانینی برای یافتن منابع، دسترسی‌ها و کنترل کیفیت داشته باشد. در سازمان پلیس که روزمره داده‌های زیادی به آن افزوده می‌شود برای تحلیل لازم است تا اطلاعات به صورت هدفمند شده مورد پایش قرار گیرد.

فناوری و زیرساخت‌ها: تهیه و گسترش فناوری و زیرساخت‌های

¹ Deloitte

استفاده می‌کنند [۱۴].

در برخی مسائل به‌منظور کاهش تلفات و آسیب‌های اجتماعی، پلیس باید سریعاً نسبت به حل مسئله و انجام کنش/واکنش اقدام کند. یکی از راهبردهایی که می‌تواند پلیس را در این مهم یاری کند هوشمندسازی مأموریت پلیس در این راستا می‌باشد. بنا به ضرورت بهتر است قبل از توصیف پلیس هوشمند به تحلیل چالش‌های پیش‌رو پلیس بپردازیم. به‌منظور واکاوی و تحلیل چالش‌های پلیس می‌توان فرآیند کشف جرم را از جنبه‌های مختلف مورد بررسی قرار داد [۱۸-۲۰].

تغییر ماهیت جرم: پدیده جرم وابسته به برخی فاکتورهای اصلی و متغیرهای اساسی مانند فناوری، اقتصاد، تورم، نرخ بیکاری، مسائل فرهنگی و ... در جامعه است. از آنجاکه این متغیرها دائماً در حال تغییر هستند، لذا ماهیت جرم نیز در حال تغییر است. همچنین شیوه و شگرد ارتکاب به جرم نیز وابسته به بسیاری متغیرهای محیطی می‌باشد که این موضوع نیز پویا بوده و نمی‌توان برای آن قالبی لحاظ کرد. در نهایت اینکه پدیده جرم می‌تواند از لحاظ ابعاد و میزان پیچیدگی و وسعت پلیس را با چالش‌های جدی مواجه کند.

تغییر محیط جرم: امروزه مجرمان بیش از گذشته گرایش به جابه‌جایی دارند و در مناطق گوناگون فعالیت می‌کنند. همچنین از فناوری عصر اطلاعات برای هماهنگی و ارتباطات استفاده می‌کنند و برای فرار از دست مجریان قانون، از راهبردهای پیچیده استفاده می‌کنند. فناوری تأثیری شگرف بر سرشت جرم در چند سال گذشته گذاشته است. به‌طور برجسته، دسترسی تقریباً بی‌پایان به تلفن همراه و خدمات پیام کوتاه برخی از روش‌های تحقیق، حفاظت و نظارت مجریان قانون را بی‌اثر کرده است. اکنون جرم را می‌توان، از خانه مجرم، برنامه‌ریزی و حتی اجرا کرد. شبکه‌های اجتماعی مجرمان در حال گسترش است و شناسایی آن‌ها دشوارتر می‌شود زیرا ارتباط بین مجرمان دیگر در خیابان‌ها یا قهوه‌خانه‌ها نیست بلکه به‌طور الکترونیکی و اغلب در مرزهای جغرافیایی گسترده صورت می‌گیرد. در نتیجه، فناوری امروز به مجرمان امکان می‌دهد در جرم‌هایی وارد شوند که با خطری کمتر روبرو شوند، که پیش‌تر در دسترس آن‌ها نبوده است [۲۱].

واقعیت‌های مالی: به سبب وجود محدودیت‌های مالی، باوجود گسترش تنوع و محیط جرائم و در نتیجه توسعه مأموریت فراجا، بودجه و اعتبارات کافی متناسب نبوده و در برخی شرایط کاهش چشمگیری نیز پیدا می‌کند که با توجه به ساختار و فرآیندهای فعلی، موجب کاهش بهره‌وری عملیاتی می‌گردد.

نیروی انسانی: مهم‌ترین عنصری که در فرآیند کشف جرم برای پلیس اهمیت دارد دانش ضمنی و تجارب عملی افسران و کارآگاهان

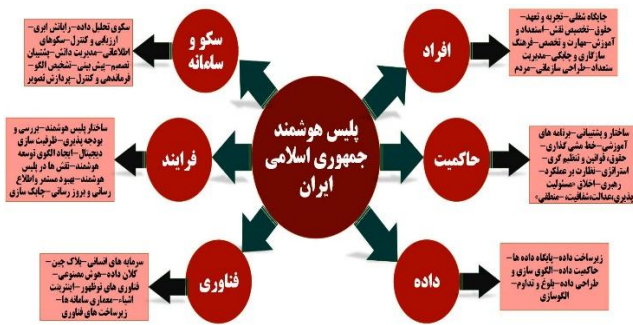
هوش مصنوعی: روندهای هیجان‌انگیز فعلی هوش مصنوعی همه صنایع را در کشورهای مختلف مورد استقبال قرار داده است. هوش مصنوعی این قدرت را دارد که چهره‌ها را بشناسد، ماشین‌های خودران را راه‌اندازی کند، نتایج آنلاین بهتری ارائه دهد و تجارت در صنایع مختلف را تقویت کند. جای تعجب نیست، کشورهای توسعه‌یافته، در حال توسعه و کشورهای توسعه‌یافته همگی بر پذیرش هوش مصنوعی برای فرصت‌ها و توسعه آینده‌نگر بهتر تمرکز کرده‌اند. در حالی که فناوری هوش مصنوعی در سراسر جهان و فراتر از آن در حال شکوفا شدن است، برخی از قدرت‌های بزرگ هوش مصنوعی به سختی برای برنده شدن در مسابقه تلاش می‌کنند [۱۵].

داده کاوی: تجزیه و تحلیل داده‌ها می‌تواند، کلیدی باشد که به توقف جرم کمک می‌کند برای مدتی، واکنش به جرم به حالت پیش‌فرض پلیس در سراسر جهان تبدیل شد. اما امروز ما شاهد بازگشت به اصولی هستیم که سر رابرت پیل در سال ۱۸۲۹ بیان کرد. این امر در وهله اول بر سخت‌تر کردن ارتکاب جرم تا حد امکان تمرکز می‌کند. همچنین، پلیس پیشگیرانه منجر به بهبود کیفیت زندگی می‌شود. کاهش وقوع جرم، تعداد قربانیان مستقیم را کاهش می‌دهد، آرامش ذهنی را برای افراد در جوامع آسیب‌دیده از جرم فراهم می‌کند و به شکستن چرخه جرمی که می‌تواند در خانواده‌ها رخ دهد، کمک می‌کند [۱۶].

یکپارچه‌سازی داده: هیچ رویکرد جهانی برای یکپارچه‌سازی داده‌ها وجود ندارد. با این حال، راه‌حل‌های یکپارچه‌سازی داده‌ها معمولاً شامل چند عنصر مشترک، از جمله شبکه‌ای از منابع داده، یک سرور اصلی، و مشتریانی هستند که به داده‌ها از سرور اصلی دسترسی دارند. در یک فرآیند یکپارچه‌سازی داده‌های معمولی، مشتری درخواستی را برای داده به سرور اصلی ارسال می‌کند. سپس سرور اصلی داده‌های مورد نیاز را از منابع داخلی و خارجی دریافت می‌کند. داده‌ها از منابع استخراج می‌شوند، سپس در یک مجموعه داده منسجم ادغام می‌شوند و برای استفاده به مشتری ارائه می‌شود [۱۷].

۵- چالش‌های پیش‌روی پلیس

محیط شناسی در بخش اصلی ماهیت سازمان پلیس، تدابیر و فرامین و ارزش‌های دینی، تدابیر فرماندهی و هیات رئیسه فراجا، چالش‌ها و محدودیت‌ها و ماهیت سازمان پلیس مورد توجه می‌باشد. پلیس امروزه علاوه بر ارائه خدمات انتظامی، رسیدگی به پرونده‌های جنایی و انجام تحقیقات تا حل مسئله را جزء رسالت خود می‌داند، اما نکته‌ای که باید به آن توجه شود عنصر زمان است. بدان معنا که



شکل ۱. مدل مفهومی تحقیق

جدول ۱. ابعاد و مولفه‌های احصایی توسط محققین

مؤلفه	بعد	مؤلفه	بعد
سرمایه‌های انسانی	فناوری	جایگاه شغلی	افراد
بلاک چین		تجربه و تعهد	
کلان داده		حقوق	
هوش مصنوعی		تخصیص نقش	
فناوری‌های نوپهور		استعداد و آموزش	
اینترنت اشیا		مهارت و تخصص	
معماری سامانه‌ها		فرهنگ سازگاری و چابکی	
زیرساخت‌های فناوری		مدیریت استعداد	
ساختار پلیس هوشمند	فرایند	طراحی سازمانی	حاکمیت
بررسی و بودجه پذیری		مردم	
ظرفیت‌سازی دیجیتال		ساختار و پشتیبانی	
ایجاد الگوی توسعه هوشمند		برنامه‌های آموزشی	
نقش‌ها در پلیس هوشمند		خط مشی گذاری	
بهبود مستمر و اطلاع‌رسانی و بروزرسانی		حقوق، قوانین و تنظیم‌گری	
چابک‌سازی		استراتژی	
سکوی تحلیل داده		نظارت بر عملکرد	
رایانش ابری	سکو و سامانه	رهبری	داده
ارزایی و کنترل		اخلاق (مسئولیت‌پذیری، عدالت، شفافیت، منطقی)	
سکوهای اطلاعاتی		زیرساخت داده	
مدیریت دانش		پایگاه داده‌ها	
پشتیبان تصمیم		حاکمیت داده	
پیش‌بینی		الگوی سازی و طراحی داده	
تشخیص الگو		بلوغ و تدوین	
فرماندهی و کنترل		الگوسازی	
پردازش تصویر			

می‌باشد. با توجه به نوع مأموریت آنان احتمال آسیب جدی و حتی فوت در حین عملیات نیز وجود دارد. افسران پلیس دارای ویژگی‌ها و توانمندی‌های متفاوتی برای حل مسئله می‌باشند و متناسب با شرایط فرهنگی، روانی و ... با روش‌های مختلف دانش لازم برای انجام کار تحقیقاتی خود را دنبال می‌کنند [۲۲].

حافظه سازمانی: پلیس در راستای انجام مأموریت خود و بهره‌مندی از ظرفیت‌های فناوری اطلاعات، فرآیندهای کاری خود را در تمام کشور مکانیزه کرده و بانکهای اطلاعاتی متمرکزی را ایجاد کرده است، اما آنچه حائز اهمیت می‌باشد این است که ممکن است یک پرونده جنایی دارای گستردگی جغرافیایی زیادی باشد، اما مظنونین در مراحل مختلف اطلاعات غلط و اشتباهی را به پلیس‌های محلی داده‌اند و لذا حل مسئله را برای پلیس مشکل کرده‌اند.

روند فزاینده چالش‌های برآمده از وقوع جرم در فضای سایبری: فضای سایبری به عرصه نوپدید برای بسیاری از اقدامات مجرمانه خلاف قانون و شرع تبدیل شده است. این جرائم در سطوح سخت، نیمه سخت و نرم در لایه‌های زیرساخت، خدمات، محتوا و انسانی-اجتماعی فضای سایبر در حال وقوع و ازدیاد است. به‌عنوان مثال انتشار انواع و اقسام اخبار کاذب در این فضا و بهره‌گیری از فضای مجازی و به‌ویژه شبکه‌های مجازی در سال‌های اخیر و استفاده بی‌رویه اقشار جوان و نوجوان از فضای مجازی و سیر آن‌ها در شبکه‌های اجتماعی که به فرموده رهبر معظم انقلاب به قتلگاه جوانان تبدیل شده است [۲۳].

افزایش سلطه علوم و فناوری‌ها بر حیات اجتماعی: افزایش سلطه علوم و فناوری‌ها بر حیات اجتماعی، زمینه‌ساز بروز مفاهیم جدیدی در بافتار مدیریت انتظامی کشورهای جهان همچون «پلیس جامعه‌محور» شده است. پلیس جامعه‌محور الگوی جدید پلیسی است که در آن امنیت با مشارکت مردم صورت می‌پذیرد. اهداف اصلی آن کاهش جرم و ارتقاء احساس امنیت در جامعه است. مشارکت شهروندان و ارائه راه حال ابتکاری برای حل مشکلات از جمله شاخصه‌های این رویکرد است.

۶- مدل مفهومی تحقیق

در این تحقیق پس از بررسی ادبیات و مبانی نظری، اسناد و مدارک و مصاحبه با خبرگان، ابعاد و مولفه‌های الگوی راهبردی بومی پلیس هوشمند ج.ا.ایران، به شرح جدول ۱ استخراج و شاکله مدل مفهومی تحقیق در ۶ بعد: «افراد»، «حاکمیت»، «داده»، «سکو و سامانه»، «فرایند» و «فناوری» و ۴۹ مؤلفه شکل گرفت که مدل مفهومی تحقیق برابر شکل ۱ رسم شده است.

۷- روش‌شناسی تحقیق

این تحقیق از نظر روش اجرا، توصیفی-تحلیلی و از نوع کاربردی-توسعه‌ای می‌باشد. رویکرد پژوهش به صورت آمیخته (کمی و کیفی) بوده و علاوه بر استفاده از روش‌های رایج کتابخانه‌ای مشتمل بر مطالعه کتب و مقالات مربوط به بحث و منابع معتبر اینترنتی، با تعدادی از خبرگان نیز به منظور جمع‌آوری اطلاعات مصاحبه به عمل آمده است.

در راستای غنی‌سازی و تکمیل ادبیات پژوهش محقق با تعداد ۵ نفر از خبرگان مصاحبه نموده است و سوالات ذیل مطرح گردید: ۱- لطفاً نظر خود را در مورد پلیس هوشمند بیان نمایید ۲- به نظر شما الگوی بومی پلیس هوشمند دارای چه ابعاد و مولفه‌هایی است؟

پس از انجام مصاحبه با خبرگان، متن مصاحبه‌ها به روش تحلیل محتوا و به کمک آمار استنباطی مورد تجزیه و تحلیل قرار گرفت.

جامعه آماری تحقیق در بخش کیفی و کمی شامل خبرگان و کارشناسان مرتبط با هوشمندی در فراجا و سازمان‌های نیروهای مسلح می‌باشند. در بخش کیفی پژوهش بدلیل مشابهت مفاهیم و مقوله‌های مرتبط با هوشمندی برای جمع‌آوری دقیق و عمیق عوامل هوشمندسازی از خبرگان سازمان‌های دارای برنامه هوشمندسازی مانند قوه قضاییه؛ شهرداری تهران بزرگ نیز استفاده شده است. نخبگان دارای ویژگی‌هایی بدین شرح می‌باشند: الف: مدیران راهبردی فراجا با سابقه مدیریت در طرح‌های هوشمندسازی

ب: مدیران راهبردی فاوا سازمان‌ها دارای تخصص در حوزه‌های مدیریت راهبردی؛ طراحی سیستم‌ها ج: مدیران اجرایی حوزه‌های تخصصی د: اعضاء هیئت علمی؛ پژوهشگران و متخصصین که تعداد ۵۰ نفر برآورد شده و بصورت هدفمند و تمام شمار به آنها مراجعه شده است.

تجزیه و تحلیل اطلاعات و داده‌های حاصل از مطالعه منابع، با پیروی از سؤالات مربوطه به روش کیفی، توصیف و تحلیل شده‌اند و اطلاعات حاصل از پرسشنامه با استفاده از روش معادلات ساختاری با رویکرد حداقل مربعات جزئی و با استفاده از نرم‌افزار اسمارت پی ال اس مورد تحلیل قرار گرفته است.

تحلیل مدل در این پژوهش با استفاده از روش مدل‌سازی معادلات ساختاری با رویکرد حداقل مربعات جزئی از طریق بررسی برازش مدل انجام می‌گردد. برازش مدل خود در سه بخش انجام می‌شود: برازش مدل اندازه‌گیری؛ برازش مدل ساختاری و برازش مدل کلی.

الف) برازش مدل اندازه‌گیری: به منظور بررسی برازش مدل

اندازه‌گیری از چهار معیار پایایی «سنجش ضرایب همبستگی، آلفای کرونباخ، پایایی ترکیبی و متوسط واریانس استخراج‌شده» استفاده می‌شود.

ب) برازش مدل ساختاری: به منظور برازش مدل ساختاری از دو معیار: مقادیر R Square و Q2 بهره‌گیری می‌شود.

پ) برازش مدل کلی: جهت برازش مدل کلی از معیار GoF استفاده می‌گردد. توسط این معیار، محقق می‌تواند پس از بررسی برازش بخش اندازه‌گیری و بخش ساختاری مدل کلی پژوهش خود، برازش بخش کلی را نیز کنترل نماید.

در بخش تحلیل و تفسیر، آزمون‌های ذکر شده به تفکیک، انجام شده است.

۸- تحلیل و تفسیر

پس از توزیع پرسشنامه بین جامعه نمونه، داده‌های گردآوری شده وارد نرم افزار اسمارت پی ال اس شد و آزمونهای مرتبط با برازش مدل انجام گردید که هر یک از آزمون‌ها به شرح ذیل آورده شده است.

الف) بررسی ضرایب همبستگی ابعاد تحقیق:

جدول ۲. ضریب همبستگی ابعاد تحقیق

ردیف	ابعاد	ضریب همبستگی	T- TEST	
			T- value	P- value
۱.	افراد	۰.۹۴۸	۶۷.۸۸۵	۰.۰۰۰
۲.	حاکمیت	۰.۹۵۵	۱۰۴.۵۹۱	۰.۰۰۰
۳.	داده	۰.۹۱۸	۳۹.۷۲۵	۰.۰۰۰
۴.	سکو و سامانه	۰.۹۷۴	۱۳۴.۲۵۲	۰.۰۰۰
۵.	فرایند	۰.۹۶۹	۱۲۴.۶۹۲	۰.۰۰۰
۶.	فناوری	۰.۹۳۳	۴۴.۵۰۹	۰.۰۰۰

تفسیر جدول ۲:

جدول بالا بیانگر این است که مطابق نظر جامعه نمونه و با توجه به اینکه مقادیر T-value برای هر شش بعد تحقیق بیشتر از ۱,۹۶ و مقادیر P-value نیز برای هر شش بعد تحقیق کمتر از ۰,۰۵ محاسبه شده است، لذا «الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران»، دارای شش بعد: «افراد»، «حاکمیت»، «داده»، «سکو و سامانه»، «فرایند» و «فناوری» است و بین این ابعاد و «پلیس هوشمند جمهوری اسلامی ایران» یک ارتباط مثبت، مستقیم و معناداری وجود دارد.

ب) برازش مدل اندازه‌گیری: جهت برازش مدل اندازه‌گیری از آزمونهای «سنجش ضرایب همبستگی، آلفای کرونباخ، پایایی

			تشخیص الگو	۰.۷۶۲
			فرماندهی و کنترل	۰.۸۱۲
			پردازش تصویر	۰.۸۴۸
			ساختار پلیس هوشمند	۰.۸۹۱
			بررسی و بودجه پذیری	۰.۸۶۲
			ظرفیت‌سازی دیجیتال	۰.۸۲۹
			ایجاد الگوی توسعه هوشمند	۰.۸۳۲
			نقش‌ها در پلیس هوشمند	۰.۸۲۸
			بهبود مستمر و اطلاع رسانی و بروز رسانی	۰.۸۷۹
			چابک‌سازی	۰.۸۶۳
			سرمایه‌های انسانی	۰.۷۱۷
			بلاک چین	۰.۸۴
			کلان داده	۰.۷۰۷
			هوش مصنوعی	۰.۸۲۲
			فناوری‌های نو ظهور	۰.۷۲۶
			اینترنت اشیا	۰.۷۰۹
			معماری سامانه‌ها	۰.۸۶۲
			زیرساخت‌های فناوری	۰.۷۷۳

تفسیر جدول ۳:

پایایی درونی نشانگر میزان همبستگی بین یک بعد و مولفه‌های مربوط به آن است. مقدار بالای واریانس تبیین شده بین بعد و مولفه‌هایش در مقابل خطای اندازه‌گیری مربوط به هر مولفه، پایایی درونی بالا را نتیجه می‌دهد. مقدار آلفای کرونباخ بالاتر از ۰/۷ نشانگر پایایی قابل قبول است. در صورتی که مقدار پایایی ترکیبی (CR) برای هر سازه بالای ۰/۷ شود، نشان از پایداری درونی مناسب برای مدل اندازه‌گیری دارد. معیار متوسط واریانس استخراج شده (AVE)، میزان همبستگی یک بعد با مولفه‌های خود را نشان می‌دهد که هر چه این همبستگی بیشتر باشد، برازش نیز بیشتر است. فورنل و لارکر اظهار داشته اند مقدار AVE بالای ۰/۵، همگرایی قابل قبول را نشان می‌دهد.

الف) آزمون آلفای کرونباخ

مقدار آلفای کرونباخ باید بزرگ‌تر از ۰,۷ باشد که برابر جدول بالا، کلیه ضرایب آلفای کرونباخ برای متغیرهای تحقیق بیشتر از ۰,۷ می‌باشد، بنابراین پایایی مدل بر اساس ضریب آلفای کرونباخ تأیید می‌شود.

ب) آزمون پایایی ترکیبی

ترکیبی و متوسط واریانس استخراج شده» به شرح جدول ۳ بهره‌گیری شده است.

جدول ۳. ضرایب آزمون‌های پایایی مدل متغیرهای پژوهش

بعد	مؤلفه	ضریب همبستگی	آلفای کرونباخ	پایایی ترکیبی	متوسط واریانس استخراج شده
افراد	جایگاه شغلی	۰.۶۹۹	۰.۹۳۳	۰.۹۴۵	۰.۶۵۶
	تجربه و تعهد	۰.۷۳۳			
	حقوق	۰.۷۶۷			
	تخصیص نقش	۰.۷۲۱			
	استعداد و آموزش	۰.۹۰۵			
	مهارت و تخصص	۰.۸۶۴			
	فرهنگ سازگاری و چابکی	۰.۸۳۶			
	مدیریت استعداد	۰.۸۴۶			
	طراحی سازمانی	۰.۸۹۲			
	مردم	۰.۸۹			
حاکمیت	ساختار و پشتیبانی	۰.۸۲۶	۰.۹۳۴	۰.۹۴۵	۰.۶۸۵
	برنامه‌های آموزشی	۰.۷۹۲			
	خط مشی گذاری	۰.۸۸۸			
	حقوق، قوانین و تنظیم گری	۰.۸۴۶			
	استراتژی	۰.۷۴۶			
	نظارت بر عملکرد	۰.۸۶۸			
	رهبری	۰.۸۴۲			
داده	اخلاق (مسئولیت پذیری، عدالت، شفافیت، منطقی)	۰.۸۰۴	۰.۹۲۱	۰.۹۳۹	۰.۷۲۳
	زیرساخت داده	۰.۸۹۶			
	پایگاه داده‌ها	۰.۹۴۶			
	حاکمیت داده	۰.۶۸۹			
	الگوی سازی و طراحی داده	۰.۸۵۵			
	بلوغ و تداوم	۰.۸۱			
سکو و سامانه	الگوسازی	۰.۸۸۲	۰.۹۵	۰.۹۵۸	۰.۷۱۸
	سکوی تحلیل داده	۰.۸۵۲			
	رایانش ابری	۰.۶۸۲			
	ارزیابی و کنترل	۰.۹۱۱			
	سکوهای اطلاعاتی	۰.۸۶۵			
	مدیریت دانش	۰.۸۷۴			
پشتیبان تصمیم	پشتیبان تصمیم	۰.۸۷۴	۰.۹۳		
	پیش بینی	۰.۹۳			

نموده‌اند که به ترتیب مقادیر ضعیف، متوسط و قوی قدرت پیش‌بینی مدل در مورد سازه‌های درون‌زا را نشان می‌دهد.

جدول ۴. آزمون کیفیت مدل اندازه‌گیری

متغیر	Q ²	نتیجه
افراد	۰.۵۴۷	بسیار قوی
حاکمیت	۰.۵۳۲	بسیار قوی
داده	۰.۶۰۸	بسیار قوی
سکو و سامانه	۰.۵۸۴	بسیار قوی
فرایند	۰.۶۶۴	بسیار قوی
فناوری	۰.۳۹۹	قوی

تفسیر جدول ۳:

با توجه به مقادیر Q² در جدول بالا، برای سازه‌های «افراد»، «حاکمیت»، «داده»، «سکو و سامانه»، «فرایند» نشان از برازش بسیار قوی مدل ساختاری دارد و سازه «فناوری»، قابلیت پیش‌بینی قوی با سازه مربوط به خودش برخوردار می‌باشد.

(۲) معیار R²:

دومین معیار برای بررسی برازش مدل ساختاری در این پژوهش، ضرایب R² مربوط به متغیرهای پنهان درون‌زای (وابسته) مدل است. R² معیاری است که نشان از تأثیر یک متغیر برون‌زا دارد. هنسلر و همکاران اظهار داشته‌اند در صورتی که در یک مدل، یک سازه درون‌زا توسط تنها یک یا دو سازه برون‌زا تحت تأثیر قرار گیرد، مقدار R² از ۰/۳۳ به بالا، نشان از قوت رابطه بین آن سازه‌ها و سازه درون‌زا است.

جدول ۵. آزمون R² مؤلفه‌های تحقیق

متغیر	R ²	نتیجه
افراد	۰.۸۹۸	بسیار قوی
حاکمیت	۰.۹۱۲	بسیار قوی
داده	۰.۸۴۳	بسیار قوی
سکو و سامانه	۰.۹۴۸	بسیار قوی
فرایند	۰.۹۳۹	بسیار قوی
فناوری	۰.۸۷۰	بسیار قوی

تفسیر جدول ۵:

مطابق جدول بالا مقدار R² برای کلیه سازه‌ها، بسیار قوی می‌باشد.

(پ) آزمون مدل کلی

برازش مدل کلی (معیار GoF)

توسط این معیار، محقق می‌تواند پس از بررسی برازش بخش اندازه‌گیری و بخش ساختاری مدل کلی پژوهش خود، برازش بخش

مطابق با نظر هنسلر پایایی ترکیبی به معنای همبستگی سؤالات یک متغیر (داخل مدل) می‌باشد. در صورتی که مقدار پایایی ترکیبی برای هر سازه بالای ۰,۷ شود، نشان از پایداری درونی مناسب برای مدل اندازه‌گیری دارد. خوشبختانه کلیه ضرایب پایایی ترکیبی برای متغیرهای پژوهش بیشتر از ۰,۷ است، لذا پایایی مدل بر اساس این آزمون نیز تأیید می‌شود.

(پ) آزمون همبستگی اسپیرمن بین سؤالات هر متغیر (آزمون rho-A)

یک آزمون تکمیلی در سال ۲۰۱۶ بر اساس مطالعات رینگر به ضرایب پایایی اضافه شد که از سؤالات همبستگی اسپیرمن گرفته می‌شود تا اگر طیف لیکرت دارای تعداد گزینه‌های اندک نیز می‌باشد (مثلاً ۵ گزینه‌ای) بتوان برای آن ضریب همبستگی تعریف کرد. این شاخص باید بزرگ‌تر از ۰,۷ باشد. خوشبختانه تمامی ضرایب اسپیرمن بالای ۰,۷ هستند، بنابراین پایایی بر اساس آزمون rho-A نیز تأیید می‌شود.

(ت) آزمون متوسط واریانس استخراج‌شده:

معیار متوسط واریانس استخراج‌شده، میزان همبستگی یک بعد با مولفه‌های خود را نشان می‌دهد که هر چه این همبستگی بیشتر باشد، برازش نیز بیشتر است. فورنل و لارکر اظهار داشته‌اند مقدار متوسط واریانس استخراج شده بالای ۰,۵، همگرایی قابل قبول را نشان می‌دهد. خوشبختانه برای همه متغیرها، متوسط واریانس استخراج شده تمامی سؤالات بالای ۰,۵ است؛ بنابراین پایایی بر اساس این آزمون نیز تأیید می‌شود.

(ب) برازش مدل ساختاری:

به‌منظور برازش مدل ساختاری از دو معیار: مقادیر R Square و Q² به شرح ذیل بهره‌گیری شده است.

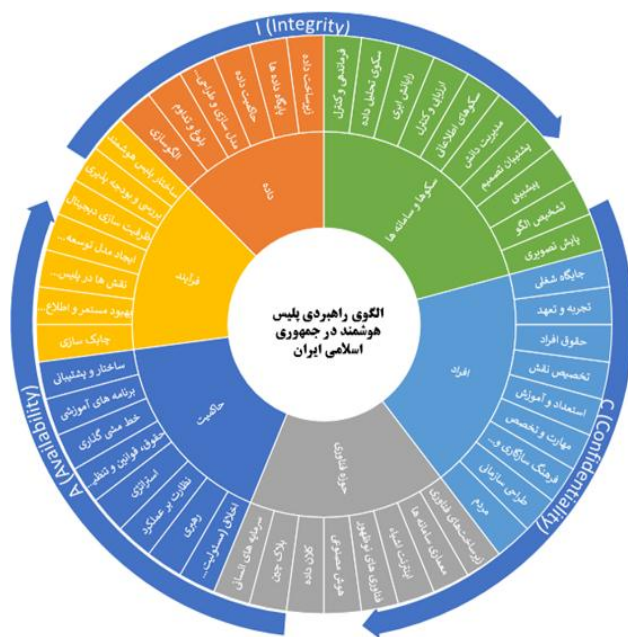
(۱) معیار Q²

معیار Q² قدرت پیش‌بینی مدل را مشخص می‌کند. این معیار که توسط استون و گیسر (۱۹۷۵) معرفی شد، قدرت پیش‌بینی مدل در سازه‌های درون‌زا را مشخص می‌کند. به اعتقاد آن‌ها مدل‌هایی که دارای برازش ساختاری قابل قبول هستند، باید قابلیت پیش‌بینی متغیرهای درون‌زای مدل را داشته باشند. بدین معنی که اگر در یک مدل، روابط بین سازه‌ها به‌درستی تعریف شده باشند، سازه‌ها تأثیر کافی بر یکدیگر گذاشته و از این راه روابط بین سازه‌ها به‌درستی تأیید شوند. هنسلر و همکاران در مورد شدت قدرت پیش‌بینی مدل در مورد سازه‌های درون‌زا، سه مقدار ۰/۰۲، ۰/۱۵ و ۰/۳۵ را تعیین

ایجاد الگوی توسعه هوشمند؛ ظرفیت‌سازی دیجیتال؛ نقش‌ها در پلیس هوشمند.

ج) بعد «فناوری»: با ۸ مؤلفه به ترتیب اولویت (با توجه به مقادیر ضرایب همبستگی) شامل: معماری سامانه‌ها؛ بلاک چین؛ هوش مصنوعی؛ زیرساخت‌های فناوری؛ فناوری‌های نوظهور؛ سرمایه‌های انسانی؛ اینترنت اشیا؛ کلان داده.

برابر آنچه که گفته شد «الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران» برابر شکل ۲ ترسیم شده است.



شکل ۲. الگوی پژوهش

تفسیر الگو: نتایج حاصله از این تحقیق نشان می‌دهد هر شش بعد الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران، شامل: «افراد»، «حاکمیت»، «داده»، «سکو و سامانه»، «فرایند» و «فناوری» به صورت مکمل بر الگوی تحقیق تأثیرگذار می‌باشند، ولی میزان تأثیرگذاری آن‌ها از نگاه جامعه نمونه متفاوت است. جامعه نمونه این اعتقاد را داشتند که بعد «سکو و سامانه» با ضریب مسیر ۰٫۹۷۴، تأثیرگذارتر از پنج بعد دیگر الگو است. تجزیه و تحلیل داده‌ها نشان می‌دهد که بین بعد «سکو و سامانه» با «الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران»، یک همبستگی بسیار زیاد، مستقیم و مثبتی وجود دارد، در واقع می‌توان گفت که این بعد نشان می‌دهد به میزان بهره‌گیری از ۹ مؤلفه بعد «سکو و سامانه» همچون: «پیش‌بینی؛ ارزیابی و کنترل؛ مدیریت دانش؛ پشتیبان تصمیم؛ سکوها؛ اطلاعاتی؛ سکوی تحلیل داده؛ پردازش تصویر؛ تشخیص الگو و رایانش ابری»، به همان میزان به تقویت، اثربخشی و کارآمدی پلیس هوشمند جمهوری اسلامی ایران، منجر خواهد شد و برعکس،

کلی را نیز کنترل نماید. وتزلس و همکاران سه مقدار ۰/۲۵، ۰/۰۱ و ۰/۳۶ را به عنوان مقادیر ضعیف، متوسط و قوی برای GoF معرفی نموده‌اند معیار GoF طبق رابطه زیر محاسبه می‌گردد:

$$GoF = \sqrt{(Communalities^*R^2)}$$

از آنجا که در حداقل مربعات جزئی مقدار Commonality با AVE برابر است وتزلس و همکاران فرمول زیر را ارائه نموده‌اند:

$$GOF = \sqrt{\text{average (AVE)} \times \text{average (R2)}}$$

در نتیجه مقدار معیار GoF برابر است با:

$$GoF = \sqrt{(.902 * 0/685)} = 786/0$$

حاصل شدن مقدار ۰/۷۸۶ برای GoF نشان از «برازش کلی بسیار قوی» مدل دارد.

۹- تحلیل نتایج

نتایج تحقیق نشان داد که از دید صاحب‌نظران و خبرگان این تحقیق «الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران» شامل ۶ بعد و ۴۹ مؤلفه و به شرح ذیل می‌باشد:

الف) بعد «افراد»: با ۱۰ مؤلفه به ترتیب اولویت (با توجه به مقادیر ضرایب همبستگی) شامل: استعداد و آموزش؛ طراحی سازمانی؛ مردم؛ مهارت و تخصص؛ مدیریت استعداد؛ فرهنگ سازگاری و چابکی؛ حقوق؛ تجربه و تعهد؛ تخصیص نقش؛ جایگاه شغلی.

ب) بعد «حاکمیت»: با ۸ مؤلفه به ترتیب اولویت (با توجه به مقادیر ضرایب همبستگی) شامل: خط مشی گذاری؛ نظارت بر عملکرد؛ حقوق، قوانین و تنظیم گری؛ رهبری؛ ساختار و پشتیبانی؛ اخلاق (مسئولیت پذیری، عدالت، شفافیت، منطقی)؛ برنامه‌های آموزشی؛ استراتژی.

پ) بعد «داده»: با ۶ مؤلفه به ترتیب اولویت (با توجه به مقادیر ضرایب همبستگی) شامل: پایگاه داده‌ها؛ زیرساخت داده؛ الگوسازی؛ الگوی سازی و طراحی داده؛ بلوغ و تداوم؛ حاکمیت داده.

ت) بعد «سکو و سامانه»: با ۱۰ مؤلفه به ترتیب اولویت (با توجه به مقادیر ضرایب همبستگی) شامل: پیش بینی؛ ارزیابی و کنترل؛ مدیریت دانش؛ پشتیبان تصمیم؛ سکوها؛ اطلاعاتی؛ سکوی تحلیل داده؛ پردازش تصویر؛ فرماندهی و کنترل؛ تشخیص الگو؛ رایانش ابری.

ث) بعد «فرایند»: با ۷ مؤلفه به ترتیب اولویت (با توجه به مقادیر ضرایب همبستگی) شامل: ساختار پلیس هوشمند؛ بهبود مستمر و اطلاع رسانی و بروز رسانی؛ چابک‌سازی؛ بررسی و بودجه پذیری؛

الزامات و راهکارهای اجرای الگوی راهبردی پلیس هوشمند در جمهوری اسلامی شامل موارد زیر است:

۱) **الزامات ساختار و سازمان:** از اهداف اصلی هوشمندسازی سازمانی مدیریت بهینه فرایندها و منابع سازمان می‌باشد. برای انجام این کار، باید ساختار و سازمان نیز متناسب با اهداف اصلاح و بهینه‌سازی گردد. الزامات ساختار و سازمان برای هوشمندسازی شامل مدیریت کلان، سازمان دهی، پایش، نظارت، اندازه‌گیری و تنظیم گری، محیط آزمایشی و... می‌باشد.

۲) **الزامات حقوقی:** برای پیاده‌سازی الگو جهت تحقق اهداف پلیس هوشمند و تعاملات مناسب و سازنده با سایر سازمان‌های موثر در انجام ماموریت‌های پلیسی و همچنین افراد جامعه، نیازمند شاخص‌های حقوقی از قبیل رعایت حقوق شهروندی، حریم خصوصی، شفافیت، مسئولیت‌پذیری سازمان‌ها، پاسخگویی پلیس و... می‌باشد پیاده‌سازی این سیاست‌های حقوقی نیازمند تهیه و تصویب قوانین؛ آیین‌نامه‌ها؛ دستورالعمل‌ها و شیوه‌نامه‌ها و رعایت دقیق آن‌ها در طرح‌های مرتبط با پیاده‌سازی الگو می‌باشد.

۳) **الزامات منابع پشتیبانی:** برای پیاده‌سازی الگو بمنظور دستیابی به پلیس هوشمند نیاز به تامین الزامات پشتیبانی زیرساختی و سامانه‌ای می‌باشد مهم‌ترین الزامات شامل اسناد راهبردی، منابع اعتباری پایدار طبق برنامه‌های سالانه؛ سامانه‌های نرم‌افزاری و زیرساختی، سکوها، تعاملی با دستگاه‌های حاکمیتی، شهروندان و رده‌های پلیس، فضا آموزشی و بسترهای آموزش مجازی، ابزارهای شبیه‌سازی مانند انواع شبیه‌سازها و مرکز تحلیل داده می‌باشد.

۴) **الزامات آموزشی:** با توجه به جامعیت الگو و ضرورت آموزش کلیه نقش‌ها در زیست بوم پلیس هوشمند یکی از مهم‌ترین الزامات پیاده‌سازی الگو حوزه آموزش در این زیست بوم می‌باشد. الزامات آموزشی شامل فراگیران (نیروهای پلیس، مردم و دستگاه‌ها)، اساتید خبرگان آموزشی متناسب با جامعه فراگیران و اهداف آموزشی، محتوای آموزشی با اولویت فناوری‌های روزآمد و مبتنی بر آموزش‌های فراگیر مجازی، شیوه‌های تحلیل داده و...؛ کانال آموزشی در بسترهای مجازی و فضای حقیقی با رویکرد تغییر پارادایم آموزشی از یاددهی به یادگیری می‌باشد.

۵) **الزامات سرمایه انسانی:** سرمایه انسانی یکی از مهم‌ترین الزامات الگو بوده و در پیاده‌سازی و عملیاتی نمودن سایر ابعاد و مولفه‌های الگو نقش اساسی دارد برای پیاده‌سازی الگو لازم است الزامات مرتبط با عامل انسانی بصورت اساسی مورد توجه قرار گیرد.

به میزان کاهش بهره‌گیری از آن‌ها، این تقویت، اثربخشی و کارآمدی کاهش خواهد یافت.

جامعه نمونه دومین بعد تأثیرگذار بر «پلیس هوشمند جمهوری اسلامی ایران» را بعد «فرایند» با ضریب مسیر ۰,۹۶۹ می‌داند. نکته مورد توجه اینکه، میزان بهره‌گیری از مؤلفه‌های این بعد است؛ که به میزان افزایش بهره‌گیری از مؤلفه‌های این بعد، بطور نسبی میزان تأثیرگذاری بر «پلیس هوشمند جمهوری اسلامی ایران» افزایش خواهد یافت. سومین بعد تأثیرگذار از دید جامعه نمونه بر «پلیس هوشمند جمهوری اسلامی ایران»، بعد «حاکمیت» با ضریب مسیر ۰,۹۵۵ می‌باشد. چهارمین بعد مؤثر بر «پلیس هوشمند جمهوری اسلامی ایران»، از نگاه جامعه نمونه، بعد «افراد» با ضریب مسیر ۰,۹۴۸ است. پنجمین بعد مؤثر بر «پلیس هوشمند جمهوری اسلامی ایران»، از نگاه جامعه نمونه، بعد «فناوری» با ضریب مسیر ۰,۹۳۳ است. ششمین بعد مؤثر بر «پلیس هوشمند جمهوری اسلامی ایران»، از نگاه جامعه نمونه، بعد «داده» با ضریب مسیر ۰,۹۱۸ است.

۱۰- پیشنهاد

با توجه به نتایج حاصل از کار پژوهشی انجام شده، پیشنهادهای ذیل ارائه می‌گردد:

الف) پیشنهاد نظری: طرح راهبردی «پلیس هوشمند جمهوری اسلامی ایران» و «الگوی ارزیابی، نظارت، کنترل و پایش پلیس هوشمند جمهوری اسلامی ایران»، توسط پژوهشگران تدوین گردند.

ب) پیشنهاد سیاستی: فرماندهی انتظامی جمهوری اسلامی ایران نسبت به در اولویت قرار دادن تامین الزامات مورد نیاز در حوزه‌های «زیرساختی، منابع انسانی، منابع مالی، سخت‌افزاری، نرم‌افزاری و ...» جهت توسعه پلیس هوشمند اقدام نماید.

پ) پیشنهاد کاربردی: محتوی و نتایج پژوهش، مورد بهره‌برداری وزارت کشور و فرماندهی انتظامی جمهوری اسلامی ایران قرار گیرد. معاونت آموزش فراجا در خصوص برگزاری دوره‌های آموزشی مورد نیاز برای کارکنان و فرماندهان، فرماندهی انتظامی جمهوری اسلامی ایران به منظور کسب دانش لازم و تقویت مهارت افزایی برگزار نماید. معاونت طرح و برنامه و بودجه فراجا با همکاری سایر معاونت‌های فراجا، نسبت به طرح‌ریزی و برنامه‌ریزی جهت پیاده‌سازی «الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران»، اقدام نمایند. معاونت‌های مهندسی و فناوری ارتباطات فراجا نسبت به فراهم آوردن زیرساخت‌های لازم جهت پیاده‌سازی «الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران»، اقدام نمایند.

هوشمند تبدیل شود، لذا هدف این تحقیق «ارائه الگوی راهبردی بومی پلیس هوشمند جمهوری اسلامی ایران» می‌باشد.

مراجع

- [۱] اردلان، محمدرضا؛ قنبری، سیروس؛ نصیری ولیک بنی، فخرالسادات؛ بهشتی‌راد، رقیه " نقش رهبری خدمتگزار در ارتقای اعتماد سازمانی با نقش میانجی توانمندسازی. " فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی، دوره سوم، شماره چهارم، ۱۳۹۲
- [۲] اوجاقی، علی. " بررسی نقش میانجی گروه‌های تسهیم تجربه بر رابطه ساختار سازمانی ارگانیک و مدیریت دانش در سازمانهای نظامی. " مدیریت نظامی، دوره سوم، شماره نوزدهم، ۱۳۹۸
- [۳] الهی، شعبان و آذر، عادل. " سیستم‌های هوشمند اطلاعاتی مدیریت، رویکرد فازی، عصبی " فصلنامه علمی پژوهشی مدرس، ۱۵۷-۱۳۵، ۱۳۷۸
- [۴] حقانی، محمود؛ شبیبانی، حسن؛ کرباسیان، سعید. " سنجش سرمایه‌های فکری به منظور توسعه منابع انسانی در شرکتهای خودروسازی. " رهیافتی نو در مدیریت آموزشی، دوره نهم، شماره چهل و سوم، ۱۳۹۷
- [۵] حمیدزاده، مهرداد، پورا بر اهی می، علیرضا، طلوعی اشلقی، عباس، معتدل، محمدرضا. " امنیت سازمان هوشمند مبتنی بر هستان شناسی با رویکرد مفهوم‌سازی شبکه " نشریه فناوری اطلاعات و ارتباطات انتظامی دوره چهارم شماره سیزدهم، ۱۴۰۲
- [۶] شاه محمدی، غلامرضا، درویشی سرنابادی، رضا، قمری، محمدرضا. الگوی پلیس هوشمند در فرماندهی انتظامی جمهوری اسلامی ایران؛ پژوهش‌های مدیریت انتظامی، دوره هجدهم، شماره اول ۱۴۰۲
- [۷] شاه محمدی، غلامرضا، اکباتانی، سمیه " آسیب‌شناسی هوشمندسازی مرزهای جمهوری اسلامی ایران " اولین همایش ملی رویکردهای نوین در مدیریت مرز " دانشگاه علوم انتظامی امین. ۱۴۰۱
- [۸] قنوتی زاده درویش، ویدا و گودرزی، علی. " هوشمندسازی و تاثیر تکنولوژی دیجیتال در حوزه‌های مدیریت منابع انسانی " دومین کنفرانس ملی آینده نگری در روانشناسی و علوم تربیتی، شیراز، ۱۴۰۱
- [۹] درویشی، صیاد. " تأثیر کیفیت خدمات پلیس هوشمند بر مشارکت شهروندان در پیشگیری از توزیع مواد مخدر " مجله علمی " مدیریت سرمایه اجتماعی، دوره هفتم، شماره چهارم، ۱۳۹۹
- [10] Katz, Charles M. & Kurtenbach, Mike & Choate, David E. & White, Michael D. 2019. Phoenix, Arizona, Smart Policing Initiative: Evaluating the Impact of Police Officer Body- Worn Cameras, Bureau of Justice Assistance, NCJRS Abstract.
- [11] Coldren Jr., James R. & Huntoon, Alissa & Medaris, Michael. "Introducing Smart Policing: Foundations, Principles, and Practice," Police quarterly, Vol(16), No(3), PP. 275-286, 2013.
- [12] Arcos, R. (2016). Public relations strategic intelligence: Intelligence analysis, communication and influence. Public Relations Review, Vol(42), No(2), PP.264-270, 2016
- [۱۳] درستکار یاقوتی، بهنام. " بازیابی ویدئو مبتنی بر محتوا با استفاده از شبکه عصبی عمیق برای کشف علمی جرائم در پلیس هوشمند " نشریه "فناوری اطلاعات و ارتباطات انتظامی، دوره چهارم، شماره سیزدهم، ۱۴۰۲.

محورهای اصلی الزامات سرمایه انسانی شامل جذب نیروی انسانی کارآمد، توانمندسازی و ارتقاء مهارت‌های حرفه‌ای کارکنان، نگهداشت کارکنان و ایجاد انگیزش جهت افزایش تعهد، مسئولیت‌پذیری، پاسخگویی و بهره‌وری حداکثری و..... می‌باشد. در سطوح مدیریتی نیز استعدادیابی، شایسته‌گزینی، هم‌تا پروری و جایگزینی بطور فرآیندی برای مشاغل مرتبط پیاده‌سازی گردد.

ب) راهکارهای اجرایی شدن الگوی راهبردی بومی پلیس هوشمند ج.ا.ایران شامل موارد ذیل می‌باشد:

جذب نیروهای خبره با دانش دیجیتال - آموزش و فرهنگ‌سازی نیروها - ارتقای جایگاه سازمانی و ساختار تشکیلاتی متناسب با الگو انجام شود - استقرار زیرساخت بلاک چین بصورت بومی - ایجاد بستر امن جهت تبادل پرونده‌ها بصورت الکترونیکی بین سازمانی - طراحی و ایجاد زیرساخت سخت افزاری و انبار داده - طراحی و پیاده‌سازی زیرساخت نرم‌افزاری پردازش و تحلیل - ایجاد و توسعه معماری پایگاه داده‌ها - کاربردهای هوش مصنوعی برای ارائه خدمات هوشمند توسط پلیس با نگاه حل مسائل واقعی و چالش‌های راه‌اندازی احصاء و پس از اولویت‌بندی متناسب با آناه پروژه تعریف شود - با استفاده از مدل‌های اندازه‌گیری میزان آمادگی هوش مصنوعی، سطح بلوغ پلیس برای بکارگیری هوش مصنوعی احصاء شود - در راستای روزآمدسازی پلیس، پروژه‌های پژوهشی مستمری برای شناسایی و رصد آخرین پیشرفتهای فناوری‌های نوظهور در دنیا، علی‌الخصوص در حوزه‌های مرتبط با ماموریت‌های پلیس، تعریف و از خروجی آن‌ها برای استخراج برنامه اقدام بکارگیری این فناوری‌ها بهره گرفته شود - استفاده از معماری‌های مدرن و به‌روز برای سامانه‌های هوشمندسازی پلیس - تهیه و استقرار زیرساخت‌ها (منابع، ذخیره‌سازی، پردازش، امنیت).

۱۱- نتیجه‌گیری

در این مقاله تلاش شد تا ابزارها و فرآیندهای مهم جهت رسیدن به پلیس خدمتگزار و هوشمند معرفی گردد. بواسطه اجرایی شدن هوشمندسازی در پلیس، مدیریت یکپارچه و فرماندهی و اشراف بر موقعیت‌ها و صدور دستورهای مناسب در ماموریت‌ها منجر خواهد شد. یکی از مولفه‌های ارزشمند و پایدار جهت افزایش امنیت، همراهی مردم با پلیس در راستای کشف علمی جرائم و همچنین بهره‌مندی پلیس از جدیدترین فناوری‌ها در سطح سازمان می‌باشد. با وجود نقش گسترده فناوری اطلاعات در فراجا برای آنکه فراجا بتواند به نحو مطلوب ماموریت‌های محوله را به انجام رساند و نظم و امنیت را برقرار کند، باید سازمان پلیس از فناوری محوری به پلیس

- [۱۴] امیری، نسرین، فرورقی، کیوان، قادری، محمد رضا. "آنتن شکاف دار موجبری با شکاف بدون خمش در دیواره باریک برای کاربردهای پهنای در باند X" نشریه "فناوری اطلاعات و ارتباطات انتظامی، دوره چهارم، شماره پانزدهم، ۱۴۰۲.
- [۱۵] گرامی، محسن، مبین علی، افسانه، هاشم پور، راهب، یزدان‌بان، وحید. تحلیل رفتار کاربران در فروشگاه‌های اینترنتی با استفاده از یادگیری عمیق "نشریه" فناوری اطلاعات و ارتباطات انتظامی، دوره چهارم، شماره سیزدهم، ۱۴۰۲.
- [۱۶] لک، بهزاد، بختیاری، سعید، محمد زاده مهنه، حمید رضا، امینی، ابوالفضل. "تحلیل تاثیر ابزارهای فناوری اطلاعات در هوشمندسازی پلیس در عرصه مبارزه با قاچاق از گمرکات مرزی" نشریه "فناوری اطلاعات و ارتباطات انتظامی، دوره چهارم، شماره پانزدهم، ۱۴۰۲.
- [۱۷] حیدری، محمدرضا، کریمی زند، مهدی، کرامتی، محمدعلی، آرائی، وحید، موسوی، سید عبدالله امین. شناسایی و اولویت‌بندی عوامل موثر بر حکمرانی نوین مبتنی بر کارکردهای هوش مصنوعی فرماندهی انتظامی ج.ا. "فناوری اطلاعات و ارتباطات انتظامی، دوره چهارم، شماره چهاردهم، ۱۴۰۲.
- [۱۸] قیصری، نورالله" اقتدار، جذابیت، هوشمندی و معماری ناجای آینده." فصلنامه ی مطالعات راهبردی ناجا، سال سوم، شماره هشتم، ۱۳۹۷
- [۱۹] قلاوند، کورش، حاجی ملی میرزایی، حامد. مسئله شناسی نظام انتظامی فضای مجازی جمهوری اسلامی ایران نشریه "فناوری اطلاعات و ارتباطات انتظامی، دوره چهارم، شماره پانزدهم، ۱۴۰۲.
- [۲۰] مجدی، علی اکبر؛ تیموری، محمود؛ میرمحمدتبار، سید احمد. "تحلیل و واکاوی عوامل مؤثر بر یادگیری سازمانی؛ فراتحلیلی از تحقیقات موجود" فصلنامه علوم اجتماعی، دوره پانزدهم، شماره اول، ۱۳۹۷
- [21] Alegre J., Chiva R.; Assessing impact of organizational learning capability on product performance; *Technovation*, No. 28, , pp.315–326. 2014
- [22] Al-Kasasbeh, M. M. (2019). The Impact of E-Management in Achieving Competitive Advantages. *International Journal of Adaeh*, 10, 101–120.
- [23] Afzal, M., & Panagiotopoulos, P. (2020). Smart policing: A critical review of the literature. In *International Conference on Electronic Government* (pp. 59-70).

Designing A New Genetic-Fuzzy Type 2 Approach to Evaluate Self-Adaptive Systems by Software Quality Indicators

Majid Abdolrazzagh-Nezhad^{1*}, Eshrat Zargari², Mahdi Kherad³

¹ Associate Professor, Department of Computer Science, Faculty of Computer & Industrial Engineering, Birjand University of Technology, Birjand, Iran

² Department of Computer Engineering, Birjand Branch, Islamic Azad University, Birjand, Iran

³ Ph.D. Student of Information Technology, Faculty of Engineering, Department of computer Engineering, University of Qom, Qom, Iran

Received: 23 May 2023, Revised: 12 March 2024, Accepted: 22 April 2024

Paper type: Research

Abstract

Self-adaptive systems are one of the types of software systems that change their behavior according to system conditions and environmental conditions and adapt themselves to it. Although the evaluation of the performance of these systems has been covered in most researches, the evaluation of their quality has remained closed. Therefore, designing an approach is an essential issue to evaluate the quality of self-adaptive systems. The first challenge, the quality indicators of these systems are not fixed, specific, and definite parameters. For example, one of the qualitative indicators for a self-adaptive system is the ability to run software on different operating systems. This indicator may have different degrees of importance for different experienced individuals. The next challenge, some qualitative indicators are not mathematical variables, but rather a linguistic variable between users and experts, indicating that these qualitative quantities are fuzzy variables and can be completely formulated by fuzzy logic. In this article, a new genetic-fuzzy type 2 approach has been proposed to evaluate the quality of the systems based on their quality indicators. The fuzzy logic type 2 is used to describe qualitative indicators, and a genetic algorithm is utilized to determine the optimal fuzzy weights of the qualitative indicators. In the proposed method, it has been tried to compare the self-adaptive systems from two dimensions, including the software dimension and their self-adaptive dimension. In order to evaluate the proposed method, a traffic control system called InSync, which is an adaptive traffic control system (ATCS) and contains multiple qualitative factors, has been used. The obtained results confirm the effectiveness of the proposed method due to the lack of scenario generation, the ability to extend the method to all software quality parameters, and its simplicity. Also, the method is more comprehensive than other existing evaluation models. In addition, it has been tried to compare the self-adaptive systems from two dimensions, including the software dimension and their self-adaptive dimension. This is despite the fact that most of the existing research deals with only one dimension.

Keywords: Self-Adaptive System, Quality Indicators of The Software, Genetic Algorithm, fuzzy Type 2, Adaptive traffic control system (ATCS).

* Corresponding Author's email: abdolrazzagh@birjandut.ac.ir

ارائه یک رویکرد ژنتیک-فازی نوع ۲ برای ارزیابی کیفیت سیستم‌های خودانطباق با استفاده از شاخص‌های کیفی نرم‌افزار

مجید عبدالرزاق نژاد^{۱*}، عشرت زرگری^۲، مهدی خرد^۳

^۱ دانشیار، گروه علوم کامپیوتر، دانشکده مهندسی کامپیوتر و صنایع - دانشگاه صنعتی بیرجند، بیرجند، ایران

^۲ کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد بیرجند، بیرجند، ایران

^۳ دانشجوی دکتری، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه قم، قم، ایران

تاریخ دریافت: ۱۴۰۲/۰۳/۰۲ تاریخ بازبینی: ۱۴۰۲/۱۲/۲۲ تاریخ پذیرش: ۱۴۰۳/۰۲/۰۳

نوع مقاله: پژوهشی

چکیده

سیستم‌های خودانطباق یکی از انواع سیستم‌های نرم‌افزاری هستند که رفتار خود را با توجه به شرایط محیطی تغییر می‌دهند. اگرچه ارزیابی عملکرد این سیستم‌ها در بیشتر تحقیقات تحت پوشش قرار گرفته شده است، ولی ارزیابی کیفیت آنها مغفول مانده است. اولین چالش در این مسیر، عدم قطعیت شاخص‌های کیفی این سیستم‌ها است که پارامترهایی ثابت، مشخص و قطعی ندارند. به عنوان مثال یکی از شاخص‌های کیفی سیستم خودانطباق، توانایی اجرای نرم‌افزار در سیستم‌عامل‌های مختلف است. این شاخص، از نظر افراد خبره مختلف، می‌تواند درجه اهمیت مختلفی داشته باشد. چالش دیگر، برخی شاخص‌های کیفی متغیر ریاضی نیستند و یک متغیر زبانی بین کاربران و کارشناسان می‌باشند، که این امر نشان می‌دهد این کمیت‌های کیفی، متغیرهای فازی بوده و با منطق فازی کاملاً قابل فرموله شدن می‌باشند. در این مقاله به منظور ارزیابی کیفیت این سیستم‌ها براساس شاخص‌های کیفی نرم‌افزار اقدام به ارائه یک رویکرد ژنتیک-فازی نوع دوم شده است. جهت توصیف شاخص‌های کیفی از منطق فازی نوع دوم و برای تعیین بهینه وزن‌های فازی شاخص‌های کیفی از الگوریتم ژنتیک استفاده شده است. در روش پیشنهادی سعی شده تا سیستم‌های خودانطباق از دو بعد شامل بعد نرم‌افزاری این سیستم‌ها و بعد خودانطباق آنها مورد مقایسه قرار گیرند. جهت ارزیابی روش پیشنهادی از یک سیستم کنترل ترافیک با نام InSync که یک سیستم خودانطباق کنترل ترافیک (ATCS) بوده و حاوی شاخص‌های متعدد کیفی می‌باشد استفاده شده است. نتایج بدست آمده با توجه به عدم نیاز به تولید سناریو، توانایی بسط روش به تمامی شاخص‌های کیفی نرم‌افزار و سادگی آن تایید کننده کارآمدی روش پیشنهادی می‌باشند. همچنین جامعیت بیشتر روش پیشنهادی را نسبت به سایر مدل‌های ارزیابی موجود نشان می‌دهد.

کلیدواژگان: سیستم خودانطباق، شاخص‌های کیفی نرم‌افزار، الگوریتم ژنتیک، فازی نوع دو، سیستم خودانطباق کنترل ترافیک

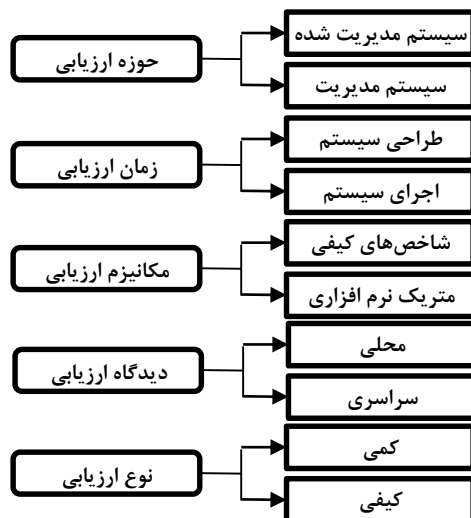
* رایانامه نویسنده مسؤول: abdolrazzagh@birjandut.ac.ir

۱- مقدمه

تصمیمات بهتر در این سیستم‌ها در مواجهه با شرایط عدم قطعیت محیط و الگوریتم ژنتیک به عنوان یک روش بهینه‌سازی می‌باشند. محققین در این مقاله همچنین، چالش‌های تحقیقاتی مرتبط با سیستم‌های خودانطباق را مورد بررسی قرار داده‌اند که عبارتند از:

۱. تعریف دقیق مفاهیم مربوط به این سیستم‌ها و نحوه ارزیابی آنها.
۲. ارائه روش‌های جدید برای پیاده‌سازی این سیستم‌ها با توجه به چالش‌های موجود.
۳. تعریف شاخص‌های جدید برای ارزیابی کارایی، اعتمادپذیری و امنیت این سیستم‌ها.
۴. پیدا کردن راه‌حل‌های مناسب برای مواجهه با چالش‌های مربوط به تصمیم‌گیری در این سیستم‌ها.
۵. تعریف روش‌های جدید برای مدیریت منابع در آنها.
۶. پیدا کردن راه‌حل‌های مناسب برای مواجهه با چالش‌های مربوط به تغییرات در شرایط محیطی.

باتوجه به چالش‌های شماره ۱ و ۳، می‌توان گفت نیاز به رهیافتی جهت سنجش و ارزیابی این سیستم‌ها یک مسئله ضروری می‌باشد. برای ارزیابی سیستم خودانطباق نگرش‌های مختلفی وجود دارد، که می‌توان آنها را از دیدگاه‌های مختلف دسته‌بندی کرد که در شکل ۱ نشان داده شده است [۶]. همانطور که در این شکل مشاهده می‌شود، ارزیابی این سیستم‌ها تنها محدود به ارزیابی عملکردی، زمانی، سازوکاری و محلی/سراسری نیست، بلکه کیفیت آنها را نیز در بر می‌گیرد. روش‌های ارزیابی کیفی این سیستم‌ها در [۱] شامل شاخص‌های مختلف مانند یادگیری، اعتمادپذیری و امنیت، تعادل بین کارایی و هزینه و ... هستند. به عنوان مثال، برای ارزیابی کارایی این سیستم‌ها، شاخص‌هایی مانند زمان پاسخ، دقت و سرعت تصمیم‌گیری مورد بررسی قرار می‌گیرند.



شکل ۱. طبقه‌بندی نگرش‌ها برای ارزیابی سیستم خودانطباق [۶]

بسیاری از تکنیک‌ها و راه‌حل‌های علوم مهندسی، از طبیعت الهام گرفته‌اند. یکی از این الهامات، خاصیت انطباق‌پذیری است که در آن، رفتارها و سازمان‌بندی‌ها در جهت سازگاری و هماهنگی بیشتر، با محیط پیرامون، تغییر می‌کنند. این تغییر ممکن است، هم در ساختار و هم در رفتار اتفاق بیافتد. هدف از تغییرات، سازگاری بیشتر با محیط پیرامون است. تغییر ممکن است به صورت فوری یا در بازه زمانی طولانی رخ دهد و نتیجه آن در طبیعت، تکامل مطلوب است. خودانطباق درجه‌ای از انطباق سیستم نرم‌افزاری در جهت عملکرد بهینه در مواجهه با تغییرات دینامیک محیط و عدم قطعیت است. خودانطباق که کاربرد وسیعی در شبکه‌های اجتماعی، اینترنت اشیا، سیستم‌های خودرو و ... دارد، اجازه می‌دهد تا سیستم نرم‌افزاری بتواند به‌طور خودکار رفتار خود را در زمان اجرا تغییر دهد و خود را سازگار کند [۱]، که این امر سبب کاهش هزینه پیچیدگی ناشی از نگهداری دستی سیستم می‌شود [۲]. به عبارت بهتر، خودانطباقی، قابلیت است که سیستم، رفتار خود را در پاسخ به محیط سازگار می‌کند، بنابراین سیستم می‌تواند بسته به زمان و محیط، تغییرات مناسب را در خود اجرا کند [۳]. این ویژگی، منجر به حضور یک سازوکار انطباق با حداقل تعامل و دخالت انسان در اجرا سیستم می‌شود [۴]. برای مثال، یک ترموستات خانه، دمای خانه را اندازه‌گیری می‌کند و آن را با مجموعه نقاط، کنترل سیستم گرما و سرمای خانه کنترل می‌کند. این یک مثال ساده از مدل ترمودینامیکی است. در سیستم بسیار پیچیده‌تر، نیاز به یک مدل صریح‌تر است. برای مثال در سیستم تهویه هوا برای ساختمان‌های بزرگ، که چندین موقعیت را نظارت و کنترل می‌کند، نیاز به یک مدل صریح از موقعیت ساختمان و دماهای آن، برای کنترل کارآمد است که مشخص کند کدام واحد سرمایه‌گذاری و در چه زمانی، روشن شود. مثال دیگر سیستم کنترل ترافیک است که یکی از معمول‌ترین انواع سیستم‌های خودانطباق است که تحقیقات بسیاری در حوزه آن انجام گرفته است. لذا باتوجه به آنچه اشاره شد می‌توان سازوکار این سیستم‌ها را به چهار فرآیند تجزیه کرد: نظارت بر نهادهای نرم‌افزاری (خودآگاهی) و محیط (آگاه به زمینه)، تجزیه و تحلیل تغییرات مهم، برنامه‌ریزی برای نحوه واکنش، و اجرا، به طوری که تصمیمات موثر بر اساس سیگنال‌های دریافتی از حسگرها به منظور خودسازگار سازی اخذ شوند [۵].

روش‌های مختلف پیاده‌سازی این سیستم‌ها در [۱] بررسی شده‌اند. این روش‌ها شامل استفاده از شبکه‌های عصبی به عنوان یک روش پردازش داده‌های غیرخطی، منطق فازی به عنوان یک روش پردازش داده‌های عددی با مقادیر بین صفر و یک به منظور کمک به اخذ

می‌گردد و اولویت و اعتبارسنجی این شاخص‌ها بررسی شود. در فاز دوم یک سیستم ژنتیک-فازی شامل تعیین بهینه وزن‌های فازی توابع عضویت، فازی‌ساز، پایگاه قواعد، موتور استنتاج و غیرفازی‌ساز طراحی می‌شود و در نهایت یک روش برای ارزیابی سیستم‌های خودانطباق ارائه می‌شود.

در ادامه این مقاله در بخش ۲، پیشینه تحقیق آورده شده است و در بخش ۳، مسئله پژوهش و مفاهیم مرتبط با آن شامل سیستم‌های خودانطباق، ارزیابی کیفی و فازی نوع ۲ تشریح شده است. روش پیشنهادی در بخش ۴ و پیاده‌سازی و ارزیابی روش پیشنهادی این مقاله در بخش ۵ ارائه شده است. نهایتاً، خلاصه و نتیجه‌گیری رویکرد پیشنهادی در بخش ۶ جمع‌بندی شده است.

۲- پیشینه تحقیق

برای ارزیابی سیستم‌های نرم‌افزاری مدل‌های کیفی مختلفی ارائه شده است. از جمله مهم‌ترین مدل‌های کیفی می‌توان به McCall [۸]، ISO9126 [۹] و مدل بوهم [۱۰] اشاره کرد. جدول ۱ به طور خلاصه، شاخص‌های کیفی مربوط به این سه مدل کیفی را نشان می‌دهد [۱۱]. خلاصه تحقیقات انجام گرفته شده در حوزه ارزیابی کیفیت نرم‌افزارها از سال ۲۰۰۸ تا کنون در جدول ۲ جهت تسهیل و مقایسه روند آنچه تاکنون ارائه شده، تشریح شده است.

جدول ۱. شاخص‌های کیفی مدل‌های McCall، بوهم و ISO9126

معیارها / اهداف	McCall	بوهم	ISO 9126
قابلیت نگهداری	✓	✓	✓
قابلیت انعطاف‌پذیری	✓	✓	
قابلیت آزمایش	✓	✓	قابلیت نگهداری
قابلیت حمل	✓	✓	✓
قابلیت استفاده مجدد	✓	✓	
قابلیت همکاری درونی	✓		
درستی (دقت)	✓	✓	قابلیت نگهداری
قابلیت اعتماد	✓	✓	✓
بهره‌وری	✓	✓	✓
یکپارچگی	✓	✓	✓
قابلیت استفاده	✓	✓	
قابلیت درک	✓	✓	
قابلیت اصلاح	✓	✓	قابلیت نگهداری

استفاده از منطق فازی برای ارزیابی سیستم‌های نرم‌افزاری، روشی جدید است ولی در برخی از تحقیقات از این روش برای ارزیابی سیستم‌های نرم‌افزار (و نه سیستم‌های خودانطباق) استفاده کرده‌اند. برای مثال در [۱۲] شاخص‌ها و گزینه‌های ابعاد کیفیت خدمات مبتنی بر رایانش ابری با استفاده از روشهای فازی بررسی شدند و در [۱۳] با استفاده از روش تحلیل پوششی داده‌ها و فازی به ارزیابی عملکرد سیستم‌های برنامه‌ریزی منابع سازمان شرکت تولیدی پرداخته شده است. با توجه به مطالعات انجام شده، هیچ استاندارد

همچنین، برای ارزیابی اعتمادپذیری و امنیت، شاخص‌هایی مانند نفوذپذیری، تحمل به خطا و ... بررسی می‌شوند. همچنین می‌توان چرایی بهره‌گیری از منطق فازی در تعریف شاخص‌های کیفی نرم‌افزار و استفاده از آنها در ارزیابی کیفیت سیستم‌های خودانطباق [۷] را در چهار دلیل ذیل تبیین نمود:

- تعامل با پیچیدگی: سیستم‌های فازی قابلیت مدل‌سازی و مدیریت پیچیدگی سیستم‌های خودانطباق را دارا بوده و قادر به در نظر گرفتن عوامل مبهم و متغیر در فرآیند ارزیابی هستند.
- قابلیت استدلال مبتنی بر قوانین: سیستم‌های فازی از قوانین و قواعد قابل فهم که یا از تجربه و دانش افراد متخصص یا از طریق الگوریتم‌های بهینه‌گر بدست می‌آیند برای استدلال استفاده می‌کنند. این استدلال برای تصمیم‌گیری در مورد انطباق و تغییرات در سیستم مورد استفاده قرار می‌گیرد.
- توانایی مدل‌سازی تعاملات پیچیده: سیستم‌های فازی قادر به مدل‌سازی تعاملات پیچیده و تأثیرات چند ورودی و خروجی هستند. در محیط خودانطباق، ورودی‌ها و خروجی‌ها ممکن است به صورت غیرخطی و غیرقطعی باشند و سیستم‌های فازی قادر به مدل‌سازی این تعاملات پیچیده هستند.
- امکان استفاده از دانش کاربر: سیستم‌های فازی قابلیت استفاده از دانش کاربر را دارند. با در نظر گرفتن ترجیحات و نظرات کاربران در قالب توابع فایده فازی، سیستم‌های خودانطباق می‌توانند تصمیماتی را بر اساس دانش کاربر انجام دهند.

در راستای ارزیابی سیستم‌های خودانطباق، مطالعات و تحقیقات کمی وجود دارد، که بیشتر آنها به عملکرد کلی سیستم و میزان موفق بودن آن در برآورده ساختن تغییرات تمرکز دارند و به کیفیت و شاخص‌های کیفی این دسته از سیستم‌ها توجه خاصی نشده است. از آن جهت که کیفیت سیستم نرم‌افزار صرف‌نظر از اینکه خودانطباق باشد یا خیر، یک چالش و مسئله مهم در مواجهه با آنها است، لذا فراهم ساختن روشی برای ارزیابی سیستم‌های خودانطباق و به خصوص ارزیابی شاخص‌های کیفی یکی از مسائل ضروری است که با گسترش این دسته از سیستم‌ها نیاز به وجود و رسیدگی به آنها پیش از پیش احساس می‌شود.

در این مقاله سعی شده تا یک روش جدید برای ارزیابی سیستم خود انطباق با توصیف شاخص‌های کیفی مناسب براساس منطق فازی نوع دوم و تنظیم بهینه وزن‌های آنها توسط الگوریتم ژنتیک ارائه شود. این روش، می‌تواند به پژوهشگران در جهت ارزیابی و مقایسه کیفی سیستم‌های نرم‌افزاری خود انطباق کمک کند. روش پیشنهادی حاوی دو فاز است. در فاز اول، شاخص‌های مؤثر در عملکرد سیستم‌های نرم‌افزاری و مشخصات نرم‌افزارهای خودانطباق شناسایی

محققان در [۱۴] چالش‌های ارزیابی کیفیت در سیستم‌های خودانطباق را به صورت زیر بیان می‌کنند:

۱. **مشکلات معماری:** اولین چالش مربوط به تنوع معماری‌های ارائه شده توسط این سیستم‌ها است. به دلیل وجود مدل‌های معماری مختلف (متمرکز، توزیع شده) و همچنان روش‌های مختلف برای انطباق این معماری‌ها در زمان اجرا، توسعه روش‌های ارزیابی کیفیت عمومی کافی برای پشتیبانی از همه ترکیب‌ها، ایجاد مجموعه‌ای از راه‌حل‌های هدایت شده به سمت معماری‌های خاص، چالش برانگیز می‌شود.

۲. **مقادیر آستانه برای شاخص‌های کیفیت:** در مورد خود وظیفه ارزیابی کیفیت، از طریق مطالعه ادبیات مشاهده می‌شود که برخی از تحقیقات توسعه‌یافته بر اساس شاخص‌های کیفیت، مقادیر مرجعی را برای تفسیر نتایج اندازه‌گیری ارائه نمی‌دهند، که ممکن است ارزیابی و تضمین کیفیت را تحت تأثیر قرار دهد.

۳. **پیچیدگی:** پیچیدگی این سیستم‌ها عاملی است که بر چندین راه‌حل ارزیابی کیفیت تأثیر می‌گذارد.

۴. **ابزارهای پشتیبانی:** چالش فعلی ناشی از پیچیدگی این سیستم‌ها، کمبود ابزارهای خودکار برای کمک به وظایف آزمون و ارزیابی کیفیت است.

آنها همچنین برخی از راه‌حل‌های این چالش‌ها را به شرح ذیل بیان کرده‌اند:

- توسعه ابزار برای پشتیبانی از آزمون سیستم‌های خودانطباق و فعالیت‌های ارزیابی کیفیت آنها
- تعریف حد آستانه برای شاخص‌های کیفیت
- ارزیابی شاخص‌های ضعیف مطالعه شده (به عنوان مثال، قابل حمل و مناسب بودن عملکرد) و تجزیه و تحلیل تاثیر آنها بر سیستم‌های خودانطباق
- تعریف ویژگی‌های جدید و شاخص‌های کیفی متناسب با جنبه‌های پویای این سیستم‌ها
- توسعه رویکردها برای اولویت‌بندی شاخص‌های کیفیت با توجه به عملیات و حوزه‌های این سیستم‌ها
- نگاشت تضاد بین شاخص‌های کیفیت

رایبولت و همکاران [۶] یک طبقه‌بندی برای ارزیابی کیفیت سیستم‌های خودانطباق ارائه می‌دهند. بر اساس این طبقه‌بندی، ارزیابی کیفیت در این سیستم‌ها می‌تواند بر اساس ویژگی‌های کیفی یا معیارهای نرم‌افزاری انجام شود. ارزیابی‌های مبتنی بر ویژگی‌های کیفی برای ارزیابی کل سیستم یا بخشی از آن با در نظر گرفتن معنادار ویژگی‌های کیفی مانند کارایی، بهینه‌سازی و قابلیت اطمینان

در خصوص مدل‌های کیفی برای ارزیابی سیستم خودانطباق وجود ندارد. در بیشتر مطالعات با توجه به مقالات و سیستم‌های خودانطباق موجود اقدام به توسعه مدل کیفی دلخواه خود نموده‌اند. مدل‌های کیفی بوهم، McCall و ISO 9126 و دیگر استانداردها نیز، حیطة عمل گسترده‌ای دارند و اگرچه می‌توانند در ارزیابی سیستم‌های خودانطباق بکار روند، اما برای این امر کافی نیستند.

جدول ۲. تحقیقات و مطالعات انجام شده برای ارزیابی کیفیت نرم‌افزار

مقاله	سال انتشار	توضیحات
[15]	۲۰۰۸	استفاده از منطق فازی، مفهوم فرآیند سلسله مراتبی تحلیل (AHP) جهت ارزیابی کیفیت نرم‌افزار
[16]	۲۰۱۰	استفاده از منطق فازی جهت ارزیابی نرم‌افزار
[17]	۲۰۱۱	استفاده از منطق فازی و مدل ISO9126 برای ارزیابی کیفیت سیستم نرم‌افزاری.
[18]	۲۰۱۴	استفاده از منطق فازی، Analytical Network Process (ANP) و Analytical Hierarchical Process (AHP) برای ارزیابی کیفیت نرم‌افزار
[19]	۲۰۱۵	استفاده از منطق فازی و تابع عضویت مثلثی جهت ارزیابی کیفیت سیستم نرم‌افزاری.
[20]	۲۰۱۵	استفاده از منطق فازی برای ارزیابی کیفیت نرم‌افزار بر طبق مدل‌ها و استانداردهای کیفی
[21]	۲۰۱۷	۱) ایجاد یک مدل کیفیت نرم‌افزار عمومی بر اساس چندین استاندارد کیفیت نرم‌افزار موجود. ۲) ارائه یک رویکرد ارزیابی سراسری جدید از مدل ارزیابی نرم‌افزار با استفاده از معیارهای اندازه‌گیری و منطق فازی. ۳) پیشنهاد یک الگوریتم نمونه‌سازی برای استخراج مدل ارزیابی نرم‌افزار خاص از مدل‌های کیفیت نرم‌افزار عمومی.
[22]	۲۰۱۹	یک مدل فازی سلسله مراتبی چهار سطحی جدید برای ارزیابی کیفیت نرم‌افزار پیشنهاد می‌کند.
[23]	۲۰۱۸	روش فرآیند تحلیل سلسله مراتبی فازی (FAHP) با روش باکلی (Buckley) برای ارزیابی تأثیر قابلیت نگهداری و زیرمجموعه‌های آن در کیفیت نرم‌افزار در سطح معماری پیشنهاد شده است.
[24]	۲۰۲۰	طراحی و توسعه یک رویکرد ترکیبی برای تجزیه و تحلیل نرم‌افزار و برآورد هزینه هر مدل نرم‌افزاری با ترکیب منطق فازی و مدل کوکومو ۲ (COCOMO-II).
[25]	۲۰۲۲	ارائه برخی از معیارهای شباهت جدید سیستم‌های فازی شهودی که با ترکیب تابع نمایی توابع عضویت و تابع منفی توابع غیرعضویت ساخته شده است. همچنین پیشنهاد یک معیار آنتروپی جدید به عنوان یک مرحله برای محاسبه وزن شاخص‌ها در مدل تصمیم‌گیری چند معیاره (MCDM) و استفاده از آن برای ارزیابی کیفیت پروژه‌های نرم‌افزاری.
[26]	۲۰۲۱	کیفیت محصول نرم‌افزار بر اساس سیستم استنتاج فازی در استاندارد ISO اندازه‌گیری می‌شود.
[27]	۲۰۲۳	پیشنهاد یک روش جدید ترکیبی برای آزمایش تأثیرگذاری پارامتر (PIT) در توسعه نرم‌افزار با تبدیل داده‌های زبانی به دست آمده از کارشناسان به اعداد فازی مثلثی و استفاده از یک تکنیک فرآیند سلسله مراتبی تحلیلی فازی- تکنیک برای اولویت سفارش با شباهت به راه‌حل ایده آل (AHP-) (TOPSIS) برای ارزیابی و اولویت‌بندی PIT ها به طور سیستماتیک.

در [۳۴] محققان به بررسی چگونگی تجمع معیارهایی که برای ارزیابی کیفیت مدل‌های ویژگی پیشنهاد شده‌اند، می‌پردازند. آن‌ها برای تجمع این معیارها از تئوری منطق فازی استفاده کردند. معیارهای تجمع‌شده جدید را می‌توان برای ارزیابی جنبه‌های مختلف و پیچیده یک مدل ویژگی، مانند اندازه، پایداری، انعطاف‌پذیری و پویایی، به کار برد. علاوه بر این، برای ارزیابی استفاده از معیارهای تجمع‌شده جدید، آن‌ها را در مدل‌های ویژگی مختلف به کار بست. یافته‌های آن‌ها نشان می‌دهد که معیارهای تجمع‌شده می‌توانند به مهندس حوزه‌ی مربوطه در ارزیابی قابلیت نگهداری مدل‌های ویژگی کمک کنند.

ذی‌نفعان انسانی اغلب از تناقضات پنهان بین شاخص‌های کیفی آگاه نیستند. به منظور رفع این مسئله، محققان در [۳۵] یک رویکرد ارائه می‌دهند که از تکنیک‌های یادگیری ماشین (کاهش بعد، خوشه‌بندی و یادگیری درخت تصمیم) برای توضیح دادن استدلال پشت برنامه‌ریزی خودانطباق استفاده می‌کنند. رویکرد آن‌ها بر تناقضات بین شاخص‌های کیفی تمرکز کرده که منجر به انتخاب وزن‌ها در توابع فایده و تولید برنامه‌های متفاوت می‌شود. آن‌ها به انسان‌ها در درک تناقضات شاخص‌های کیفی کمک می‌کنند.

در [۳۶] یک رویکرد جدید به نام CHESS ارائه شده که با استفاده از مهندسی آشوب، سیستم‌های خودانطباق را ارزیابی می‌کند. معماری ارائه شده در این مقاله یک نمای کامل از استفاده از شرط‌نچ از طریق دو مورد مطالعه مبتنی بر میکروسرویس را ارائه می‌دهد: یک مورد مطالعه مربوط به دفتر هوشمند و یک برنامه نمونه موجود به نام Yelb. این معماری شامل یک سرویس مدیریت سامانه، یک سرویس خودنظارتی و پنج سناریوی تزریق خطا که شامل خطاهای زیرساختی و خطاهای عملکردی است، می‌باشد. هر یک از این اجزا قابلیت گسترش یا جایگزینی آسان جهت استفاده از رویکرد CHESS در مطالعه موردی جدید را دارند.

همانطور که مشاهده می‌شود نگرش‌های ارزیابی متنوعی تاکنون ارائه شده‌اند که باید گفت، روش ارزیابی اولیه از جمله روش‌های مبتنی بر سناریو نمی‌توانند از سیستم خودانطباق پشتیبانی کنند. با توجه به روش‌ها و مطالعات موجود نیز، اگرچه، روش‌های مختلفی، با توجه به جنبه‌های مختلفی از سیستم خودانطباق اقدام به ارزیابی اینگونه از سیستم‌ها کردند، اما هیچ یک از آنها مختص این سیستم‌ها نیستند [۳۷]. از طرف دیگر، بیشتر نگرش‌های ارزیابی استفاده شده برای سیستم‌های خودانطباق در تحقیقات موجود، به عملکرد سیستم توجه دارند و عملکرد سیستم را مورد ارزیابی قرار می‌دهند و روشی برای ارزیابی کیفیت سیستم نرم‌افزاری خودانطباق ارائه نمی‌دهند. با این اوصاف، تلاش برای ارزیابی کیفیت سیستم خودانطباق و همچنین مشخص کردن یک مدل کیفی برای ارزیابی این دسته از

به کار می‌روند. ارزیابی‌های مبتنی بر معیارهای نرم‌افزاری برای ارزیابی کیفیت سیستم از یک دیدگاه محلی (مانند معیاری مرتبط با یک جزء یا عامل) یا از یک دیدگاه کلی (مانند معیاری مرتبط با کل سیستم) مورد استفاده قرار می‌گیرند.

کاری که در [۲۸] ارائه شده، رویکردی را برای انجام تطبیق‌پذیری در سیستم‌های خودانطباق مبتنی بر کامپوننت با معماری متمرکز بر اساس تعیین ویژگی‌ها، اندازه‌گیری و بهینه‌سازی خواص و ویژگی‌های کیفی در مدل‌های ویژگی ارائه می‌کند. نویسندگان گزارش می‌دهند که برخورد با یک سیستم توزیع‌شده پیچیده خواهد بود، زیرا مدیریت ویژگی‌های کیفی (مانند امنیت، در دسترس بودن، قابلیت مقیاس‌پذیری و مصرف باتری) دشوارتر خواهد شد.

سرال و همکاران [۲۹] ادعا می‌کنند که ارزیابی مکانیزم‌های خودانطباق که تضمین‌کننده حداکثر رضایت کاربر (مانند کارایی، راحتی، صرفه‌جویی در انرژی و غیره) هستند، به دلیل تفاوت‌های ذاتی بین کاربران و تغییر زمینه عملیاتی سیستم، یک چالش عمده است.

معین‌فر و همکاران [۳۰] گزارش می‌کنند که یکی از چالش‌های دستیابی به قابلیت اطمینان در سیستم‌های خودانطباق، مربوط به رسیدگی به مسائل محدودیت زمانی و فضایی است که محدودیت‌های مهمی برای این سیستم‌ها به شمار می‌روند. برای ارزیابی سیستم‌های خودانطباق، باید توجه داشت که معیارها و ویژگی‌های قابل ارزیابی متعددی وجود دارد. بر همین اساس، فرصت‌های پژوهشی همچنین می‌تواند به مقادیر مرجع برای معیارهای کیفیت، همانطور که در [۳۱] ذکر شده است، مرتبط باشد. از این منظر، انجام مطالعاتی مانند آزمایش‌ها و کاربرد تکنیک‌های یادگیری ماشین برای تعریف آستانه‌هایی برای معیارهای کیفیت در سیستم‌های خود-انطباق را می‌توان پیشنهاد کرد.

در مواجهه با سیستم‌های خودانطباق، ارزیابی مکانیزم‌های تطبیق‌پذیری می‌تواند به ویژگی‌ها و معیارهای کیفیتی بستگی داشته باشد که در زمان طراحی و زمان اجرا قابل اندازه‌گیری هستند [۳۲]. ارزیابی در زمان طراحی به راه‌حلی می‌پردازد که به معماری یک سیستم نرم‌افزاری برای پشتیبانی از قابلیت خود-انطباق‌پذیری اشاره می‌کند. ارزیابی ویژگی‌های خود-انطباق‌پذیر در زمان اجرا بیشتر به معیارهایی مرتبط است که به دنبال ارائه ارزیابی کیفی یا کمی از نرم‌افزار هستند. به عنوان مثال، کاری که توسط کادوم و همکاران [۳۳] ارائه شده است، مجموعه‌ای از معیارها و اندازه‌گیری‌ها را برای سیستم‌های خود-انطباق‌پذیر نشان می‌دهد، که از میان آن‌ها می‌توان به تأخیر، معیاری برای سنجش عملکرد در دیدگاه زمان اجرا، اشاره کرد که نشان‌دهنده تغییر بین زمان پاسخ با وجود قابلیت خودانطباقی و بدون وجود مکانیزم‌های خودانطباقی است.

حالی که در سیستم‌های نرم‌افزاری دقیق مدل‌سازی تقریباً غیرممکن است. دوم اینکه، طراحی کنترل‌کننده‌ها و استراتژی‌های کنترل نیاز به دانش بسیار زیاد دارد که این امر برای محققان و کارکنان علوم کامپیوتر بسیار سخت است [۳۸]. لذا با توجه به این امر که کیفیت نرم‌افزار کمیتی کیفی است، در نتیجه منطق فازی می‌تواند یک راه-حل بسیار مناسب برای این مسئله باشد. به همین دلیل تحقیقاتی نیز در این خصوص فراهم آمده است که در جدول ۲ به آن اشاره شده است. اما تاکنون تحقیق و مطالعه‌ای برای ارزیابی سیستم خودانطباق توسط منطق فازی نوع دوم صورت نگرفته است و تحقیقات فوق برای سیستم‌هایی غیر از سیستم‌های خودانطباق مورد استفاده قرار گرفتند. روش ارزیابی فازی می‌تواند مشکل عامل‌های چندگانه، مشکلات سیستماتیک، علمی، عملیاتی و غیره را حل کند. بنابراین، یک نوع روش ارزیابی پیشرفته جامع را فراهم می‌آورد [۴۴]. منطق فازی در کنار مزیت‌های فراوانی که به همراه می‌آورد، معایبی نیز دارد. همین امر سبب ایجاد فازی نوع ۲ شده است. استفاده از فازی نوع ۲ در ارزیابی سیستم‌های نرم‌افزاری می‌تواند با رفع مشکل عدم قطعیت در سیستم فازی نوع ۱، به دقیق‌تر شدن ارزیابی کمک بسیاری کند.

سیستم‌ها امری ضروری و لازم است.

دلیل اصلی استفاده از منطق فازی، سهولت و انعطاف‌پذیری آن است. منطق فازی می‌تواند مشکلات ناشی از اطلاعات مبهم و ناتمام را حل کند و توابع غیرخطی پیچیده را مدل نماید. منظور از اطلاعات مبهم در واقع عدم قطعیت و نسبی بودن مقادیر کیفی است. منطق کلاسیک درک بالایی از سیستم را نیاز دارد. در حالی که منطق فازی برای مدلسازی یک سیستم پیچیده توسط سطح بالایی از انتزاع از دانش و اطلاعات کلی بدون درک عمیق از سیستم مورد استفاده قرار می‌گیرد. در مدل‌های ارزیابی، برخی از داده‌های ورودی به طور معقول تعیین می‌شوند (به عنوان مثال ارزیابی بر طبق شاخص‌های کیفی، توابع ارزیابی جزئی برای شاخص‌های کمی، انتخاب یک نوع مناسب تجمیع، وزن‌دهی به شاخص‌ها یا در نهایت، توصیف مبتنی بر قانون بین مقادیر شاخص‌ها و ارزیابی کلی آنها). از آنجا که عدم قطعیت ویژگی متداول هرگونه اطلاعات خبره است، نظریه مجموعه فازی یک ابزار ریاضی مناسب برای ایجاد چنین مدل‌هایی می‌باشد. محدودیت‌های متعددی از تحقیقات کنونی در ارزیابی سیستم‌های خودانطباق وجود دارد. اول اینکه، به مدل نرم‌افزاری نیاز است، در

جدول ۳. مستندات و اطلاعات منتشر شده درباره ارزیابی و عملکرد سیستم InSync

اطلاعات	سال انتشار	نام مستند
سیستم‌های کنترل ترافیک مختلفی توسعه یافتند و اکنون در حال استفاده هستند از جمله آنها می‌توان به سیستم‌های زیر اشاره کرد: <ul style="list-style-type: none"> • SCATS: در دهه ۱۹۹۰ تولید شد. این سیستم یک سیستم انطباق ترافیک است، که ترافیک را در جاده‌ها شناسایی می‌کند. • ACS-Lite: در سال ۲۰۰۶ برای کنترل و بهبود ترافیک چهار راه‌ها معرفی شد. • InSync: در سال ۲۰۰۹ کاربرد شد. از پردازش تصویر و پردازش فیلم استفاده می‌کند تا ترافیک را شناسایی کند. این مستند سیستم‌های کنترل ترافیک را که تاکنون طراحی، ساخته و اجرایی شدند مورد بررسی و ارزیابی قرار داده است. این مستند، بر طبق نوع عملکرد و همچنین برخی از معیارهای کیفی این سیستم‌ها را مورد مقایسه و تحلیل قرار داده است. شاخص پایداری، ثبات، هزینه را از این مستند استخراج کرد.	۲۰۱۳	Criteria For the Selection and Application of Advanced Traffic Signal Systems [۳۹]
این مستند به سیستم insync پرداخته است. این گزارش ابتدا به بررسی عملکرد سیستم insync می‌پردازد و سپس عملکرد آن را مورد ارزیابی قرار می‌دهد. متریک‌هایی مانند کارایی، رضایت از سیستم، را از این مستند استخراج کرد.	۲۰۱۲	Comparative Evaluation of Insync and Time-Of-Day Signal Timing Plans Under Normal and Varied Traffic Conditions [۴۰]
این مستند به بررسی سیستم insync و تشریح عملکرد و موارد کاربرد آن می‌پردازد. همچنین متریک‌های حسگر فعال، استفاده از منابع، قابلیت استفاده را استخراج کرد.	۲۰۱۲	InSync Adaptive Traffic Control System for the Veterans Memorial Hwy Corridor on Long Island, NY [۴۱]
این مستند یک ارزیابی کلی از عملکرد سیستم‌های کنترل ترافیک ارائه می‌دهد. از آن می‌توان مقادیر شاخص‌ها و یا متریک‌های هزینه، قابلیت اطمینان، قابلیت دسترسی، قابلیت نگهداری، قابلیت نصب، قابلیت پیکربندی (مجدد)، تاخیر را استخراج کرد.	۲۰۱۰	Adaptive-Traffic-Signal-Control Systems in the United States: updated summary and comparison [۴۲]
این مستند نیز به بررسی سیستم‌های کنترل ترافیک از جمله Scats و InSync می‌پردازد. این مستند علاوه بر تشریح عملکرد سیستم به ارزیابی آن نیز می‌پردازد. از جمله شاخص‌ها یا متریک‌های قابل استخراج در این سیستم می‌توان به ایمنی، قابلیت یکپارچگی، سودمندی سیستم اشاره کرد.	۲۰۱۳	Analysis of Adaptive Traffic Control Systems and design of a Decision Support System for better choice [۴۳]

عملیاتی می‌تواند شامل ورودی کاربر نهایی، حسگرها و دستگاه‌های سخت‌افزاری خارجی، یا برنامه دیگر شود [۴۵]. محققان معمولاً از تعاریف زیر برای سیستم‌های خودانطباق استفاده می‌کنند:

- خودانطباقی، توانایی سیستم برای تنظیم رفتار خود در پاسخ به محیط است. پیشوند «خود» نشان می‌دهد که سیستم‌ها به طور مستقل و خود مختار (یعنی، با حداقل دخالت یا بدون دخالت) تصمیم می‌گیرند چگونه بر طبق زمینه و محیط، خود را انطباق و یا سازماندهی کنند [۴۶].
- در چارچوب سیستم‌های چند مدله، انطباق یک رویه یا روش برای تعویض بین مدل‌ها است. بنابراین انطباق می‌تواند به عنوان توانایی یک سیستم برای دستیابی به اهداف خود در یک محیط متغیر تعریف شده و با انتخاب مجدد و تغییر بین مدل‌ها، تعریف شود [۴۷].
- یک سیستم خودانطباق شامل یک سیستم حلقه بسته است (یعنی خود را در زمان اجرا با استفاده از بازخورد ناشی از تغییرات مداوم سیستم اصلاح می‌کند). مفهوم سیستم خودانطباق به نیازمندی‌های کاربر، خصوصیات سیستم و خصوصیات محیطی بستگی دارد. نرم‌افزار خودانطباق مستلزم قابلیت اطمینان بالا، استحکام زیاد، انطباق پذیری زیاد و قابلیت دسترس پذیری بالا است [۴۸].
- سیستم خودانطباق، رفتار خود را ارزیابی می‌کند و عملکرد خود را بر طبق ارزیابی‌ها تغییر می‌دهد [۴۵].

از نرم‌افزار خودانطباق، انتظار می‌رود که نیازمندی‌های زمان اجرا (یعنی اهداف) خود را انجام دهند، حتی زمانی که با تغییرات داخلی و خارجی رو به رو می‌شود. از یک سیستم خودانطباق انتظار می‌رود که تغییراتی بر اهدافش اثر دارند، کشف کند و فعالیت مناسب را با هزینه و زمان قابل قبول انجام دهد. این اهداف می‌توانند از طریق نظارت سیستم بر محیط برای کشف تغییرات جهت گرفتن تصمیم مناسب، تحقق یابد [۴۹].

۳-۲- ارزیابی کیفیت نرم‌افزار

کیفیت نرم‌افزار یک جنبه مهم برای توسعه‌دهندگان، کاربران و مدیران پروژه است. محققان بسیاری سعی در طراحی مدلی مناسب برای تعریف کیفیت نرم‌افزار از دیدگاه‌های مختلف داشتند. مطالعات کیفیت نرم‌افزار یک مجموعه سیستماتیک و برنامه‌ریزی شده از فعالیت‌ها برای اطمینان از سودمند بودن نرم‌افزار را فراهم می‌کند و شامل زیرعنوان‌های مختلف مانند تامین کیفیت نرم‌افزار^۱، کنترل کیفیت^۲ و مهندسی کیفیت^۳ است. بر طبق IEEE، کیفیت نرم‌افزار

شایان ذکر است که باتوجه به عدم دسترسی به جزئیات سیستم‌های خودانطباق موجود در تحقیقات مختلف، در این مقاله از سیستم کنترل ترافیک ترافیک با نام InSync که یک سیستم خودانطباق کنترل ترافیک (ATCS) بوده و حاوی شاخص‌های متعدد کیفی می‌باشد استفاده شده است. این سیستم حاوی شاخص‌ها و متریک‌های کیفی متعددی است که با کمک فرد خبره از مستندات موجود برای این سیستم [۳۹، ۴۰، ۴۲، ۴۳] جمع‌آوری شده که این اطلاعات به اختصار در جدول ۳ بیان شده‌اند. از این مستندات ۳۲ شاخص شامل پایداری، زمان تنظیم کوتاه، پایان دهی، مقیاس پذیری، دانه‌بندی، پیشگویی، سرعت پاسخ خودمختار، هزینه، درجه خودمختاری، خودپیکربندی، خودبهینگی، خود حفاظتی، بازبودن، قابلیت حمل، قابلیت نگهداری، قابلیت عملیاتی، دقت، تخطی کم، سازگاری، حساسیت، پرهیز از شکست، استحکام، امنیت، دقت تصمیم‌گیری، سطح خودمختاری، خود بهبودی، خودآگاهی، آگاهی به زمینه، ثبات، رضایت، قابلیت استفاده و قابلیت اعتماد بدست آمده است. همچنین ۴۷ متریک شامل حسگر فعال، زمان پاسخ (اجرا)، زمان انتظار، استفاده از منابع (بهره‌وری منابع)، سادگی، قابلیت گسترش، یکپارچگی، مناسب بودن، قابلیت همکاری درونی، تاخیر، توان عملیاتی، ظرفیت، رفع شکست و نقص، تشخیص شکست و نقص، پرهیز از شکست و نقص، قابلیت دسترسی، وفق‌پذیری ساختار داده، وفق‌پذیری محیطی سخت افزاری، وفق‌پذیری محیط اجرایی، وفق‌پذیری محیط نرم‌افزاری، درگاه‌های کاربر پسند نصب آسان، آسانی تنظیمات مجدد، همزیستی قابل دسترس، زمان تصمیم، پیچیدگی رفتار، دانه‌بندی ریزادداشت، ثبات، رضایت، فراگیری عملیات، دانش، قدرت اتصال، پیچیدگی تعاملات، محرمانگی، ساختار دهی مجدد، عدم انکار، سرعت تصمیم‌گیری، زمان شکست، زمان بازیابی، سودمندی، قابلیت بازسازی و شروع مجدد، قابلیت تحلیل، قابلیت تغییر، قابلیت آزمایش، قابلیت یادگیری، قابلیت یادسپاری و قابلیت درک از این مستندات استخراج می‌شوند.

۳- مسئله پژوهش

هر مسئله تحقیق، ادبیات و مفاهیم منحصر بفرد خود را دارا می‌باشد و در این بخش به تشریح مفاهیم مرتبط با مسئله شامل سیستم‌های خودانطباق، ارزیابی کیفیت نرم‌افزار و فازی نوع ۲ می‌پردازیم.

۳-۱- سیستم‌های خودانطباق

بر طبق تعریف آژانس DARPA نرم‌افزار خودانطباق، رفتار و تغییرات رفتار خود را برای بهبود کارایی و عملکرد بهتر، ارزیابی کرده و آنرا برای پاسخ به تغییرات در محیط عملیاتی اصلاح می‌کند. محیط

³ Quality engineering

¹ Software quality assurance

² Quality control

۳-۳- فازی نوع ۲

دکتر لطفی‌زاده ابداع‌کننده منطق فازی^۱ معتقد به لزوم در نظر گرفتن ابهام در ماهیت علم بود. برخلاف دیگران که معتقد بودند که باید تقریب‌ها را دقیق‌تر کرد تا بهره‌وری افزایش یابد، ایشان معتقد به لزوم ساخت مدل‌هایی که ابهام را به عنوان بخشی از سیستم مدل نماید بودند. برخلاف منطق ارسطویی که بر درست یا نادرست بودن گزاره‌ها استوار است، در منطق فازی، جملاتی هستند که مقداری درست و مقداری نادرست هستند. برای مثال، جمله «هوا سرد است» یک گزاره منطقی فازی می‌باشد که درستی آن گاهی کم و گاهی زیاد است. گاهی همیشه درست و گاهی همیشه نادرست و گاهی تا حدودی درست است.

پایگاه قواعد فازی، مجموعه‌ای از قواعد اگر-آنگاه فازی را نشان می‌دهد. موتور استنتاج فازی این قواعد را به یک نگاهت از مجموعه‌های فازی در فضای ورودی به مجموعه‌های فازی در فضای خروجی بر اساس اصول منطق فازی ترکیب می‌کند. در سیستم‌های فازی با فازی‌ساز ابتدا در ورودی سیستم متغیرهای عددی تبدیل به متغیرهای فازی می‌شوند و به موتور استنتاج وارد می‌شوند و خروجی موتور استنتاج که یک کمیت فازی است توسط غیر فازی‌ساز به یک عدد ثابت تبدیل می‌شود. توابع عضویت فازی بیان‌کننده میزان تعلق یک عدد به یک متغیر زبانی می‌باشد و این میزان تعلق همواره بین ۰ تا ۱ می‌باشد. یک مجموعه فازی نوع ۱، یک متغیر منفرد $x \in X$ است که می‌تواند به صورت زیر ارائه شود:

$$A = \{x, \mu_A(x) \mid \forall x \in X\} \quad (1)$$

A می‌تواند به صورت زیر نیز تعریف شود:

$$A = \int_{x \in X} \mu_A(x) / x \quad (2)$$

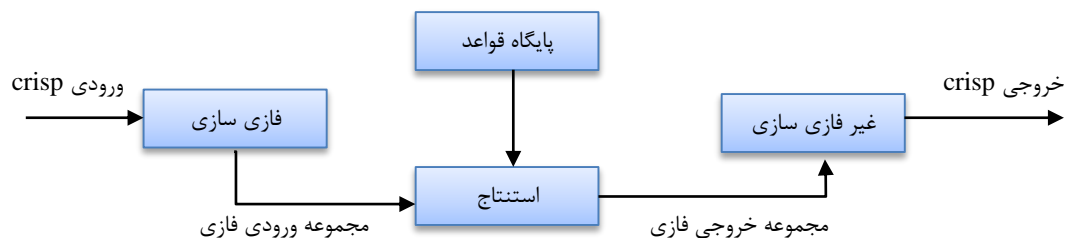
توابع عضویت انواع مختلفی دارند که به تابع عضویت مثلثی، دوزنقه‌ای، گوسی و سیگموئید می‌توان اشاره کرد. روش کار سیستم فازی نوع ۱ در شکل ۲ نشان داده شده است. پروفیسور لطفی‌زاده در سال ۱۹۷۵ مجموعه‌های فازی نوع ۲ را به عنوان توسعه‌ای از مجموعه‌های فازی معرفی کرد. مجموعه‌های فازی نوع ۲ دارای درجه عضویت‌های فازی می‌باشند؛ از این رو، به آنها مجموعه‌های فازی - فازی نیز می‌گویند که در برخورد با عدم قطعیت‌ها توانایی کاهش اثر و مدل کردن آنها را دارند. یک مجموعه فازی نوع ۲ در مواجهه با عدم قطعیت‌های زبانی برای ارائه‌ی اطلاعاتی بیشتر از یک درجه عضویت ساده، اندازه‌ای از پراکندگی را حول یک عدم قطعیت زبانی در نظر می‌گیرد. درحقیقت، مجموعه فازی نوع ۱ تقریب درجه اول عدم قطعیت و مجموعه فازی نوع ۲ تقریب درجه دوم عدم قطعیت هستند.

مجموعه‌ای از شاخص‌های نرم‌افزار است و به صورت زیر تعریف می‌گردد [۱۷]:

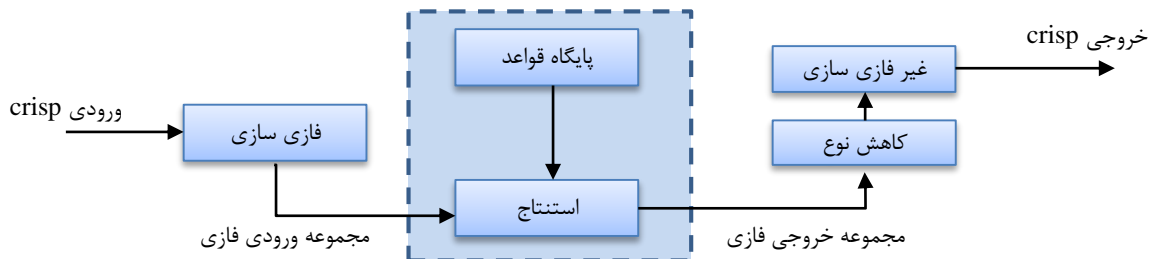
۱. درجه‌ای که سیستم، مؤلفه‌ها یا محصول، نیازمندی‌ها را تامین می‌کند.
 ۲. درجه‌ای که سیستم، مؤلفه‌ها یا فرآیند، نیازهای مشتریان و کاربران را تامین می‌کند.
 ۳. کیفیت شامل هدایت شاخص‌ها برای رضایت نیازمندی‌هایشان می‌شود.
- در IEEE ارزیابی اینگونه توصیف شده است: «فرآیند تحلیل یک بخش از نرم‌افزار برای تعیین اختلاف بین شرایط مورد نیاز و وضعیت موجود آن بخش (برای تعیین خطاها) و تست مؤلفه‌های نرم‌افزاری آن است». از این توصیف ما می‌توانیم ببینیم ارزیابی را می‌توانیم رویکردی در ارزیابی و سنجش نرم‌افزارها و ازجمله نرم‌افزارهای خودانطباق در نظر بگیریم.

نیازمندی‌های کیفی مانند کارایی و بهینه‌سازی، در معماری و نرم‌افزار سیستم‌های خودانطباق بسیار مهم هستند. سیستم‌های خودانطباق خواهان تکامل مداوم هستند و هدف آنها اطمینان از کیفیت انجام وظایف زمان اجرا و پیکربندی مجدد است. در نتیجه، ارزیابی کیفی از انطباق‌ها و موفقیت در انجام تغییرات زمان اجرا مسئله مهمی است که باید در این سیستم‌ها در نظر گرفته شود [۶]. در این راستا، مهندسی نرم‌افزار چندین مکانیزم و استاندارد برای ارزیابی کیفیت سیستم‌ها از مرحله طراحی معماری تا پیاده‌سازی راه‌حل، فراهم کرده است [۶]. مکانیزم‌های ارزیابی در سیستم‌های خودانطباق می‌توانند وابسته به موارد زیر باشند:

- **شاخص‌های کیفی:** برای ارزیابی کیفیت سیستم یا بخش‌هایی از آن استفاده می‌شود. که به طور خاص بر ویژگی‌های کیفی معنی‌دار برای سیستم‌های خودانطباق تمرکز می‌کند. از جمله این ویژگی‌ها می‌توان به کارایی و بهینه‌سازی اشاره کرد.
- **متریک‌های نرم‌افزار:** برای ارزیابی کیفیت انطباق از دیدگاه محلی یا سراسری استفاده می‌شود. دیدگاه محلی می‌تواند شامل انطباق یک محصول باشد و دیدگاه سراسری می‌تواند به زمان برای انطباق یا درجه عدم تمرکز اشاره کند. دو مرحله مهم در مکانیزم‌های انطباق وجود دارد:
 - الف) در زمان طراحی، در طول توسعه سیستم که مکانیزم انتخاب می‌شود.
 - ب) در زمان اجرا، که سیستم در حالت تمام عملیاتی است و کیفیت انطباق باید به طور پیوسته نظارت شود.



شکل ۲. روش کار سیستم فازی نوع ۱ [۵۰]



شکل ۳. سیستم فازی نوع ۲ [۵۱]

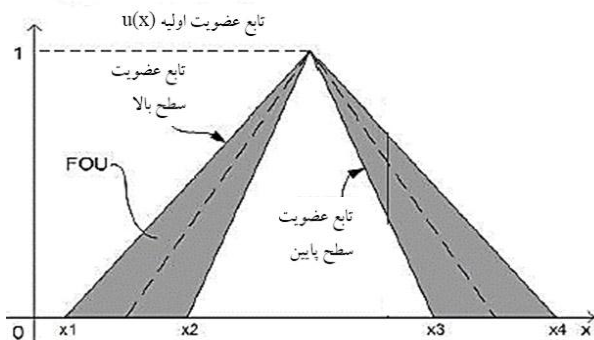
که \iint اجتماع روی همه x و u های قابل قبول را مشخص می‌کند.

$$\tilde{A} = \{((x, u), \mu_{\tilde{A}}(x, u)) \mid \forall x \in X, \forall u \in J_x \subseteq (0,1)\} \quad (6)$$

هنگامی که همه توابع عضویت ثانویه، از مجموعه فازی نوع ۲ یکی شدند، مجموعه فاصله فازی نوع ۲ به صورت زیر مشخص می‌شود.

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} 1 / (x, u) \quad J_x \subseteq (0,1) \quad (7)$$

عدم قطعیت در تابع عضویت اولیه متشکل از یک منطقه محدود که به عنوان FOU^2 شناخته می‌شود، که اجتماع همه توابع عضویت اولیه است [۵۳]. شکل ۴ نمایی از تابع عضویت مثلثی در منطق فازی نوع ۲ را نمایش می‌دهد.



شکل ۴. نمایی از تابع عضویت مثلثی در فازی نوع ۲ [۵۰]

به این ترتیب، مجموعه‌های فازی نوع ۲ در کاهش اثر عدم قطعیت در قوانین فازی بهتر عمل می‌کنند، علاوه بر کاهش اثر عدم قطعیت در قوانین فازی، به علت فازی بودن توابع عضویت در مجموعه‌های فازی نوع ۲ امکان مدل کردن عدم قطعیت‌های زبانی و داده‌ها به نحو مؤثری بهبود پیدا می‌کند [۵۲]. مشابه با فازی نوع ۱، سیستم فازی نوع ۲ نیز شامل فازی‌ساز، پایگاه قوانین، موتور واسط فازی، و پردازش‌گر خروجی است. شکل ۳ نمای یک سیستم فازی نوع ۲ را نشان می‌دهد.

پردازش‌گر خروجی، شامل کاهش دهنده نوع و غیر فازی‌ساز است و مجموعه خروجی فازی نوع ۱ یک عدد درست را (از غیر فازی‌کننده) تولید می‌کند. فازی نوع ۲ توسط قوانین if-then مشخص می‌شود. اما مجموعه‌های استنتاجی خود را دارد. مجموعه فاصله^۱، تا حد زیادی توابع عضویت ثانویه را ساده می‌کند. مجموعه فازی نوع ۲ که با \tilde{A} مشخص می‌شود، توسط توابع عضویت نوع ۲ یعنی $\mu_{\tilde{A}}(x, u)$ مشخص می‌شوند، که $x \in X$ و $u \in J_x \subseteq (0,1)$ است یعنی

$$\mu_{\tilde{A}}(x, u): X \rightarrow [0,1] \quad (3)$$

همچنین J_x تابع عضویت اولیه در فاصله (۰ و ۱) است و u مقدار تابع عضویت است. در نتیجه:

$$0 \leq \mu_{\tilde{A}}(x, u) \leq 1 \quad (4)$$

علاوه بر این \tilde{A} به صورت زیر نیز بیان می‌شود:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x, u) / (x, u) \quad J_x \subseteq (0,1) \quad (5)$$

² Foot-print of uncertainty

^۱ interval sets

۴-۳- شاخص‌های کیفی سیستم‌های خودانطباق

شاخص‌های کیفی که برای ارزیابی نرم‌افزارهای خودانطباق در نظر گرفته شده‌اند، مؤلفه‌های انتزاعی مانند سودمندی، عملکرد، خودانطباق، استحکام و مؤلفه پیشتیبانی می‌باشند که نمایش‌دهنده مفاهیمی سطح بالا هستند و به راحتی قابل اندازه‌گیری نیستند. برای ارزیابی این مولفه‌ها نیاز است که آن‌ها به کمیت‌های سطح پایین و قابل اندازه‌گیری نگاشت شوند، که در این نگاشت هر مؤلفه به چند متریک قابل اندازه‌گیری تبدیل می‌شود. شکل ۵ نمای کلی از شاخص‌های مهم در ارزیابی سیستم خودانطباق را نشان می‌دهد.

۴- روش پیشنهادی

روند عملکرد رویکرد پیشنهادی که ترکیب جدید ژنتیک-فازی نوع ۲ است در شکل ۶ ترسیم شده و مشتمل بر هفت مرحله اصلی می‌باشد. الگوریتم ژنتیک وظیفه تنظیم پارامترهای ۳ مرحله را برعهده داشته و نتیجه مرحله ششم یعنی محاسبه خطا مقدار تخمینی ارزیابی سیستم توسط رویکرد پیشنهادی از مقدار واقعی بعنوان تابع برازش و تعیین کیفیت پارامترهای پیشنهادی به منظور اعمال در فرایند بهبود الگوریتم ژنتیک در نظر گرفته می‌شود. در ادامه مراحل ششگانه رویکرد فازی پیشنهادی تشریح شده است.

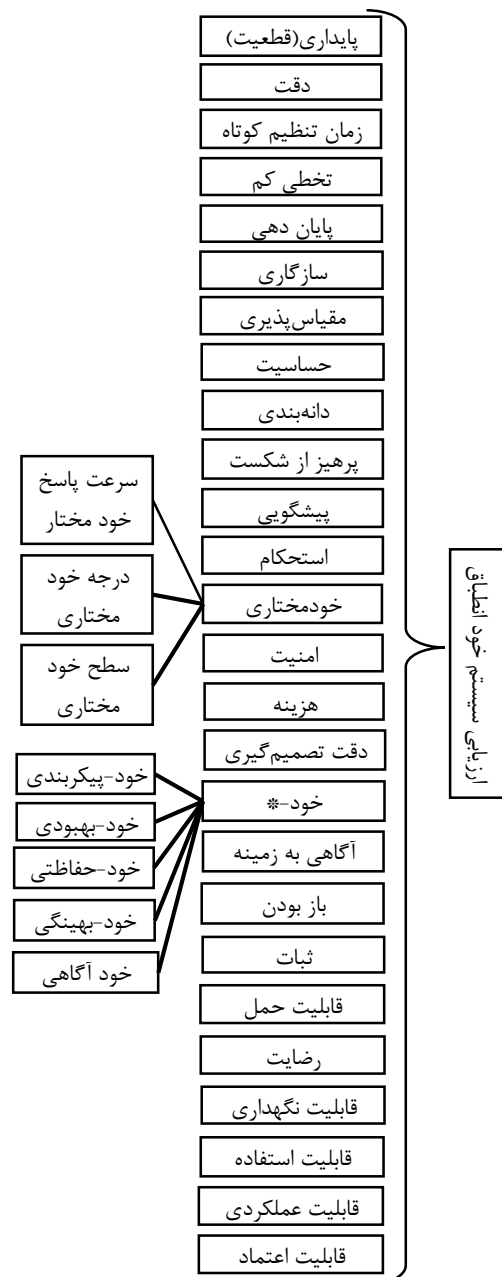
مرحله اول: شناسایی شاخص‌های کیفی

در نخستین مرحله رویکرد پیشنهادی می‌بایست تمامی نیازمندی‌ها و شاخص‌های کیفی برای ارزیابی سیستم خودانطباق شناسایی شوند. این شاخص‌های کیفی از مدل‌های مک‌کال، بوهم و سایر استانداردهای موجود در این زمینه و همچنین از مقالات ارائه‌دهنده سیستم خودانطباق جمع‌آوری شده است. سپس متریک‌های مؤثر در بدست آوردن مقدار این شاخص‌ها مشخص می‌شوند. هر شاخص می‌تواند توسط چند متریک مورد ارزیابی و محاسبه قرار گیرد. از آن جهت که شاخص‌ها و متریک‌های بسیاری جهت ارزیابی کیفی وجود دارد، شاخص‌ها و متریک‌های مؤثر بر شاخص‌های کیفی مورد استفاده در این مقاله با توجه به جدول ۳ از مقالات ارائه‌دهنده سیستم خودانطباق جمع‌آوری شده است.

به منظور تعریف و نمایش شاخص‌ها و متریک‌ها از اعداد فازی نوع ۲ در رویکرد پیشنهادی استفاده می‌شود. همانطور که شکل ۷ نشان می‌دهد، تابع عضویت فازی نوع ۲ از دو سطح بالایی و پایینی تشکیل شده است. مقادیر سطح بالا و پایین از چهار مقدار تشکیل شده‌اند. مطابق شکل، مقدار تابع عضویت سطح بالا برابر است با: (A, D, C, B) و مقدار سطح پایین برابر است با (A_2, D, C_2, B_2) . همانطور که در شکل ۶ مشاهده می‌شود مقادیر مربوط به پارامترهای توابع عضویت فازی با استفاده از الگوریتم ژنتیک بدست می‌آیند.

مرحله دوم: تخصیص رتبه به شاخص

در این مرحله باید به شاخص‌های انتخاب شده یک رتبه اختصاص داده شود که میزان اهمیت آن را نمایش می‌دهد. بدین ترتیب، شاخص‌های کیفی مهم‌تر شناسایی می‌شوند. میزان اهمیت هر شاخص توسط مقادیر خیلی کم، کم، متوسط، زیاد و خیلی زیاد مشخص می‌شود که تنظیم بهینه پارامترهای توابع عضویت آنها توسط الگوریتم ژنتیک انجام می‌پذیرد. برای این مهم هر یک از پارامترهای توابع عضویت ۵ مقدار کیفی فوق را به عنوان یک ژن در ساختار کروموزم در نظر می‌گیریم.



شکل ۵. شاخص‌های مؤثر در ارزیابی کیفی سیستم خودانطباق

نیست. جهت رتبه‌دهی به متریک‌ها نیز مانند مرحله قبل عمل می‌شود. یعنی هر متریک با توجه به مقادیر خیلی کم، کم، متوسط، زیاد و خیلی زیاد رتبه‌دهی می‌شوند. ۳۵ ژن بخش دوم یک کروموزوم به تعیین مقادیر پارامترهای توابع عضویت این ۵ مقدار کیفی براساس نمایش عدد فازی نوع ۲ اختصاص می‌یابد.

نتیجه این مرحله دو ماتریس با چهار ستون و m سطر است. که m نشان دهنده تعداد متریک‌های مورد استفاده برای ارزیابی شاخص مذکور می‌باشد. یکی از این ماتریس‌ها به مقادیر سطح بالا و دیگری به مقادیر سطح پایین اشاره دارد که به ترتیب با rateup و ratelow نشان داده می‌شوند.

$$\text{rateup} = \begin{bmatrix} rup_{1,1} & rup_{1,2} & rup_{1,3} & rup_{1,4} \\ \vdots & \vdots & \vdots & \vdots \\ rup_{m,1} & rup_{m,2} & rup_{m,3} & rup_{m,4} \end{bmatrix}$$

$$\text{ratelow} = \begin{bmatrix} rlow_{1,1} & rlow_{1,2} & rlow_{1,3} & rlow_{1,4} \\ \vdots & \vdots & \vdots & \vdots \\ rlow_{m,1} & rlow_{m,2} & rlow_{m,3} & rlow_{m,4} \end{bmatrix}$$

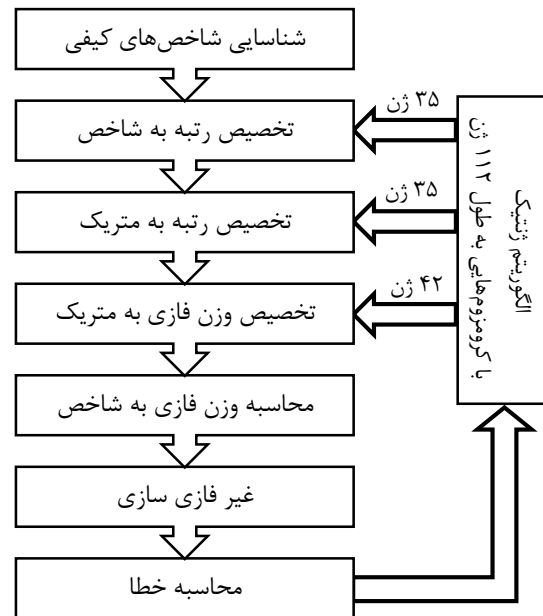
مرحله چهارم: تخصیص وزن فازی به متریک

این مرحله مربوط به انتخاب وزن فازی مناسب برای هر متریک است. این وزن‌ها نیز توسط الگوریتم ژنتیک محاسبه می‌شوند. هر متریک دارای نوع وزن‌دهی مخصوص به خود است. اما به طور کلی می‌تواند شامل مقادیر برآورده نشدن، خیلی کم، کم، متوسط، زیاد و خیلی زیاد قرار گیرد. به منظور کد کردن مقادیر پارامترهای توابع عضویت این ۶ مقدار کیفی براساس ساختار فازی نوع ۲ به تعداد ۴۲ ژن در قالب بخش سوم یک کروموزوم در نظر گرفته می‌شود.

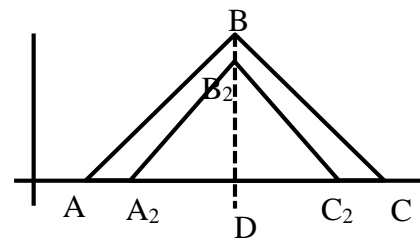
خروجی این مرحله دو ماتریس است که مانند ماتریس رتبه‌دهی حاوی ۴ ستون و m سطر است که m تعداد متریک‌های مؤثر در ارزیابی شاخص را نمایش می‌دهد. ماتریس upperWeightValue به مقادیر سطح بالا و ماتریس lowerWeightValue به مقادیر سطح پایین اشاره دارد.

$$\text{upperWeightValue} = \begin{bmatrix} wup_{1,1} & wup_{1,2} & wup_{1,3} & wup_{1,4} \\ \vdots & \vdots & \vdots & \vdots \\ wup_{m,1} & wup_{m,2} & wup_{m,3} & wup_{m,4} \end{bmatrix}$$

$$\text{lowerWeightValue} = \begin{bmatrix} wlow_{1,1} & wlow_{1,2} & wlow_{1,3} & wlow_{1,4} \\ \vdots & \vdots & \vdots & \vdots \\ wlow_{m,1} & wlow_{m,2} & wlow_{m,3} & wlow_{m,4} \end{bmatrix}$$



شکل ۶. روند عملکرد رویکرد پیشنهادی



شکل ۷. نمونه تابع عضویت مثلثی برای مقادیر فازی شاخص‌ها و متریک‌ها

باتوجه به شکل ۷، برای کد کردن این توابع عضویت نیاز به ۳۵ ژن می‌باشد. هر ژن نیز با مقدار حقیقی کد می‌شود. نتیجه این مرحله، دو بردار ۴ بیتی است که مقدار رتبه شاخص را به شکل فازی بیان کرده و به صورت زیر نمایش داده می‌شود:

$$\text{rateLow} = [rlow1, rlow2, rlow3, rlow4]$$

$$\text{rateUp} = [rup1, rup2, rup3, rup4]$$

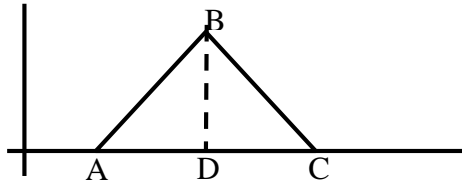
مرحله سوم: تخصیص رتبه به متریک

برای ارزیابی هر شاخص کیفی ممکن است چند متریک وجود داشته باشد، سیستم ممکن است با توجه به نوع عملیات خود به برخی از متریک‌ها اهمیت بیشتری دهد. به همین جهت وزن و تأثیر هر متریک برای ارزیابی شاخص، باید مشخص باشد. لذا در این مرحله به منظور تعیین میزان اهمیت متریک‌هایی که برای ارزیابی شاخص وجود دارند، به آنها رتبه تخصیص داده می‌شود. این امر می‌تواند در کاهش پیچیدگی محاسبات کمک نماید. در صورتی که یک شاخص تنها از یک متریک جهت ارزیابی استفاده کند، نیازی به تعیین رتبه

مرحله پنجم: محاسبه وزن فازی شاخص

$$\text{Value} = [A \ D \ C \ B] \quad (12)$$

این بردار، یک مثلث را در نمودار نمایش می‌دهد که در شکل ۸ نشان داده شده است. مقدار سطح زیر این نمودار، مشخص کننده مقدار غیرفازی است که برای بدست آوردن مقدار آن، از انتگرال استفاده می‌شود.



شکل ۸. حاصل عملیات کاهش نوع

$$\begin{aligned} \text{main value} &= \frac{\int_A^B \left(\frac{D-A}{B-A}\right) z \, dz + \int_B^C \left(\frac{C-D}{C-B}\right) z \, dz}{\int_A^B \left(\frac{D-A}{B-A}\right) dz + \int_B^C \left(\frac{C-D}{C-B}\right) dz} \quad (13) \end{aligned}$$

بدین ترتیب مقدار حاصل از عملیات غیرفازی سازی بدست می‌آید. این مقدار بین ۰ تا ۱،۲ است و هرچه مقدار بیشتر باشد به معنای آن است که شاخص مورد نظر بیشتر برآورده شده است.

مرحله هفتم: محاسبه خطا

از آنجایی که الگوریتم ژنتیک وظیفه انتخاب بهینه پارامترهای توابع عضویت مقادیر فازی رتبه‌های شاخص، رتبه‌های متریک و وزن‌های فازی متریک‌ها را توسط یک کروموزوم به طول ۱۱۲ ژن برعهده دارد، لازم است کیفیت هر کروموزوم در فرایند الگوریتم ژنتیک توسط یک تابع برازش محاسبه گردد. لذا خطای هر شاخص را از طریق تفاضل مقدار بدست آمده از عملیات غیرفازی سازی (رابطه ۱۳) بعنوان مقدار تخمین زده شده هر شاخص از مقدار واقعی بدست آورده و کمینه‌سازی مجموع آنها ($\sum |e|$) را بعنوان تابع برازش در نظر می‌گیریم. یعنی هرچه مقدار خروجی تخمین زده شده توسط سیستم فازی پیشنهادی به مقدار واقعی که توسط فرد خبره داده می‌شود، نزدیک‌تر باشد، برازش بیشتر خواهد بود. برای این منظور از مقادیری که قبلاً توسط فرد خبره داده شده استفاده می‌شود.

۵- پیاده‌سازی و ارزیابی

این بخش به منظور پیاده‌سازی و ارزیابی روش پیشنهادی مشتمل بر سه بخش اصلی می‌باشد. بخش اول به معرفی داده مورد استفاده یعنی یک سیستم کنترل ترافیک می‌پردازد. بخش دوم به تشریح پیاده‌سازی روش پیشنهادی و ابزارهای مورد استفاده در پیاده‌سازی آن اشاره دارد. بخش سوم به ارزیابی روش پیشنهادی بر روی سیستم کنترل ترافیک تمرکز دارد.

با توجه به مقادیر بدست آمده در مرحله چهارم، برای هر شاخص یک وزن فازی نوع ۲ محاسبه می‌شود. در این مرحله متریک‌ها و مقادیر آنها به عنوان ورودی و شاخص کیفی به عنوان خروجی در نظر گرفته شده است و مقادیر داده شده برای متریک‌ها، تعیین کننده مقدار شاخص کیفی هستند. شاخص کیفی نیز می‌تواند یکی از مقادیر برآورده نشده، خیلی کم، کم، متوسط، زیاد و خیلی زیاد باشد.

$$\begin{aligned} \text{weightFactorUp} &= \text{rateup.upperWeightValue} \quad (8) \end{aligned}$$

برای ماتریس سطح پایین نیز به صورت زیر عمل می‌شود:

$$\begin{aligned} \text{weightFactorLow} &= \text{ratelow.lowerWeightValue} \quad (9) \end{aligned}$$

لازم به ذکر است که در هر دو فرمول بالا از ضرب نقطه‌ای استفاده شده است. یعنی درآیه‌های متناظر در هر دو ماتریس با هم ترکیب می‌شوند.

مرحله ششم: غیرفازی سازی

در این مرحله میزان کیفی شاخص به مقدار کمی تبدیل می‌شود و بدین ترتیب مشخص می‌شود که تا چه اندازه سیستم توانسته شاخص کیفی مورد نظر را برآورده سازد. برای غیرفازی سازی در فازی نوع ۲، دو مرحله اساسی وجود دارد:

- کاهش نوع: تبدیل فازی نوع ۲ به فازی نوع ۱
 - غیرفازی سازی توسط یکی از روش‌های مربوط به فازی نوع ۱
- برای کاهش نوع مینیمم مقدار هر ستون از ماتریس‌های مربوط به وزن فازی بدست می‌آید و در نهایت یک ماتریس با ۱ سطر و چهار ستون ایجاد می‌شود. هر درآیه از این ماتریس نشان دهنده کمترین مقدار در ستون متناظر است. این عمل برای هر دو ماتریس weightFactorLow و weightFactorUp انجام می‌شود. بدین ترتیب وزن فازی شاخص مشخص می‌شود. سپس رتبه و میزان اهمیت این شاخص در این وزن ضرب می‌شود:

$$\text{FactorUpValue} = \min(\text{factorlowvalue}, [], 1) * \text{ratefactorLow}; \quad (10)$$

$$\text{FactorLowValue} = \min(\text{factoruppvalue}, [], 1) * \text{ratefactorUp}; \quad (11)$$

سپس از دو بردار فوق برای هر ستون مقدار ماکزیمم محاسبه می‌شود. در نهایت یک بردار با یک سطر و چهار ستون تولید می‌گردد، مانند:

۵-۱- داده مورد استفاده

در ادامه به منظور محاسبه وزن فازی شاخص‌ها، لازم است به ۴۷ متریک موثر در ارزیابی سیستم وزن‌های فازی اختصاص یابد که در جدول الف-۵ (پیوست الف) پرسشنامه انجام این مهم آمده است. در این جدول ستون‌های No به معنی برآورده نشدن، VL به معنای خیلی کم برآورده شدن، L به معنای کم برآورده شدن، M برآورده شدن در حد متوسط، و H برآورده شدن در حد زیاد و VH برآورده شدن در حد خیلی زیاد را نشان می‌دهد.

جهت پیاده‌سازی سیستم پیشنهادی از نرم‌افزار متلب ۲۰۲۰ استفاده شده است. برای اجرای برنامه نیز از سیستمی با ویندوز ۷، core i5 و RAM 8GB استفاده شده است. به منظور تنظیم بهینه مقادیر رتبه‌بندی شاخص‌های کیفی، متریک‌ها و وزن‌های فازی برای متریک‌ها از نسخه کلاسیک الگوریتم ژنتیک استفاده شد که تعداد جمعیت اولیه ۱۰۰ کروموزوم، تعداد تکرار ۱۰۰۰ نسل و نرخ جهش و بازترکیب به ترتیب ۰٫۲ و ۰٫۸ استفاده شده است. در جدول ۴ پارامترهای توابع عضویت فازی بدست آمده توسط الگوریتم ژنتیک برای رتبه‌بندی شاخص‌ها نشان داده شده است. همچنین برای رتبه‌دهی به متریک‌ها مقادیر جدول ۵ و مقادیر وزن فازی برای هر متریک نیز در جدول ۶ بدست آمده است.

جدول ۴. مقادیر رتبه‌بندی شاخص‌های کیفی

مقدار کیفی	محدوده در جدول	
	سطح بالایی	سطح پایینی
خیلی کم	[0 0 0.32 0.95]	[0 0 0.2 0.45]
کم	[0.1 0.44 0.75 1]	[0.32 0.44 0.61 0.5]
متوسط	[0.4 0.67 0.89 1]	[0.49 0.67 0.8 0.5]
زیاد	[0.81 0.9 1.1 1]	[0.78 0.9 1 0.5]
خیلی زیاد	[0.9 1 1.2 1]	[0.96 1 1.06 0.5]

جدول ۵. مقادیر رتبه‌بندی متریک‌ها

مقدار کیفی	محدوده در جدول	
	سطح بالایی	سطح پایینی
خیلی کم	[0 0 0.49 0.92]	[0 0 0.27 0.44]
کم	[0.12 0.39 0.61 0.97]	[0.29 0.39 0.55 0.5]
متوسط	[0.47 0.7 0.89 1]	[0.45 0.7 0.82 0.54]
زیاد	[0.73 0.9 1.12 1]	[0.81 0.9 1 0.54]
خیلی زیاد	[0.9 1 1.2 1]	[0.93 1 1.1 0.5]

جدول ۶. مقادیر وزن فازی برای متریک‌ها

وزن فازی	محدوده در جدول	
	سطح بالایی	سطح پایینی
برآورده نشدن	[0 0 0.09 1]	[0 0 0.04 0.5]
خیلی کم	[0 0.1 0.41 1]	[0.04 0.1 0.36 0.5]
کم	[0.12 0.4 0.72 1]	[0.3 0.4 0.57 0.5]
متوسط	[0.44 0.7 0.9 1]	[0.54 0.7 0.81 0.5]
زیاد	[0.72 0.9 1.12 1]	[0.82 0.9 1 0.5]
خیلی زیاد	[0.9 1 1.2 1]	[0.95 1 1.075 0.5]

سیستم کنترل ترافیک یکی از معمول‌ترین انواع سیستم‌های خودانطباق است که تحقیقات بسیاری در حوزه آن انجام گرفته است. جزئیات اطلاعات سایر سیستم‌های خودانطباق موجود در تحقیقات دیگر یا در دسترس نبودند یا اطلاعات کمی در مورد آنها وجود داشت و یا اینکه نمی‌توانستند شاخص‌های کیفی زیادی را شامل شوند. به همین دلیل، در خصوص ارزیابی، یک سیستم کنترل ترافیک به نام InSync انتخاب شده است. این سیستم، حاوی چند دوربین است که در مکان‌های خاصی از خیابان نصب می‌شوند. دوربین‌ها به عنوان حسگر یا سنسور عمل می‌کنند. اطلاعات ترافیک و ازدحام خودروها را جمع‌آوری کرده و سپس این اطلاعات را برای سیستم مرکزی ارسال می‌کنند. با توجه به تصاویر ارسال شده و دیگر اطلاعات موجود، سیستم مرکزی تصمیم می‌گیرد که چراغ راهنمایی و رانندگی را برای کدام مسیر سبز و برای کدام مسیر قرمز کند. در نتیجه سیستم فوق، یک سیستم حاوی بخش‌های سخت‌افزاری و نرم‌افزاری بسیاری است. برای ارزیابی این سیستم از مستنداتی که حاوی اطلاعات جامع به همراه ارزیابی‌ها و عملکردهای مربوط به آن بوده استفاده شده است. InSync دو جنبه مهم دارد:

- به طور خودکار توسط سیگنال‌های محلی تنظیم می‌شود.
- با توجه به اطلاعات سیگنال‌ها رانندگان را نسبت به ترافیک و مسائل ترافیکی در جاده آگاه می‌سازد.

۵-۲- پیاده‌سازی

بر طبق روش پیشنهادی که بیان شد، پرسشنامه یکی از موارد مهم و کلیدی در این روش محسوب می‌گردد. با توجه به شاخص‌ها و متریک‌ها، می‌توان یک دسته از شاخص‌های کیفی را به عنوان متریک‌های مهم در ارزیابی سیستم‌های خودانطباق مطرح کرد. با این حال، بسته به نوع سیستم و نیازمندی‌های آن، برخی از شاخص‌های کیفی دارای اهمیت بیشتری هستند. بنابراین مشخص کردن رتبه شاخص‌های کیفی می‌تواند به منزله اولویت و اهمیت آن شاخص در سیستم باشد.

رتبه و مقدار فازی هر شاخص کیفی توسط پرسشنامه مشخص می‌گردد. جداول الف-۱ و الف-۲ (پیوست الف) فهرست ۳۲ شاخص کیفی و ۴۷ متریک موثر در ارزیابی سیستم و نحوه رتبه‌بندی آنها نمایش داده شده و پرسشنامه‌های مرتبط میزان رتبه این شاخص‌ها و متریک‌ها نیز در جداول الف-۳ و الف-۴ (پیوست الف) ارائه شده‌اند. در این جداول ستون No به معنی بی اهمیت بودن، ستون VL به معنای خیلی کم اهمیت، ستون L به معنای کم اهمیت است. همچنین ستون M اهمیت متوسط، ستون H اهمیت زیاد و ستون VH اهمیت خیلی زیاد را نشان می‌دهد.

۵-۳- ارزیابی

روش پیشنهادی را می‌توان از دو جهت با سایر روش‌ها و مقالات موجود مورد مقایسه قرار داد:

۱. شاخص‌های کیفی و متریک‌های مورد استفاده
۲. چارچوب و روش کار

از جهت شاخص‌های کیفی مورد استفاده، این شاخص‌ها با سایر مقالات و مدل‌های کیفی در جدول ۷ مقایسه شده‌اند. بطور آشکار این جدول نشان می‌دهد که روش پیشنهادی جامعیت بیشتری نسبت به سایر مدل‌ها و مقالات موجود دارد و طی آن سعی شده تا سیستم‌های خودانطباق از دو بُعد شامل بُعد نرم‌افزاری این سیستم و بُعد خودانطباق آنها مورد مقایسه قرار گیرد. این در حالی است که اکثر تحقیقات موجود تنها به یک بعد می‌پردازند. از طرفی اگر سیستم Insync را تنها با مدل‌های کیفی مطرح در جدول ۷ ارزیابی کنیم، یا به عبارت بهتر از شاخص‌های کیفی موجود استفاده کنیم، نمی‌توان به طور جامع این سیستم را مورد ارزیابی قرار داد. لذا به منظور جامعیت بخشی بیشتر به فرایند ارزیابی روش پیشنهادی از نمودار رادار استفاده کرده‌ایم. برای این منظور اطلاعات سیستم Insync را با توجه به تنها شاخص‌های کیفی مطرح شده در یک مقاله، مورد ارزیابی قرار می‌دهیم و روش پیشنهادی را تنها با همان شاخص‌های کیفی اجرا می‌کنیم. برای اینکار سه مقاله [۵۴]، [۲] و

[۵۵] انتخاب شدند. نتیجه در نمودار رادار شکل‌های ۹ تا ۱۱ و نمودار رادار روش پیشنهادی نیز در شکل ۱۲ آمده است.

همانطور که شکل‌های ۹ تا ۱۲ و جدول ۷ نمایش می‌دهند، روش پیشنهادی از جامعیت بیشتری برخوردار است و شاخص‌ها کیفیت متعددی را در بر می‌گیرد. اما نکته حائز اهمیت در ارزیابی کیفیت نرم‌افزار (چه خودانطباق باشد و چه نباشد) این است که تمام معیارها و شاخص‌های کیفی برای ارزیابی دارای اهمیت نیستند. به عبارت بهتر، همه این معیارها برای ارزیابی کیفیت نرم‌افزار به کار گرفته نمی‌شوند و در سیستم‌های مختلف دارای اولویت و اهمیت متفاوتی هستند. در واقع، شاخص‌های کیفی، دارای قطعیت نیستند و مهم بودن یک شاخص کیفی در یک یا چند سیستم نرم‌افزاری به منزله اهمیت آن در تمام سیستم‌های نرم‌افزاری نیست. اگرچه، در سیستم‌های خودانطباق مجموعه‌ای از شاخص‌های کیفی هستند که ویژگی و اساس این گونه از سیستم‌ها را فراهم می‌آورند (مانند پایداری، شاخص‌های خود-حساسیت، قابلیت پیشگویی، کارایی، هزینه، نفوذپذیری، تحمل به خطا و ...)، اما این شاخص‌ها بسته به نوع و عملکرد سیستم می‌توانند اهمیت متفاوتی در سیستم‌های مختلف داشته باشند. در نتیجه لازم است مکانیزم و راهکاری برای در نظر گرفتن میزان اهمیت شاخص‌های کیفی در چارچوب و روش کار در نظر گرفت.

جدول ۷. مقایسه شاخص‌های کیفی در روش پیشنهادی با سایر مقالات و مدل‌های کیفی

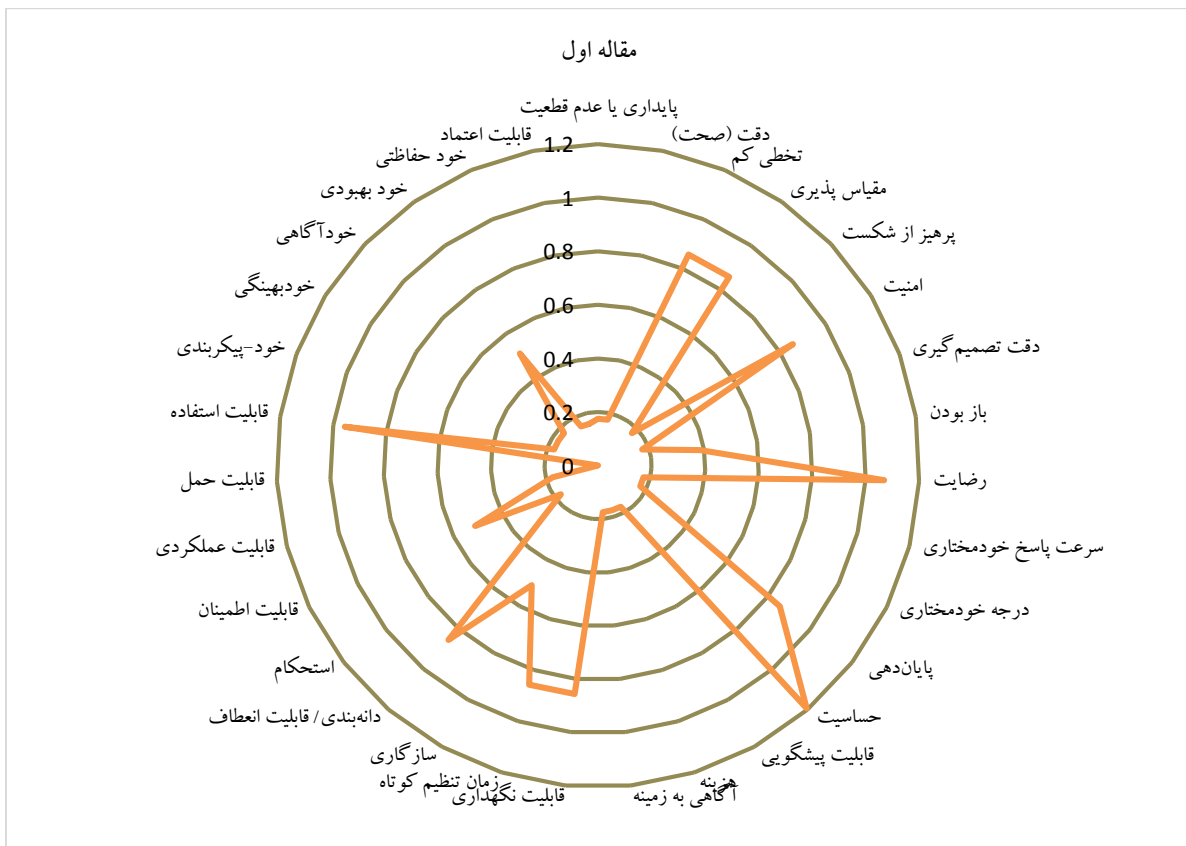
روش پیشنهادی	[۴]	[۲]	[۵۴]	[۵۷]	[۵۵]	[۳]	[۵۶]	ISO 9126	نوع	MCCall	مدل‌های کیفی / تحقیقات شاخص‌های کیفی / معیارها
✓	✓		✓	✓							پایداری یا عدم قطعیت
✓	✓		✓				✓	قابلیت نگهداری	✓	✓	دقت (صحت)
✓								قابلیت نگهداری	✓		قابلیت اصلاح
✓	✓		✓								تخطی کم
✓	✓		✓				✓				مقیاس پذیری
✓	✓	✓	✓								پرهیز از شکست
✓	✓		✓				✓				امنیت
✓	✓										دقت تصمیم‌گیری
✓	✓										باز بودن
✓	✓	✓				✓					رضایت
✓											سطح خودمختاری
✓	✓										سرعت پاسخ خودمختاری
✓	✓	✓	✓								درجه خودمختاری
✓	✓			✓							پایان‌دهی
✓	✓		✓								حساسیت
✓	✓										قابلیت پیشگویی
✓	✓	✓	✓								هزینه
✓	✓				✓						آگاهی به زمینه

روش پیشنهادی	[۵۴]	[۶]	[۵۵]	[۵۷]	[۴۵]	[۲]	[۵۶]	ISO 9126	نوع	MCCall	مدل‌های کیفی / تحقیقات شاخص‌های کیفی / معیارها
✓											ثبات
✓	✓			✓		✓	✓	✓	✓	✓	قابلیت نگهداری
✓	✓			✓							زمان تنظیم کوتاه
✓	✓			✓							سازگاری
✓	✓		✓				✓		✓	✓	دانه‌بندی / قابلیت انعطاف
✓	✓			✓							استحکام
✓	✓			✓							قابلیت اطمینان
✓	✓					✓	✓	✓			قابلیت عملکردی
✓	✓					✓	✓	✓	✓	✓	قابلیت حمل
✓	✓					✓	✓		✓	✓	قابلیت استفاده
✓	✓				✓						خود-پیکربندی
✓	✓				✓						خودبهبودی
✓	✓				✓						خودآگاهی
✓	✓				✓						خود بهبودی
✓	✓				✓						خود حفاظتی
✓	✓			✓		✓	✓	✓	✓	✓	قابلیت اعتماد
قابلیت نگهداری	✓							قابلیت نگهداری	✓	✓	قابلیت آزمایش
قابلیت اصلاح	✓								✓	✓	قابلیت استفاده مجدد
قابلیت عملکردی	✓									✓	قابلیت همکاری درونی
قابلیت استفاده / قابلیت پیشگویی / خودبهبودی	✓					✓	✓	✓	✓	✓	بهره‌وری
امنیت / سازگاری / پایداری	✓			✓				✓	✓	✓	یکپارچگی
قابلیت استفاده	✓								✓		قابلیت درک
هزینه / صحت / مقیاس‌پذیری / پایان دهی / تخطی کم / زمان تنظیم کوتاه / پایداری	✓			✓				✓			کارایی
خودبهبودی / قابلیت اطمینان / امنیت / استحکام	✓			✓						✓	قابلیت دسترسی
قابلیت استفاده	✓					✓					سودمندی
استحکام / پرهیز از شکست / پایداری	✓	✓		✓							ایمنی
				✓							جداناپذیری
			✓								کیفیت خدمات
			✓								قابلیت انطباق
کارایی	✓	✓	✓								زمان انطباق / زمان واکنش
سودمندی		✓									انطباق بهینه

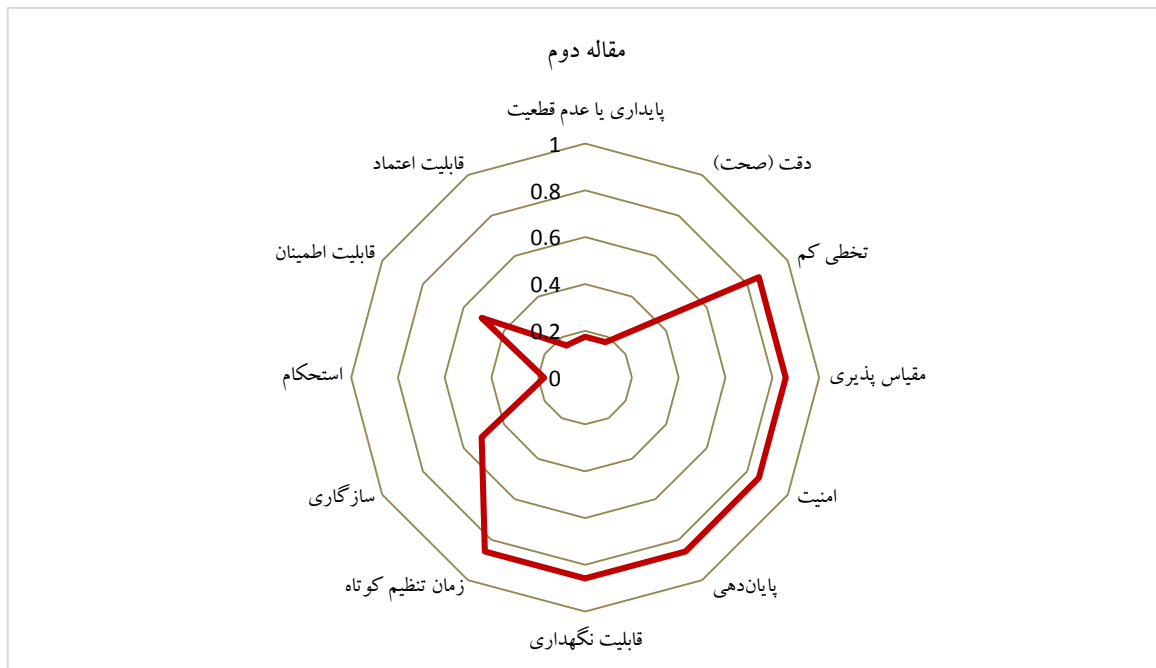
ارزیابی شود.

روش پیشنهادی ابتدا یک چک لیست از شاخص‌های کیفی موثر در ارزیابی سیستم خودانطباق فراهم می‌آورد. سپس این امکان را می‌دهد تا ارزیاب با توجه به عملکرد سیستم یا نظر ذینفعان سیستم، شاخص‌های کیفی مهم را مشخص کند و اولویت آنها را تعیین نماید. این امر می‌تواند در ارزیابی بهتر سیستم تاثیر بسزایی داشته باشد.

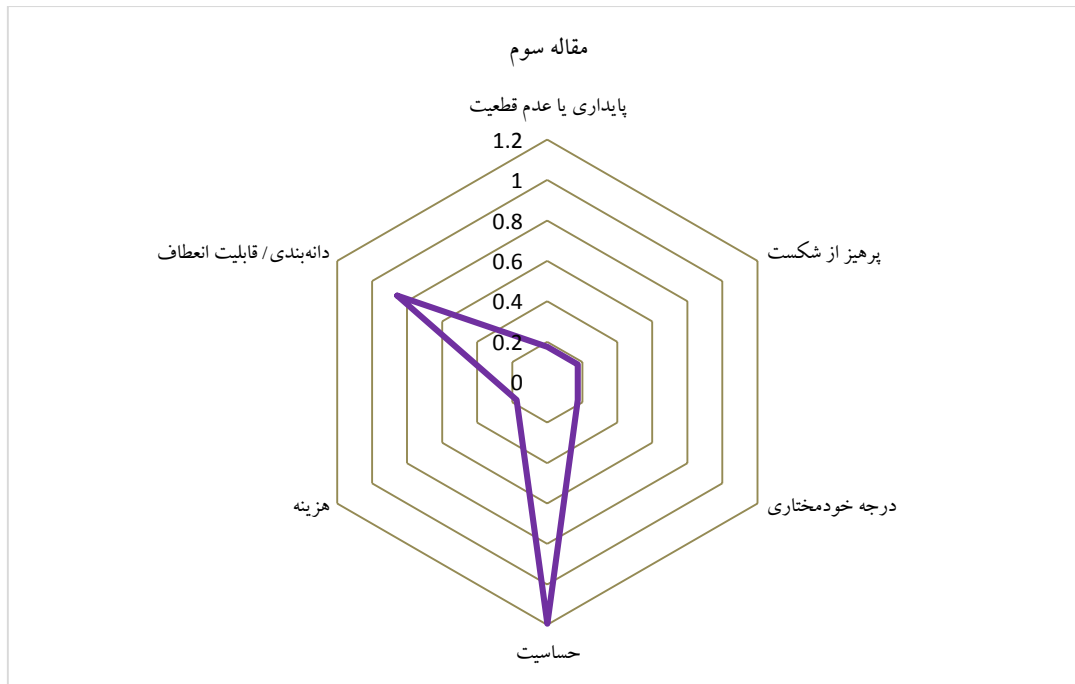
از دیدگاه چارچوب و روش کار نیز، جدول ۸ روش‌های بکار گرفته شده در مقالات موجود در حیطه ارزیابی کیفیت سیستم نرم‌افزاری خودانطباق نشان می‌دهد. با توجه به جدول ۸، مشاهده می‌شود که بیشتر مقالات توجهی به موضوع میزان اهمیت و اولویت شاخص‌های کیفی نسبت به یکدیگر نداشتند. استفاده از یک الگوی ثابت در ارزیابی سیستم‌های نرم‌افزاری می‌تواند سبب نادقیق شدن نتایج



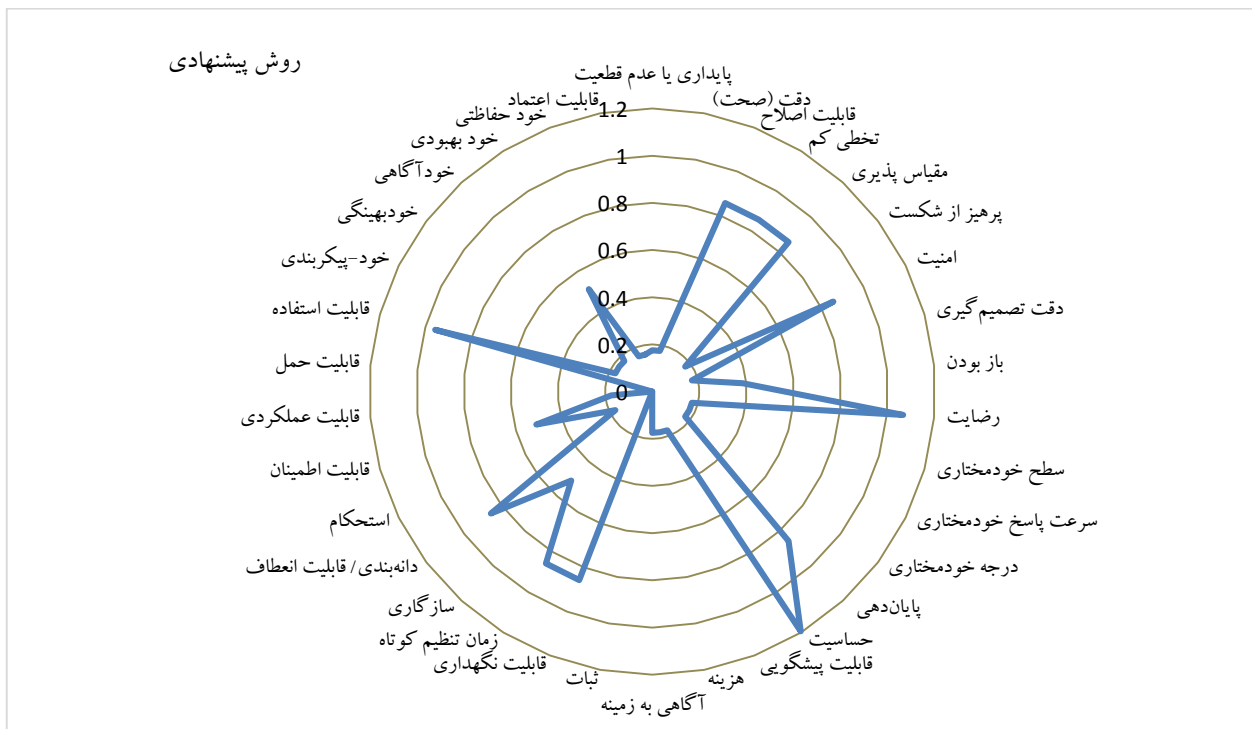
شکل ۹. نمودار رادار شاخص‌های کیفی در [۵۴]



شکل ۱۰. نمودار رادار شاخص‌های کیفی در [۲]



شکل ۱۱. نمودار رادار شاخص‌های کیفی در [۵۵]



شکل ۱۲. نمودار رادار شاخص‌های کیفی در روش پیشنهادی

از طرف دیگر، گاه بدست آوردن مقادیر لازم جهت انجام محاسبات دشوار و یا زمان‌بر است. روش مبتنی بر سناریو نیز یکی از روش‌های پرکاربرد در ارزیابی است، اما زمانی که کیفیت و شاخص‌های کیفی مطرح می‌شوند، این روش نیز نمی‌تواند کارایی مناسبی داشته باشد، زیرا تولید سناریو برای ارزیابی کیفی دشوار و یا گاه ناممکن است. روش لایه‌ای را نیز نمی‌توان برای تمام شاخص‌های کیفی بسط داد.

علاوه بر این، اکثر مقالات از فرمول‌های مشخص شده برای شاخص‌های کیفی استفاده کرده‌اند. حال اینکه، تمامی شاخص‌های کیفی را نمی‌توان با استفاده از فرمول و مدل‌های ریاضی بدست آورد. در بسیاری از موارد یافتن یک فرمول برای شاخص کیفی بسیار دشوار و گاه ناممکن است. برای مثال برای تعیین مقدار شاخص کیفی قابلیت استفاده، بدست آوردن فرمول تقریباً ناممکن است.

جدول ۸. مقایسه روش کار مقالات حوزه ارزیابی کیفیت سیستم

خودانطباق با روش پیشنهادی

مقالات	معیارها	میزان شاخص‌های کیفی مورد استفاده	استفاده از رتبه و اولویت	روش کار
[۵۷]	تعداد محدودی از شاخص‌های کیفی را در نظر گرفته است.	x	از فرمول استفاده	
[۲]	شش شاخص هزینه، پرهیز از شکست، درجه خودمختار بودن، انطباق بهینه، زمان انطباق/واکنش، ایمنی و پایداری	x	از فرمول استفاده	
[۵۴]	مجموعه کاملی از شاخص‌های کیفی در سیستم‌های خودانطباق	✓	از فرمول استفاده	
[۳۷]	قابلیت دسترسی	x	از سناریو استفاده	
[۵۸]	اعتبار و قابلیت اعتماد داده‌های ورودی، واسط‌های مورد استفاده با کاربر و محیط، تصمیمات انطباق، تراکنش‌های سیستم	x	از روش لایه‌ای استفاده	
[۶]	مجموعه ای از شاخص‌های کیفی	x	از فرمول استفاده	
[۳۳]	مجموعه ای از شاخص‌های کیفی	x	از فرمول استفاده	
روش پیشنهادی	مجموعه ای از شاخص‌های کیفی	✓	از منطق فازی استفاده	

* در جدول ۸ علامت ✓ به معنای وجود مکانیزم اولویت‌دهی و ✗ به معنای عدم وجود این مکانیزم است.

شاخص‌های کیفی، دارای مقادیر کیفی هستند، بنابراین اندازه‌گیری آنها دشوار و کمی‌سازی آنها سخت است. منطق فازی به علت توانایی کار با مقادیر کیفی، و مقادیر دارای ابهام بعنوان گزینه مناسب برای اندازه‌گیری کیفیت نرم‌افزار انتخاب شده و فازی نوع ۲ به علت ویژگی‌های متمایزی که نسبت به فازی نوع ۱ داشته و توانایی رفع عدم قطعیت را دارد، بعنوان نوآوری این پژوهش در ارزیابی کیفیت نرم‌افزار خودانطباق مد نظر قرار گرفت.

این مقاله، علاوه بر طرح یک روش برای ارزیابی سیستم خودانطباق مبتنی بر الگوریتم ژنتیک و فازی نوع ۲، یک مجموعه از شاخص‌های کیفی مؤثر در آن را نیز جمع‌آوری کرده است که می‌تواند به عنوان یک چک لیست در زمان طراحی این دسته از سیستم‌ها مورد استفاده قرار گیرد. با این حال هنوز مراحل و روش‌هایی جهت بهبود طرح پیشنهادی جاری وجود دارد. از جمله اینکه، روش پیشنهادی، تنها کیفیت سیستم پیشنهادی را می‌سنجد و بر عملکرد این دسته از سیستم‌ها توجه ندارد. برخلاف تحقیقات و مقالات دیگر که بر عملکرد سیستم توجه دارند تا کیفیت آن. بعنوان یکی از مسیرهای تحقیقاتی باز در پیشروی این مقاله می‌توان به طراحی یک مکانیزم ارزیابی پیشنهادی که هم به کیفیت و هم به عملکرد سیستم‌های خودانطباق بپردازد، اشاره نمود.

دوم اینکه، روش پیشنهادی برای ارزیابی کیفیت سیستم خودانطباق از تابع عضویت مثلثی استفاده کرده است، اما بررسی دیگر توابع عضویت در این راستا بعنوان دومین مسیر باز تحقیقات آتی می‌تواند مشخص نماید که آیا تابع عضویت مثلثی بهتر انتخاب بوده یا خیر.

مراجع

- [1] F. D. Macías-Escrivá, R. Haber, R. Del Toro, and V. Hernandez, "Self-adaptive systems: A survey of current approaches, research challenges and applications," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7267-7279, 2013.
- [2] F. Kneer and E. Kamsties, "A Framework for Prototyping and Evaluating Self-adaptive Systems - A Research Preview," 2016.
- [3] S. Sucipto and R. S. Wahono, "A Systematic Literature Review of Requirements Engineering for Self-Adaptive Systems," *Software Engineering & Applications*, vol. 1, no. 1, 2015.
- [4] M. Abufouda, "A Framework for Enhancing Performance And Handling Run-Time Uncertainty in Self-Adaptive Systems," *Software Engineering & Applications*, vol. 5, no. 1, 2014.
- [5] M. Salehie and L. Tahvildari, "Towards a goal-driven approach to action selection in self-adaptive software," *Software: Practice and Experience*, vol. 42, no. 2, pp. 211-233, 2012.
- [6] C. Raibulet, F. Arcelli Fontana, R. Capilla, and C. Carrillo, "An Overview on Quality Evaluation of Self-Adaptive Systems," 2016.
- [7] J. D. Paraiba and L. E. G. Martins, "PERSA: a requirements specification process for self-adaptive systems based on fuzzy logic and NFR-framework," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 25, no. 01, pp. 145-178, 2017.
- [8] J. A. McCall, P. K. Richards, and G. F. Walters, "Factors in software quality. volume i. concepts and definitions of software quality," DTIC Document 1977.

۶- نتیجه‌گیری و پیشنهادات

در این مقاله هدف ارائه یک رویکرد برای ارزیابی کیفیت سیستم‌های خودانطباق است. سیستم‌های خودانطباق به سیستم‌هایی اطلاق می‌گردد که رفتار خود را بر طبق اطلاعات محیطی و یا زمینه خود تغییر می‌دهند. این دسته از سیستم‌ها، عموماً بدون دخالت انسان و یا با حداقل دخالت انسان کار می‌کنند. مقاله حاضر به بررسی شاخص‌های کیفی و در واقع کیفیت نرم‌افزار خودانطباق در یک سیستم کنترل ترافیک به نام InSync می‌پردازد. به همین دلیل، ابتدا، شاخص‌های اصلی جهت ارزیابی سیستم‌های خودانطباق از مقالات و منابع مختلف در دسترس برای این سیستم استخراج شده و سپس شیوه اندازه‌گیری این شاخص‌ها معرفی گردید.

- [29] E. Serral, P. Sernani, and F. Dalpiaz, "Personalized adaptation in pervasive systems via non-functional requirements," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 6, pp. 1729-1743, 2018.
- [30] R. MoeinFar and A. A. Barforoush, "Using models at run-time to measure quality of SAS in the large-scale software systems," in *2017 9th International Conference on Information and Knowledge Technology (IKT)*, 2017, pp. 99-103: IEEE.
- [31] R. Edwards and N. Bencomo, "DeSiRE: further understanding nuances of degrees of satisfaction of non-functional requirements trade-off," in *Proceedings of the 13th International Conference on Software Engineering for Adaptive and Self-Managing Systems*, 2018, pp. 12-18.
- [32] C. Raibulet, "Hints on quality evaluation of self-systems," in *2014 IEEE Eighth International Conference on Self-Adaptive and Self-Organizing Systems*, 2014, pp. 185-186: IEEE.
- [33] E. Kaddoum, C. Raibulet, J.-P. Georgé, G. Picard, and M.-P. Gleizes, "Criteria for the evaluation of self-* systems," in *Proceedings of the 2010 ICSE Workshop on Software Engineering for Adaptive and Self-Managing Systems*, 2010, pp. 29-38.
- [34] C. I. Bezerra, R. M. Andrade, J. M. Monteiro, and D. Cedraz, "Aggregating measures using fuzzy logic for evaluating feature models," in *Proceedings of the 12th International Workshop on Variability Modelling of Software-Intensive Systems*, 2018, pp. 35-42.
- [35] R. Wohlrab, J. Cámara, D. Garlan, and B. Schmerl, "Explaining quality attribute tradeoffs in automated planning for self-adaptive systems," *Journal of Systems and Software*, vol. 198, p. 111538, 2023.
- [36] S. Malik, M. A. Naqvi, and L. Moonen, "CHESS: A Framework for Evaluation of Self-adaptive Systems based on Chaos Engineering," *arXiv preprint arXiv:2303.07283*, 2023.
- [37] A. Parvizi-Mosaed, S. Moaven, J. Habibi, and A. Heydarnoori, "Towards a Tactic-Based Evaluation of Self-Adaptive Software Architecture Availability," in *SEKE*, 2014, pp. 168-173.
- [38] Q. Yang, J. Lü, J. Li, X. Ma, W. Song, and Y. Zou, "Toward a fuzzy control-based approach to design of self-adaptive software," in *Proceedings of the Second Asia-Pacific Symposium on Internetware*, 2010, p. 15: ACM.
- [39] Y. Wang, J. Corey, Y. Lao, K. Henrickson, and X. Xin, "Criteria for the Selection and Application of Advanced Traffic Signal Systems," 2013.
- [40] A. Stevanovic and P. M. Zlatkovic, "Comparative Evaluation of InSync and Time-of-Day Signal Timing Plans under Normal and Varied Traffic Conditions," ed: February, 2013.
- [41] S. Elkins and G. Niehus, "Insync adaptive traffic control system for the veterans memorial hwy corridor on long island, ny," 2012.
- [42] M. Selinger and L. Schmidt, "Adaptive traffic control systems in the united states: Updated summary and comparison," *HDR Engineering*, 2010.
- [43] M. KETABDARI, "Analysis of adaptive traffic control systems and design of a decision support system for better choice," 2013.
- [44] W. Min, Z. Jun, and Z. Wei, "The application of fuzzy comprehensive evaluation method in the software project risk assessment," in *Proceedings of the 2017 International Conference on Management Engineering, Software Engineering and Service Sciences*, 2017, pp. 76-79: ACM.
- [45] M. Salehie and L. Tahvildari, "Self-Adaptive Software: Landscape and Research Challenges," *ACM Transactions on Autonomous and Adaptive Systems*, 2009.
- [46] Y. Brun *et al.*, "Engineering Self-Adaptive Systems through Feedback Loops," *Software engineering for self-adaptive systems*, vol. 5525, pp. 48-70, 2009.
- [47] M. Ravindranathan and R. Leitch, "Heterogeneous intelligent control systems," *IEE Proceedings-Control Theory and Applications*, vol. 145, no. 6, pp. 551-558, 1998.
- [9] S. Valenti, A. Cucchiarelli, and M. Panti, "Computer based assessment systems evaluation via the ISO9126 quality model," *Journal of Information Technology Education*, vol. 1, no. 3, pp. 157-175, 2002.
- [10] B. W. Boehm, J. R. Brown, and M. Lipow, "Quantitative evaluation of software quality," in *Proceedings of the 2nd international conference on Software engineering*, 1976, pp. 592-605: IEEE Computer Society Press.
- [11] P. Berander *et al.*, "Software quality attributes and trade-offs," *Blekinge Institute of Technology*, 2005.
- [12] Z. Abbasi, S. Fatahi, and M. J. Ershadi, "An approach to prioritize quality dimensions of based on cloud computing using Multiple Criteria Decision Making method," *Journal of Information and Communication Technology*, vol. 53, no. 14, pp. 224-236, 1401.
- [13] A. Amini and A. Alinezhad, "A Combination Method of DEA, DEMATEL and ANP for Evaluation of ERP Systems," *Journal of Information and Communication Technology*, vol. 45, no. 12, pp. 21-42, 1399.
- [14] A. O. de Sousa, C. I. Bezerra, R. M. Andrade, and J. M. Filho, "Quality evaluation of self-adaptive systems: Challenges and opportunities," in *Proceedings of the XXXIII Brazilian Symposium on Software Engineering*, 2019, pp. 213-218.
- [15] C.-W. Chang, C.-R. Wu, and H.-L. Lin, "Integrating fuzzy theory and hierarchy concepts to evaluate software quality," *Software Quality Journal*, vol. 16, no. 2, pp. 263-276, 2008.
- [16] P. HOLECEK and J. TALAŠOVÁ, "FuzzME: a new software for multiple-criteria fuzzy evaluation," *Acta Universitatis Matthiae Belii ser. Mathematica*, vol. 16, pp. 35-51, 2010.
- [17] J. S. Challa, A. Paul, Y. Dada, V. Nerella, P. R. Srivastava, and A. P. Singh, "Integrated Software Quality Evaluation: A Fuzzy Multi-Criteria Approach," *JIPS*, vol. 7, no. 3, pp. 473-518, 2011.
- [18] E. Letier, D. Stefan, and E. T. Barr, "Uncertainty, risk, and information value in software requirements and architecture," in *Proceedings of the 36th International Conference on Software Engineering*, 2014, pp. 883-894.
- [19] F. Haryana, "Software Quality Evaluation using Fuzzy Multi Criteria Decision Method," 2015.
- [20] A. Mansoor, D. Streitferdt, and F.-F. Füßl, "Fuzzy Based Evaluation of Software Quality Using Quality Models and Goal Models."
- [21] M. Kara, O. Lamouchi, and A. Ramdane-Cherif, "Ontology software quality model for fuzzy logic evaluation approach," *Procedia Computer Science*, vol. 83, pp. 637-641, 2016.
- [22] A. S. Abdygalievich, A. Barlybayev, and K. B. Amanzholovich, "Quality evaluation fuzzy method of automated control systems on the LMS example," *IEEE Access*, vol. 7, pp. 138000-138010, 2019.
- [23] U. Dayanandan and V. Kalimuthu, "A fuzzy analytical hierarchy process (FAHP) based software quality assessment model: maintainability analysis," *International Journal of Intelligent Engineering and Systems*, vol. 11, no. 4, pp. 88-96, 2018.
- [24] D. Manikavelan and R. Ponnusamy, "Software quality analysis based on cost and error using fuzzy combined COCOMO model," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-11, 2020.
- [25] N. X. Thao and S.-Y. Chou, "Novel similarity measures, entropy of intuitionistic fuzzy sets and their application in software quality evaluation," *Soft Computing*, pp. 1-12, 2022.
- [26] A. Barzegar, "Measuring software quality Product based on Fuzzy Inference System techniques in ISO standard," 2021.
- [27] V. Singh, V. Kumar, and V. Singh, "A hybrid novel fuzzy AHP-TOPSIS technique for selecting parameter-influencing testing in software development," *Decision Analytics Journal*, vol. 6, p. 100159, 2023.
- [28] L. E. Sanchez, J. A. Diaz-Pace, A. Zunino, S. Moisan, and J.-P. Rigault, "An approach based on feature models and quality criteria for adapting component-based systems," *Journal of Software Engineering Research and Development*, vol. 3, pp. 1-30, 2015.

- [53] Q. Liang and J. M. Mendel, "Interval type-2 fuzzy logic systems," in *Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on*, 2000, vol. 1, pp. 328-333: IEEE.
- [54] A. Farahani, E. Nazemi, G. Cabri, and A. Rafizadeh, "An evaluation method for Self-Adaptive systems," in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, 2016, pp. 002814-002820: IEEE.
- [55] J. A. McCann and M. C. Huebscher, "Evaluation issues in autonomic computing," in *International Conference on Grid and Cooperative Computing*, 2004, pp. 597-608: Springer.
- [56] Danny Weyns, M. Usman Iftikhar, Didac Gil de la Iglesia, and T. Ahmad, "A Survey of Formal Methods in Self-Adaptive Systems," 2012.
- [57] N. M. Villegas, H. A. Müller, G. Tamura, L. Duchien, and R. Casallas, "A Framework for Evaluating Quality-Driven Self-Adaptive Software Systems," 2011.
- [58] S. Weibelzahl and G. Weber, "Advantages, opportunities and limits of empirical evaluations: Evaluating adaptive systems," *KI*, vol. 16, no. 3, pp. 17-20, 2002.
- [48] M. Naqvi, "Claims and supporting evidence for self-adaptive systems—A literature review," ed, 2012.
- [49] R. Almeida and M. Vieira, "Changeloads for resilience benchmarking of self-adaptive systems: a risk-based approach," in *Dependable Computing Conference (EDCC), 2012 Ninth European*, 2012, pp. 173-184: IEEE.
- [50] S. M. Abuelenin, "Decomposed interval Type-2 fuzzy systems with application to inverted pendulum," in *Engineering and Technology (ICET), 2014 International Conference on*, 2014, pp. 1-5: IEEE.
- [51] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE transactions on Fuzzy Systems*, vol. 7, no. 6, pp. 643-658, 1999.

[۵۲] ح. م. فراهانی، ج. عسگری و م. ذکری، "مروری بر منطق فازی نوع-۲: از پیدایش تا کاربرد"، محاسبات نرم، دوره ۲، شماره ۱، شماره پیاپی ۳۳، صفحه ۲۲-۴۳، ۱۳۹۲.

پیوست‌ها

جدول الف-۱. رتبه شاخص‌های مؤثر در ارزیابی سیستم کنترل ترافیک

VH	H	M	L	VL	No	نام شاخص	VH	H	M	L	VL	No	نام شاخص
*						دقت	*						پایداری
*						تخطی کم	*						زمان تنظیم کوتاه
*						سازگاری	*						پایان دهی
*						حساسیت	*						مقیاس پذیری
*						پرهیز از شکست	*						دانه بندی
*						استحکام	*						پیشگویی
*						امنیت	*						سرعت پاسخ خود مختار
*						دقت تصمیم گیری	*						هزینه
	*					سطح خود مختاری	*						درجه خود مختاری
	*					خود بهبودی	*						خود پیکربندی
*						خود آگاهی	*						خود بهینگی
*						آگاهی به زمینه	*						خود حفاظتی
					*	ثبات	*						باز بودن
*						رضایت						*	قابلیت حمل
*						قابلیت استفاده	*						قابلیت نگهداری
	*					قابلیت اعتماد	*						قابلیت عملیاتی (عملکرد)

جدول الف-۲. رتبه متریک‌های مؤثر در ارزیابی سیستم

VH	H	M	L	VL	No	نام متریک	VH	H	M	L	VL	No	نام متریک
*						وقف پذیری ساختار داده	*						حسگر فعال
*						وقف پذیری محیطی سخت افزار	*						زمان پاسخ (اجرا)
*						وقف پذیری محیط اجرایی	*						زمان انتظار
*						وقف پذیری محیط نرم افزاری	*						بهره‌وری از منابع
*						درگاه‌های کاربر پسند	*						رضایت
				*		نصب آسان	*						سادگی
				*		آسانی تنظیمات مجدد	*						قابلیت گسترش
		*				همزیستی قابل دسترس	*						یکپارچگی
*						زمان تصمیم			*				مناسب بودن
			*			پیمچیدگی رفتار	*						قابلیت همکاری درونی
		*				دانه‌بندی ریز/ درشت	*						تاخیر

VH	H	M	L	VL	No	نام متریک	VH	H	M	L	VL	No	نام متریک
	*					ثبات	*						توان عملیاتی
	*					زمان شکست	*						ظرفیت
		*				زمان بازیابی	*						فراگیری عملیات
	*					سودمندی	*						رفع شکست و نقص
	*					قابلیت یادگیری	*						تشخیص شکست و نقص
	*					قابلیت یادسپاری	*						پرهیز از شکست و نقص
						پیچیدگی تعاملات							دانش
						محرمانگی							قدرت اتصال
						ساختار دهی مجدد							عدم انکار
						قابلیت بازیابی و شروع دوباره							سرعت تصمیم‌گیری
						قابلیت تغییر							قابلیت تحلیل
	*					قابلیت درک	*						قابلیت دسترسی
													قابلیت آزمایش

جدول الف-۳. پرسشنامه رتبه متریک‌های مؤثر در ارزیابی سیستم

VH	H	M	L	VL	No	سوال	نام متریک
						چه میزان معیار حسگرهای فعال (برای مشخص کردن حساسیت سیستم) برای سیستم فوق دارای اهمیت است؟	حسگر فعال
						چه میزان زمان پاسخ در سیستم مورد نظر دارای اهمیت است؟	زمان پاسخ(اجرا)
						چه میزان زمان انتظار کاربر برای دریافت جواب از سیستم دارای اهمیت است؟	زمان انتظار
						چه میزان امکان استفاده اشتراکی از منابع و مؤلفه‌های سیستم برای سیستم دارای اهمیت است؟	بهره‌وری در منابع
						چه میزان سادگی مؤلفه‌های مورد استفاده در سیستم دارای اهمیت است؟	سادگی
						چه میزان امکان بسط در طراحی سیستم دارای اهمیت است؟	قابلیت گسترش
						چه میزان ورودی غیر مجاز و تغییر غیر مجاز در داده‌ها دارای اهمیت است؟	یکپارچگی
						چه میزان مناسب بودن سیستم برای اجرای اهداف دارای اهمیت است؟	مناسب بودن
						چه میزان توانایی همکاری در بین مؤلفه‌های درونی سیستم دارای اهمیت است؟	قابلیت همکاری درونی
						چه میزان تاخیر در سیستم دارای اهمیت است؟	تاخیر
						چه میزان توان عملیاتی و موفق شدن وظایف در یک دوره زمانی معین توسط سیستم دارای اهمیت است؟	توان عملیاتی
						چه میزان، ظرفیت و میزان دریافت وظیفه جهت اجرا توسط سیستم دارای اهمیت است؟	ظرفیت
						چه میزان رفع خودکار نقص و شکست در سیستم دارای اهمیت است؟	رفع شکست و نقص
						چه میزان تشخیص نقص و شکست در سیستم دارای اهمیت است؟	تشخیص شکست و نقص
						چه میزان پرهیز از نقص و شکست در سیستم دارای اهمیت است؟	پرهیز از شکست و نقص
						چه میزان دسترسی بودن سیستم دارای اهمیت است؟	قابلیت دسترسی
						چه میزان مدت زمان از کار افتادن سیستم به علت شکست دارای اهمیت است؟	زمان شکست
						چه میزان مدت زمان بازیابی سیستم دارای اهمیت است؟	زمان بازیابی
						چه میزان برآورده شدن اهداف کاربر توسط سیستم دارای اهمیت است؟	سودمندی
						چه میزان سادگی یادگیری سیستم توسط کاربر دارای اهمیت است؟	قابلیت یادگیری
						چه میزان به یادسپاری توابع و فرامین سیستم توسط کاربر دارای اهمیت است؟	قابلیت یادسپاری
						چه میزان قابل درک بودن سیستم توسط کاربر نهایی دارای اهمیت است؟	قابلیت درک
						چه میزان سازگاری داده‌های موجود با محیط جدید و سیستم جدید دارای اهمیت است؟	وفق‌پذیری ساختار داده
						چه میزان وفق‌پذیری نرم‌افزار با سخت‌افزار (جدید) دارای اهمیت است؟	وفق‌پذیری محیطی سخت‌افزار
						چه میزان وفق‌پذیری نرم‌افزار با محیط اجرایی مختلف دارای اهمیت است؟	وفق‌پذیری محیط اجرایی
						چه میزان وفق‌پذیری نرم‌افزار با محیط دارای اهمیت است؟	وفق‌پذیری محیط نرم‌افزاری
						چه میزان کاربرپسند بودن درگاه‌های سیستم دارای اهمیت است؟	درگاه‌های کاربر پسند
						چه میزان آسان بودن نصب سیستم توسط کاربر نهایی دارای اهمیت است؟	سهولت آسان
						چه میزان سادگی تنظیمات مجدد سیستم توسط کاربر نهایی دارای اهمیت است؟	سهولت تنظیمات مجدد
						چه میزان همزیستی سیستم با دیگر نرم‌افزارها، دارای اهمیت است؟	همزیستی

VH	H	M	L	VL	No	سوال	نام متریک
						چه میزان مدت زمان لازم برای تصمیم‌گیری سیستم دارای اهمیت است؟	زمان تصمیم
						چه میزان پیچیده بودن رفتار سیستم دارای اهمیت است؟	پیچیدگی رفتار
						چه میزان ریز یا درشت بودن مؤلفه‌های سیستم دارای اهمیت است؟	دانه بندی ریز/ درشت
						چه میزان ثبات نتایج سیستم دارای اهمیت است؟	ثبات
						چه میزان رضایت کاربر نهایی از سیستم اهمیت دارد؟	رضایت
						چه میزان امکان جایگزینی و مشابه بودن توابع در سیستم نرم‌افزاری دارای اهمیت است؟	فراگیری عملیات
						چه میزان دانش در زمان تصمیم‌گیری در سیستم اهمیت دارد؟	دانش
						چه میزان قدرت اتصال و در نتیجه میزان تاخیر در بین مؤلفه‌ها در سیستم دارای اهمیت است؟	قدرت اتصال
						چه میزان پیچیدگی تعامل، و در نتیجه رفتار یک مؤلفه با سایر مؤلفه‌ها در سیستم دارای اهمیت است؟	پیچیدگی تعاملات
						چه میزان محرمانگی داده‌ها و عدم امکان دسترسی غیر مجاز به داده‌ها دارای اهمیت است؟	محرمانگی
						چه میزان ساختاردهی مجدد سیستم دارای اهمیت است؟	ساختاردهی مجدد
						چه میزان ردیابی فعالیت‌ها دارای اهمیت است؟	عدم انکار
						چه میزان سرعت تصمیم‌گیری در سیستم دارای اهمیت است؟	سرعت تصمیم‌گیری
						چه میزان قابلیت بازسازی و شروع دوباره در سیستم دارای اهمیت است؟	قابلیت بازسازی و شروع دوباره
						چه میزان تشخیص شکست یا شناسایی آیت‌هایی که باید اصلاح شوند، در سیستم دارای اهمیت است؟	قابلیت تحلیل
						چه میزان انجام و پیاده‌سازی تغییرات در سیستم دارای اهمیت است؟	قابلیت تغییر
						چه میزان آزمایش در سیستم بعد از انجام تغییرات دارای اهمیت است؟	قابلیت آزمایش

جدول الف-۴. پرسشنامه رتبه شاخص‌های مؤثر در ارزیابی سیستم

VH	H	M	L	VL	No	نام شاخص	VH	H	M	L	VL	No	نام شاخص
						دقت							پایداری
						تخطی کم							زمان تنظیم کوتاه
						سازگاری							پایان دهی
						حساسیت							مقیاس پذیری
						پرهیز از شکست							دانه بندی
						استحکام							پیشگویی
						امنیت							سرعت پاسخ خود مختار
						دقت تصمیم‌گیری							هزینه
						سطح خود مختاری							درجه خود مختاری
						خود بهبودی							خود پیکربندی
						خود آگاهی							خود بهینگی
						آگاهی به زمینه							خود حفاظتی
						ثبات							باز بودن
						رضایت							قابلیت حمل
						قابلیت استفاده							قابلیت نگهداری
						قابلیت اعتماد							قابلیت عملیاتی (عملکرد)

جدول الف-۵. پرسشنامه مقادیر فازی متریک‌های مؤثر در ارزیابی سیستم

VH	H	M	L	VL	No	سوال	نام متریک
						چه میزان معیار حسگرهای فعال (برای مشخص کردن حساسیت سیستم) برای سیستم مورد نظر برآورده شده است؟	حسگر فعال
						چه میزان زمان پاسخ در سیستم مورد نظر برآورده شده است؟	زمان پاسخ (اجرا)
						چه میزان زمان انتظار کاربر برای دریافت جواب از سیستم مطلوب است؟	زمان انتظار
						چه میزان امکان استفاده اشتراکی از منابع و مؤلفه‌های سیستم برای سیستم برآورده شده است؟	بهره‌وری از منابع
						چه میزان سادگی مؤلفه‌های مورد استفاده در سیستم برآورده شده است؟	سادگی
						چه میزان امکان بسط در طراحی سیستم وجود دارد؟	قابلیت گسترش
						چه میزان عدم امکان ورود غیر مجاز و تغییر غیر مجاز در داده‌ها برآورده شده است؟	یکپارچگی

VH	H	M	L	VL	No	سوال	نام متریک
						چه میزان سیستم برای اجرای اهداف خود مناسب است؟	مناسب بودن
						چه میزان توانایی همکاری در بین مولفه‌های درونی سیستم برآورده شده است؟	قابلیت همکاری درونی
						چه میزان تاخیر در سیستم قابل تحمل است؟	تاخیر
						چه میزان توان عملیاتی و موفق شدن وظایف در یک دوره زمانی معین توسط سیستم مطلوب است؟	توان عملیاتی
						چه میزان، ظرفیت و میزان دریافت وظیفه جهت اجرا توسط سیستم مطلوب است؟	ظرفیت
						چه میزان رفع خودکار نقص و شکست در سیستم مطلوب است؟	رفع شکست و نقص
						چه میزان تشخیص نقص و شکست در سیستم برآورده شده است؟	تشخیص شکست و نقص
						چه میزان پرهیز از نقص و شکست در سیستم برآورده شده است؟	پرهیز از شکست و نقص
						چه میزان سیستم دسترس است؟	قابلیت دسترسی
						چه میزان مدت زمان از کار افتادن سیستم به علت شکست مطلوب است؟	زمان شکست
						چه میزان مدت زمان بازیابی سیستم مطلوب است؟	زمان بازیابی
						چه میزان اهداف کاربر توسط سیستم برآورده شده است؟	سودمندی
						چه میزان یادگیری سیستم توسط کاربر ساده است؟	قابلیت یادگیری
						چه میزان به یادسپاری توابع و فرامین سیستم توسط کاربر ساده است؟	قابلیت یادسپاری
						چه میزان سیستم توسط کاربر نهایی قابل درک بودن است؟	قابلیت درک
						چه میزان داده‌های موجود با محیط جدید و سیستم جدید سازگار است؟	وقف پذیری ساختار داده
						چه میزان نرم‌افزار با سخت‌افزار (جدید) وقف پذیر است؟	وقف پذیری محیطی سخت‌افزار
						چه میزان نرم‌افزار با محیط اجرایی مختلف وقف پذیر است؟	وقف پذیری محیط اجرایی
						چه میزان نرم‌افزار با محیط وقف پذیر است؟	وقف پذیری محیط نرم‌افزاری
						چه میزان درگاه‌های سیستم کاربرپسند است؟	درگاه‌های کاربر پسند
						چه میزان نصب سیستم توسط کاربر نهایی آسان است؟	نصب آسان
						چه میزان تنظیمات مجدد سیستم توسط کاربر نهایی ساده است؟	سهولت تنظیمات مجدد
						چه میزان سیستم با دیگر نرم‌افزارها، همزیستی دارد؟	همزیستی
						چه میزان مدت زمان لازم برای تصمیم‌گیری سیستم مطلوب است؟	زمان تصمیم
						چه میزان رفتار سیستم پیچیده است؟	پیچیدگی رفتار
						چه میزان مؤلفه‌های سیستم ریز است؟	دانه بندی ریز/ درشت
						چه میزان نتایج سیستم دارای ثبات است؟	ثبات
						چه میزان کاربر نهایی از سیستم رضایت دارد؟	رضایت
						چه میزان امکان جایگزینی و مشابه بودن توابع در سیستم نرم‌افزاری وجود دارد؟	فراگیری عملیات
						چه میزان دانش مناسب در زمان تصمیم‌گیری در سیستم وجود دارد؟	دانش
						چه میزان قدرت اتصال و در نتیجه میزان تاخیر در بین مولفه‌ها در سیستم مطلوب است؟	قدرت اتصال
						چه میزان تعاملات و در نتیجه رفتار یک مولفه با سایر مولفه‌ها در سیستم پیچیده است؟	پیچیدگی تعاملات
						چه میزان محرمانگی داده‌ها و عدم امکان دسترسی غیر مجاز به داده‌ها برآورده شده است؟	محرمانگی
						چه میزان سیستم دارای قابلیت ساختاردهی مجدد است؟	ساختاردهی مجدد
						چه میزان امکان ردیابی فعالیت‌ها وجود دارد؟	عدم انکار
						چه میزان سرعت تصمیم‌گیری در سیستم مطلوب است؟	سرعت تصمیم‌گیری
						چه میزان سیستم قابلیت بازسازی و شروع دوباره دارد؟	قابلیت بازسازی و شروع دوباره
						چه میزان تشخیص شکست یا شناسایی ایت‌هایی که باید اصلاح شوند، در سیستم برآورده شده است؟	قابلیت تحلیل
						چه میزان انجام و پیاده‌سازی تغییرات در سیستم برآورده شده است؟	قابلیت تغییر
						چه میزان امکان آزمایش در سیستم بعد از انجام تغییرات برآورده شده است؟	قابلیت آزمایش

Trust Management Based on User Feedback in Cloud Computing Environment by Using Cuckoo Optimization Algorithm

Muqtada Soleimani Mobarake¹, Golnaz Aghaee Ghazvini^{2*}, Babak Nikmard³

¹Masters student, Computer Engineering, Department of Computer Engineering, Islamic Azad University, Dolatabad Branch, Isfahan, Iran

²Assistant Professor, Department of Computer Engineering, Islamic Azad University, Dolatabad Branch, Isfahan, Iran

³Assistant Professor, Department of Computer Engineering, Islamic Azad University, Dolatabad Branch, Isfahan, Iran

Received: 28 September 2023, Revised: 20 November 2023, Accepted: 01 May 2024

Paper type: Research

Abstract

Cloud computing provides computational services such as servers, memory, storage space, databases, networks, software, analytics, and information as virtualized resources through the internet to offer faster innovation, flexible resources, and cost savings at scale. Although cloud computing service providers are innovatively expanding their services, trust is one of the major obstacles to the progress of this matter. Trust is the biggest issue in cloud computing since trust is an effective guarantor during interactions between the users and the providers. Trust is one of the most fundamental methods for increasing confidence in resources provided in the cloud environment and is important in cloud business environments. With the increasing number of cloud services providers in the cloud computing environment and the number of users, the selection of provider has become a major challenge. The Coa algorithm has a higher convergence speed, at least by 5.9%, compared to the studied algorithms. In this research, an optimization approach based on a metaheuristic process using the COA algorithm combined with the K-means clustering algorithm is proposed to solve the optimization problem of selecting the best provider in the trust management third-party component layer based on parameters. In this method, while reducing trust evaluation time, the accuracy of user access to the most trusted provider based on user priorities has increased compared to previous methods. This can increase user confidence and improve the service quality of providers.

Keywords: Cloud Computing, Trust Management, User Feedback, Cuckoo Optimization Algorithm, K-Means Algorithm

* Corresponding Author's email: g.ghaee@iauda.ac.ir

مدیریت اعتماد مبتنی بر بازخورد کاربران در محیط ابری با استفاده از الگوریتم بهینه فاخته

مقتدا سلیمانی مبارکه^۱، گلناز آقایی قزوینی^{۲*}، بابک نیکمرد^۳

^۱دانشجوی کارشناسی ارشد، گروه کامپیوتر، واحد دولت آباد، دانشگاه آزاد اسلامی، اصفهان، ایران

^۲استادیار، گروه کامپیوتر، واحد دولت آباد، دانشگاه آزاد اسلامی، اصفهان، ایران

^۳استادیار، گروه کامپیوتر، واحد دولت آباد، دانشگاه آزاد اسلامی، اصفهان، ایران

تاریخ دریافت: ۱۴۰۲/۰۷/۰۶ تاریخ بازبینی: ۱۴۰۲/۰۸/۲۹ تاریخ پذیرش: ۱۴۰۳/۰۲/۱۲

نوع مقاله: پژوهشی

چکیده

مدیریت اعتماد مبتنی بر بازخورد کاربران در محیط ابری از اهمیت زیادی برخوردار است. در محیط ابری انتخاب تأمین‌کننده برای کاربران ابر، چالش برانگیز است. این موضوع که آیا انتخاب تأمین‌کننده بر مبنای اولویت‌های کاربر و پارامترهای ثبت شده تا چه میزان دقیق است به عوامل زیادی بستگی دارد. در مطالعات پیشین چارچوب‌های زیادی در خصوص نحوه محاسبه اعتمادهای عینی و درونی ارائه شده است. در روش‌های موجود با استفاده از الگوریتم‌های جستجوی فاخته، ژنتیک و مگس میوه بهره برده شده است. قالب تحقیقات انجام شده به رتبه‌بندی، محاسبه پارامترها و یا سرعت محاسبه و میزان دقت در ارزیابی پارامترها پرداخته شده است که معمولاً یا در بهینه محلی گیر کرده و یا زمان پاسخ بسیار کند بوده است. در این روش کاهش زمان ارزیابی اعتماد، نسبت به الگوریتم‌های قبلی همچون ژنتیک به دلیل این که پارامترهای کمتری برای تنظیم دارد مشهود است. با تغییر در شعاع تخمگذاری و افزایش بررسی در فضای بیشتری از مسئله، الگوریتم بهینه فاخته نسبت به الگوریتم مرجع، سرعت همگرایی بیشتر، حداقل به میزان ۵٫۹ درصد را دارد. در خصوص میزان دقت دسترسی کاربر به قابل اعتمادترین ارائه‌دهنده نیز با تغییر در جمعیت و پارامترهای همچون تعداد فراهم‌کنندگان، کاربران و تکرار الگوریتم همچنان نتیجه بهتری حاصل شده است. در مجموع، نتایج حاصل شده نشان می‌دهد، مسئله با استفاده از الگوریتم بهینه فاخته در زمان بسیار کمتر نسبت به سایر الگوریتم‌ها به نقطه بهینه همگرا می‌شود و نتیجه‌ای دقیق‌تر بدست می‌آید.

کلیدواژگان: رایانش ابری، مدیریت اعتماد، بازخورد کاربران، انتخاب بهترین تأمین‌کننده، الگوریتم بهینه فاخته، الگوریتم خوشه‌بندی K-Means

* رایانامه نویسنده مسؤول: g.agmaee@iauda.ac.ir

۱- مقدمه

بازخورد کاربران و کیفیت سرویس در فضای رایانش ابری، در مدت زمان کمتر و ارائه‌دهنده‌ای با کارایی بالاتر دست یافته شود. بصورت کلی کاربران ابری خواهند توانست در سریع‌ترین زمان ممکن با در نظر گرفتن نیازمندی‌ها و اولویت‌های خود سرویس ابری را انتخاب نمایند.

ادامه مقاله به شرح زیر سازماندهی شده است: بخش دوم مروری بر مطالعات انجام شده و در بخش سوم پیشینه نیز توضیح داده می‌شود. در بخش چهارم چارچوب پیشنهادی مورد بحث قرار گرفته است. رویکرد ارزیابی اعتماد با استفاده از الگوریتم بهینه فاخته در بخش پنجم ارائه شده است. در بخش ششم نتایج شبیه‌سازی و پیاده‌سازی بیان شده است. و در بخش انتهایی با ارائه نتیجه و پیشنهادات آتی مقاله به پایان می‌رسد.

۲- مروری بر مطالعات انجام شده

در پژوهش‌های انجام شده اعتماد بر پایه بازخورد، دامنه و توافق مورد محاسبه و ارزیابی قرار گرفته‌اند. این ارزیابی بصورت عینی و یا درونی انجام و هر کدام دارای نحوه عملکرد منحصر بفرد خود هستند. نحوه محاسبه بر مبنای تحلیل سلسله مراتبی، اعتماد چند نوبتی، سابقه تعامل، اعتماد درون و برون دامنه انجام شده است. بعنوان نمونه در مطالعات انجام شده با تبادل فایل‌ها میان نودها بر مبنای سابقه تعاملات و گزارشات میان نودها و حفظ جامعیت دیتا فرایند ارزیابی صورت می‌گیرد و یا با استفاده از پارامترهای قابل اندازه‌گیری مثل اطمینان، هزینه، امنیت و بهره بردن از مکانیزم رتبه‌بندی در ارزیابی با تحلیل سلسله مراتبی راهکارهایی پیشنهاد شده، اما ضعف این موارد در عدم کاربرد برای بعضی پارامترهایی است که قابل اندازه‌گیری نیستند. در راهکارهایی که فرایند اعتماد در چند مرحله صورت می‌گیرد وابستگی به تامین‌کنندگان سرویس ابری بسیار زیاد است. در نمونه‌های جامع‌تر نیز همچون ارزیابی بر مبنای دسترسی پذیری، قابلیت اطمینان و یکپارچگی داده و کارایی زمان پاسخ نیز با وجود اینکه بهترین منبع انتخاب می‌شود اما زمان‌بندی و نوبت درخواست‌های کاربران نادیده گرفته می‌شود.

در بعضی از پژوهش‌ها محاسبه مقدار اعتماد بین گره‌ها بر اساس تبدلات، جدول اعتماد، جدول اعتماد درون دامنه و جدول اعتماد برون دامنه بود با وجود این که برای دامنه‌ها از سیاست‌های امنیتی مختلفی استفاده می‌شود و یا از تاریخچه تراکنش و سابقه بهره می‌برد اما ممکن است گره مربوطه به چندین دامنه تعلق داشته باشد و یا موضوع تراکنش که یک عامل ضروری در تشخیص مقدار اعتماد است در نظر گرفته نشود. در جدیدترین مطالعات انجام شده در این خصوص مدیریت اعتماد دوطرفه در محیط محاسبات

رایانش ابری ارائه سرویس‌های محاسباتی مانند سرورها، حافظه، فضای ذخیره‌سازی، پایگاه‌های اطلاعاتی، شبکه‌ها، نرم‌افزارها، تجزیه و تحلیل‌ها و اطلاعات بصورت منبع‌های مجازی شده توسط تامین‌کنندگان خدمت از طریق اینترنت برای ارائه نوآوری سریع‌تر و منابع انعطاف پذیر و صرفه جویی در مقیاس است. با وجود اینکه تامین‌کنندگان خدمات رایانش ابری به گسترش روز افزون خدمات خود بصورت مبتکرانه می‌پردازند اما اعتماد یکی از موانع مهم بر سر راه پیشرفت این امر می‌باشد. اعتماد بزرگترین مسئله در رایانش ابری است، چرا که تضمین وجود اعتماد داده‌ها بطور موثر در طول تعاملات بین کاربر و ابر می‌باشد که به سبب آن یک ارتباط قابل اعتماد متقابل بین کاربر و ارائه‌دهنده خدمات ابر با روش‌های کنترل دسترسی در محیط ابری بوجود می‌آید. اگرچه برای انتخاب ارائه‌دهنده خدمات ابری طیف وسیعی از معیارها وجود دارد با این حال انتخاب نهایی در اختیار کاربر است. استفاده از چارچوبی مناسب در مدیریت اعتمادهای عینی و درونی با بهره بردن از معیارها و ارزیابی آن می‌تواند تاثیر بسزایی در تعاملات بین تامین‌کنندگان و کاربران ابری بگذارد. در این مقاله با توجه به اهمیت موضوع، جهت بهبود و افزایش اعتماد و انتخاب بهترین تامین‌کننده با در نظر گرفتن پارامترهای عینی و درونی در محاسبات یک روش مبتنی بر الگوریتم بهینه فاخته ارائه شده است تا علاوه بر سرعت در مدیریت داده‌ها و دریافت پاسخ در کمترین زمان ممکن میزان دقت در معرفی بهترین تامین‌کننده نیز نسبت به الگوریتم‌های بکارگرفته شده در پژوهش‌های قبلی بر مبنای اولویت‌های کاربران افزایش یافته است. در این مقاله یک الگوریتم بهینه‌سازی تکاملی جدید معرفی می‌شود که از سبک زندگی یک خانواده پرندگان به نام فاخته الهام گرفته شده است. تخمگذاری و پرورش خاص فاخته‌ها اساس این الگوریتم بهینه‌سازی جدید است. تلاش برای زنده ماندن در میان فاخته‌ها اساس الگوریتم بهینه‌سازی فاخته را تشکیل می‌دهد. با توجه به اهمیت موضوع و جهت بهبود و افزایش اعتماد در محاسبات ابری در مدیریت داده‌ها در بستر رایانش ابری یک روش مبتنی بر الگوریتم تکاملی فاخته ارائه شده است. در روش پیشنهادی، بازخورد کاربران و کیفیت سرویس ارائه‌دهندگان در سیستم ثبت می‌شود و این اطلاعات به عنوان پارامترهای ورودی الگوریتم فاخته استفاده می‌شود. تلاش می‌شود کمک شایانی به کاهش اندازه مسئله مورد نظر، همچنین کاهش فضای جستجو برای کاربران و کاهش پراکندگی بازخورد کاربران انجام شود. همچنین با بهره‌مندی از این الگوریتم، باید بتوان به بهترین و قابل‌اعتمادترین ارائه‌دهنده بر مبنای

دست و پنجه نرم کنیم [۲]. مفهوم اعتماد موضوع پیچیده‌ای است در واقع اعتقاد راسخ به ویژگی‌هایی مانند قابلیت اطمینان، صداقت و شایستگی نهاد مورد اعتماد مربوط می‌شود [۳]. برآورد اعتماد مبتنی بر رفتاری است که شامل تاریخچه ارتباط مستقیم بین کاربر و ارائه‌دهنده خدمات و همچنین بر اساس توصیه‌های دوستان و شخص ثالث است [۴]. ارزش اعتماد منابع به کاربران ابر کمک می‌کند تا خدمات یک ارائه‌دهنده ابر را برای پردازش و ذخیره اطلاعات ضروری خود انتخاب کنند. همچنین، ارائه‌دهندگان خدمات می‌توانند به کاربران بر اساس ارزش اعتماد دسترسی داشته باشند تا منابع ابری را از کاربران مخرب ایمن کنند. همگی این جنبه‌های ایجاد اعتماد و روش‌های محاسبه آن‌ها معمولاً با نام «مدل‌های اعتماد» در مراجع علمی شناخته می‌شود. مدل‌های اعتماد در رایانش ابری بسیار گسترده هستند به طوری که هر مدل امکانات مختلفی را پشتیبانی می‌کند و سرویس‌های ابری را بر اساس پارامترها و نیازمندی‌های مختلفی ارزیابی می‌کند. در رایانش ابری، معیارها و کیفیت سطح سرویس از یک مصرف‌کننده سرویس ابر تا مصرف‌کننده دیگر ابر متفاوت است. چرا که، یکی ممکن است یکپارچگی داده‌های ارائه شده توسط تامین‌کننده سرویس ابر را ترجیح دهد در حالی که مصرف‌کننده دیگر سرویس ابری را انتخاب کند که بیشترین پهنای باند موجود و کارایی زمان پاسخ را داشته باشد. بنابراین، بسیار اهمیت دارد که مدل اعتماد مناسبی انتخاب شود که بیشترین پوشش را از تمامی پارامترها داشته باشد [۵]. صالحی و همکاران، مقاله‌ای تحت عنوان «انتخاب سرویس ابری مبنای بر پایداری در اعتلاف بازخورد کاربران»، انجام دادند [۶]. در این مقاله، یک مدل محاسباتی مبتنی بر اعتبار بازخورد کاربران، به منظور محاسبه اعتماد سرویس‌ها ارائه می‌شود. این مدل علاوه بر تجربیات کاربران از اعتبار کاربران به همراه میزان پایداری بازخوردهای درست آنها استفاده می‌نماید. در انتها، میزان کارایی مدل بررسی شده و ارزیابی‌ها نشان می‌دهد که مدل ارائه شده با وجود کاربران بدخواه، توانسته است اعتماد سیستم را به درستی محاسبه نماید. مجاور و باجانتی، مقاله‌ای تحت عنوان «برآورد اعتماد مبتنی بر رفتار و بازخورد در فضای ابری»، انجام دادند [۷]. این مقاله روشی را برای ارزیابی قابلیت اعتماد ارائه‌دهنده خدمات ابری بر اساس رفتار و بازخورد ارائه شده توسط کاربران ارائه می‌کند. آقای قزوینی و همکاران، مقاله‌ای تحت عنوان «چارچوب جدید مدیریت اعتماد چند سطحی برای حل مشکلات نامعتبر و پراکنده رتبه‌بندی بازخورد کاربر در محیط‌های ابری» انجام دادند [۸]. در چارچوب این پژوهش، مؤلفه‌های جدیدی برای حل مشکلات نامعتبر و پراکنده ذخیره‌سازی بازخورد تعریف شدند. مطمئناً ارزیابی اعتماد

تک‌ابری و چندابری به همراه راهکارهایی برای بازخوردهای نامعتبر و پراکنده، رتبه‌بندی بازخورد کاربران پیشنهاد شده است که البته مقادیر بازخوردها ایستا در نظر گرفته شده است.

در روش پیشنهادی مدیریت اعتماد از طریق مولفه شخص ثالث قابل اطمینان انجام می‌شود. این عامل با به کارگیری سرویس‌های تعبیه شده در مولفه‌های خود به ارزیابی قابلیت اعتماد تامین‌کنندگان سرویس ابری به صورت عینی و درونی، خوشه‌بندی و رتبه‌بندی سرویس‌های ابری مورد اعتماد بر اساس سلاقی و اولویت‌های شخصی‌سازی شده کاربران، جمع‌آوری و یکپارچه سازی مقادیر اعتماد ارزیابی شده از ارائه‌دهندگان در محیط ابری، ارزیابی اعتماد کاربر ابری مبتنی بر بازخوردهای ثبت شده در سیستم مدیریت اعتماد، پیشنهاد ارائه‌دهندگان با بیشترین شباهت با اولویت‌های کاربران بر مبنای پارامترهای عینی و درونی می‌پردازد. به نحوی که علاوه بر حفظ ویژگی‌های مذکور سرعت و دقت بیشتری در ارزیابی اعتماد دارد.

۳- پیشینه تحقیق

در این مقاله پیشینه مفهوم اعتماد در رایانش ابری و همچنین رویکردهایی که اخیراً در ارزیابی اعتماد صورت پذیرفته است، مورد بررسی می‌گردد. رایانش ابری بدلیل تسهیل دستیابی مشترک به منابع محاسباتی در یک محیط جغرافیایی توزیع شده، از اهمیت تحقیقاتی فراوانی برخوردار است. یک کاربر قادر است که از مجموعه‌ای از منابع استفاده کند. در محیط ابری دسترسی به صورت درخواستی و به صورت سلف سرویس با حداقل تعامل با ارائه‌دهنده اعطا می‌شود. منابع محاسباتی مبتنی بر معماری چند مستاجر مجازی شده و بین مشتریان به اشتراک گذاشته می‌شوند و به صورت خودکار مدیریت می‌شوند. مکان فیزیکی منابع برای کاربران مبهم است. مجازی بودن این بستر به خاطر حساسیت و اهمیت داده‌ها یا اطلاعات ممکن است به لحاظ ذهنیتی برای کاربران قابل اعتماد نباشد، به همین دلیل می‌بایست مساله اعتماد حل شود تا یک موجودیت بتواند از منبع‌ها استفاده کند یا خدمات را بر روی آن مستقر کند. در یک چنین شرایطی مصرف‌کننده و تامین‌کننده بر روی یکدیگر کنترل کامل ندارند [۱]. هنوز مسائل چالش‌برانگیز بسیاری مانند کیفیت ارائه منابع، امنیت و جامعیت و نهان‌سازی داده‌های ذخیره شده در ابر و عدم اطمینان به ارائه‌دهندگان وجود دارد. انتخاب تامین‌کننده قابل اعتماد یکی از مسائل چالش بر انگیز محیط ابر است. اهمیت روزافزون رایانش ابری این امر را به طور فزاینده‌ای ضروری می‌کند که ما با معنای اعتماد در محیط ابر و نحوه ایجاد این اعتماد توسط مشتری، ارائه‌دهنده و به طور کلی تر جامعه

پیشنهاد کردند براساس گواهینامه های قبلی و قابلیت های کنونی تامین کنندگان سرویس ابری، مدل پیشنهاد شده را مدل اعتماد کیفیت سطح سرویس نامگذاری کردند [۱۵]. در این مدل مقدار اعتماد با استفاده از چهار پارامتر، دسترسی پذیری^۳، قابلیت اطمینان^۴، یکپارچگی داده ها^۵ و کارایی زمان پاسخ^۶ محاسبه شده است.

۴- روش ها

در این پژوهش سعی بر آن است تا با استفاده از الگوریتم بهینه فاخته، بتوان به بهترین و قابل اعتمادترین تأمین کننده بر مبنای بازخورد کاربران و کیفیت سرویس در فضای رایانش ابری، در مدت زمان کمتر و کارایی بالاتر دست یافت. از الگوریتم پیشنهادی می توان در مولفه های عامل بخش ثالث مورد اعتماد به منظور بهینه استفاده نمود بدین صورت که بهینه سازی انتخاب، یافتن راه حل های بهینه به طور کامل و دقیق کاملاً بر فرآیند جستجوی ذاتی متکی است. فرآیند جستجوی مورد استفاده برای انتخاب ارائه دهنده بر اساس نگاهت معیارهای درخواست توسط مصرف کننده و کیفیت خدمات منتشر شده توسط ارائه دهندگان پس از مرحله کشف است. در این بخش، ما یک راه حل فراابتکاری به طور دقیق تر بر روی الگوریتم بهینه فاخته برای حل مشکل انتخاب ارائه دهنده در رایانش ابری پیشنهاد می کنیم.

۴-۱- چارچوب پیشنهادی مدیریت اعتماد

مدیریت اعتماد از دو دیدگاه متفاوت تعریف شده است، دیدگاه فراهم کننده سرویس ابری که میزان تطابق پارامترهای کیفیت سرویس از جمله در دسترس پذیری، زمان پاسخ، قابلیت اعتماد، صحت و شفافیت با مقادیر تعیین شده در قرارداد سطح سرویس SLA ارزیابی و سنجیده می شود. دوم دیدگاه کاربر ابری که به ارزیابی و سنجش رفتار کاربر ابری با تامین کنندگان سرویس می پردازد. با در نظر گرفتن چارچوب مدیریت اعتماد در این پژوهش چهار عامل اصلی زیر تقسیم می شوند، فراهم کننده سرویس ابری در حوزه محاسبات ابری، این عامل می تواند محدوده وسیعی از سرویس ها را با مدل های مختلف از جمله SaaS^۶، PaaS^۷، IaaS^۸ ارائه کنند. کاربر ابری می تواند از سرویس های ارائه شده توسط تامین کنندگان استفاده کند و با ارسال درخواست به عامل مدیریت

بر اساس آن دقیق تر خواهد بود. مقادیر تحلیلی نشان داد که رویکرد پیشنهادی این پژوهش، حتی با کیفیت پایین داده ذخیره سازی بازخورد به طور قابل توجهی بهتر از سایر رویکردها عمل می کند. حسن و همکاران، مقاله ای تحت عنوان «مدل پیشرفته مبتنی بر کیفیت سرویس برای ارزیابی اعتماد در محیط ابری» انجام دادند [۹]. مدل پیشنهادی ارزش انباشته اعتماد را محاسبه می کند که به صورت پویا در هر تراکنش به روز می شود و منعکس کننده تراکنش فعلی یا آخرین تراکنش ارائه دهنده در ابر است. آقای قزوینی و همکاران، مقاله ای تحت عنوان «MMLT» یک چارچوب اعتماد چند سطحی متقابل بر اساس اشخاص ثالث قابل اعتماد در محیط های چند ابری»، انجام دادند [۱۰]. این پژوهش را با هدف توسعه چارچوب مدیریت اعتماد موثر، رویکرد جدیدی برای بهبود تعاملات مبتنی بر اعتماد ارائه کرده است. پیریا و لاتا، مقاله ای تحت عنوان «چارچوبی برای مدیریت اعتماد در فضای رایانش ابری»، انجام دادند [۱۱]. در این مقاله چارچوب «اعتماد به عنوان سرویس» (Taas)^۲، برای بهبود روش های مدیریت اعتماد در محیط های ابری پیشنهاد شده است. در این مدل تطبیقی همچنین، تمایزی میان بازخوردهای معتبر و بازخوردهای مخرب در نظر گرفته شده است. احمد و همکارانش، یک مدل اعتماد میان کاربران و تأمین کنندگان ابری پیشنهاد کرده اند که اعتماد را در سه نوبت برقرار می کند و وقتی کاربران ابری در دو نوبت اول قانع شدند، سپس در نوبت سوم آنها می توانند روی تأمین کننده ابری حساب کنند [۱۲]. کومار و همکارانش، چارچوبی را برای اندازه گیری کیفیت و اولویت سرویس های ابر پیشنهاد کردند [۱۳]. آنها پیشنهاد کردند فرآیند تحلیل سلسله مراتبی بر پایه مکانیزم رتبه بندی که سرویس های ابر بر اساس برنامه های کاربردی مختلف مربوط به نیازمندی های کیفیت سرویس می تواند ارزیابی نماید. این روش پیشنهادی فقط برای مشخصه های قابل اندازه گیری کیفیت سرویس استفاده می شود مانند: پاسخگویی، مهارت، اطمینان از سرویس، هزینه، کارایی، امنیت، حریم خصوصی و قابلیت استفاده، استفاده می شود. ظفر و همکارانش، یک مدلی را پیشنهاد کردند که به کاربران سرویس ابر، در پیدا کردن تامین کنندگان سرویس ابر قابل اعتماد و کارآمد، بر مبنای داده های گرفته شده از مسئولان قانونگذار و عملکرد تامین کنندگان سرویس ابر در یک سال گذشته و بازخوردهای گرفته شده از مشتریان کمک می کند [۱۴]. مانول و همکارانش، مدل اعتمادی

⁶ Turnaround Efficiency

⁷ Infrastructure as a Service

⁸ platform as a service

⁹ Service as a service

¹ Multi-level trust management

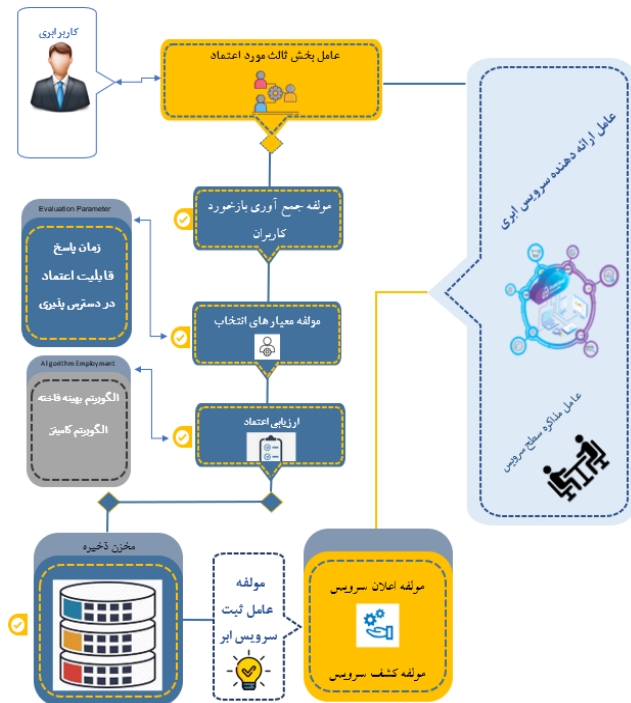
² Trust as a service

³ Availability

⁴ Reliability

⁵ Data Integrity

فرایند انجام مذاکره برای اعتبارسنجی آینده ثبت و نگهداری می‌شود. مولفه دیگر مولفه‌های عامل کاربر ابر، مولفه ناظر قرارداد SLA در سمت کاربران، مانیتورینگ رفتار و کارایی یک سرویس ابری را به منظور تایید و تضمین قراردادهای SLA در سطح مطلوب بر عهده دارد. مولفه عامل بخش ثالث مورد اعتماد یعنی عامل TTP که در حالت کلی واسطی بین کاربران و تامین‌کنندگان سرویس ابری است، به عبارتی دیگر این عامل‌ها نمایندگانی هستند که در سطح ابرهای مختلف توزیع شده‌اند این عامل شامل مولفه‌های جمع‌آوری بازخورد کاربران، مولفه معیارهای انتخاب^۳، یعنی پارامترهای مورد نظر کاربران بر مبنای اولویت آنها، مولفه ارزیابی اعتماد^۴ که وظیفه ارزیابی اعتماد تامین‌کنندگان و کاربران ابری را بر اساس اطلاعات جمع‌آوری شده بر مبنای الگوریتم پیشنهادی بر عهده دارد. مولفه مخزن^۴ ذخیره مقادیر اعتماد حاوی مقادیر اعتماد ارزیابی شده تامین‌کنندگان می‌باشد. عامل TTP برای ارزیابی اعتمادی عینی و درونی تامین‌کنندگان ابری از اطلاعات فراهم شده توسط مولفه جمع‌آوری اطلاعات ناظر و به منظور ارزیابی اعتماد تامین‌کنندگان ابری از بازخوردهای ارسال شده توسط کاربران که توسط مولفه جمع‌آوری بازخوردهایی معتبر و به روز جمع‌آوری شده است استفاده می‌کند. چارچوب پیشنهادی در شکل ۱ نمایش داده شده است.



شکل ۱. چارچوب پیشنهادی مدیریت اعتماد

اعتماد همان ابر درباره قابلیت اعتماد سرویس ابری اطلاعاتی کسب نماید تا از این طریق بهترین سرویس ابری مورد اعتماد را با توجه به نیازمندی‌های وظیفه‌مندی و سلاقی شخصی خود انتخاب کند. بخش ثالث مورد اعتماد^۱، این عامل با به کارگیری سرویس‌های تعبیه شده در مولفه‌های خود وظایف که در ادامه بیان می‌شود را بر عهده دارد: ارزیابی قابلیت اعتماد تامین‌کنندگان سرویس ابری به صورت عینی و درونی، در دسته اول از کیفیت سرویس بعنوان معیار استفاده می‌شود شایان ذکر است کیفیت سرویس قابل مشاهده و اندازه‌گیری است [۱۶]. دسته دوم روش‌هایی هستند که برای ارزیابی و انتخاب سرویس ابری از بازخوردهای ارسال شده توسط کاربران بهره می‌برد. ارزیابی مبتنی بر بازخورد که به آن ارزیابی اعتماد درونی هم گفته می‌شود به طور گسترده در سیستم‌های تجارت الکترونیک از جمله آمازون و ای‌بی به کار گرفته می‌شود [۱۷]. همچنین، خوشه‌بندی سرویس‌های ابری مورد اعتماد بر اساس سلاقی و اولویت‌های شخصی‌سازی شده کاربران. جمع‌آوری و یکپارچه‌سازی مقادیر اعتماد ارزیابی شده از تامین‌کنندگان در محیط ابری. ارزیابی اعتماد کاربر ابری مبتنی بر بازخوردهای ثبت شده در سیستم مدیریت اعتماد. آخرین عامل نیز سرویس ثبت ابری است، این عامل با دو هدف اساسی در چارچوب مدیریت اعتماد قرار گرفته است که شامل اعلان سرویس و کشف سرویس است بدین معنا که یک فراهم‌کننده سرویس ابری قادر است سرویس‌های خود را از طریق این مولفه ثبت نماید و از طرف دیگر کاربران با دسترسی به سیستم مدیریت ابری به کشف سرویس ابری مناسب می‌پردازند. کاربر ابری، می‌تواند از سرویس‌های ارائه شده توسط تامین‌کنندگان استفاده کند و با ارسال درخواست به عامل مدیریت اعتماد همان ابر درباره قابلیت اعتماد سرویس ابری اطلاعاتی کسب نماید تا از این طریق بهترین سرویس ابری مورد اعتماد را با توجه به نیازمندی‌های وظیفه‌مندی و سلاقی شخصی خود انتخاب کند.

۴-۲- مولفه‌های چارچوب مدیریت اعتماد

هر کدام از عوامل ذیل نیازمند مولفه‌هایی به منظور مدیریت اعتماد در محیط محاسبات چند ابری هستند که در ادامه به آنها پرداخته می‌شود. اولین مولفه‌های عامل فراهم‌کننده سرویس ابری است، این مولفه مذاکره SLA را بین کاربر و فراهم‌کننده ابری بر عهده دارد، به بیانی دیگر می‌توان جزئیات قرارداد SLA درباره یک فراهم‌کننده سرویس ابری را از طریق این مولفه به دست آورد، پس از انجام مذاکرات موفقیت‌آمیز بین کاربر و فراهم‌کننده قرارداد و

³ Trust evaluation

⁴ Feedback Repository

¹ Trusted Third Party Age

² Cloud attribute

۴-۳- الگوریتم بهینه فاخته COA

الگوریتم فاخته با الهام از زندگی و رفتار گروهی فاخته‌ها تعریف شده است. شکل ۲ فلوجارت الگوریتم پیشنهادی را نشان می‌دهد.

همانند الگوریتم‌های تکاملی، الگوریتم پیشنهادی با جمعیت اولیه فاخته‌ها آغاز می‌شود. این فاخته‌های در لانه برخی از پرندگان تخم گذاری می‌کنند. برخی از این تخم‌ها که بیشتر شبیه تخم‌های پرند میزبان هستند، امکان زنده ماندن و تبدیل شدن به فاخته بالغ را دارند. تخم‌های دیگر توسط پرندگان میزبان شناسایی شده و از بین می‌روند. تخم‌های رشد یافته نیز بیانگر مناسب بودن لانه‌های آن منطقه است. هر چه تعداد تخم‌های رشد کرده در یک منطقه بیشتر باشد، سود بیشتری در آن منطقه به دست می‌آید. بنابراین مکان‌هایی که در آن تخم‌های بیشتری بالغ می‌شوند پارامتری خواهد بود که الگوریتم بهینه فاخته بهینه‌سازی آن را انجام می‌دهد.

فاخته‌ها به دنبال بهترین نقاط برای تخم گذاری می‌گردند تا درصد زنده ماندن تخم‌هایشان به بالاترین میزان برسد. پس رشد تخم‌ها در بهترین مکان و تبدیل به یک فاخته بالغ، جوامعی را تشکیل می‌دهند. هر جامعه منطقه زیستگاه خود را برای زندگی دارد. بهترین زیستگاه در میان این جوامع، مقصد فاخته‌ها در سایر جوامع خواهد بود. سپس فاخته‌ها به سمت این ناحیه مهاجرت می‌کنند. آنها در جایی نزدیک به بهترین زیستگاه ساکن خواهند شد. با توجه به تعداد تخم‌های هر فاخته و همچنین فاصله فاخته تا نقطه هدف که همان بهترین زیستگاه است، شعاع‌های تخم گذاری به آن اختصاص داده شده است. سپس فاخته در چند لانه تصادفی در داخل شعاع تخم‌گذاری خود شروع به تخم گذاری می‌کند. این روند تا زمانی ادامه می‌یابد که بهترین موقعیت با حداکثر ارزش سود بدست آید و بیشتر جمعیت فاخته حول همان موقعیت جمع شوند.

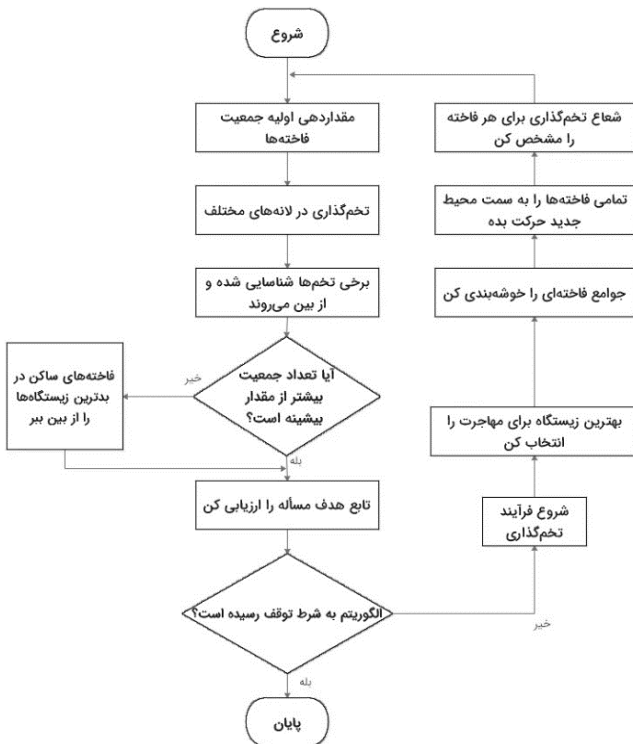
۴-۴- ایجاد سکونت گاه اولیه فاخته

در یک مسئله N_{var} بعد، یک سکونت گاه $N_{var} + 1$ آرایه‌ای است که نشان‌دهنده محل فعلی زندگی فاخته است. این آرایه بصورت زیر است.

$$habitat = [x_1, x_2, \dots, x_{N_{var}}] \quad (1)$$

میزان مناسب بودن یا مقدار سود در $habitat$ فعلی با ارزیابی تابع سود (fp) در $habitat$ به دست می‌آید.

$$profit = fp(habitat) = fp[x_1, x_2, \dots, x_{N_{var}}] \quad (2)$$



شکل ۲. فلوجارت الگوریتم فاخته [۱۸]

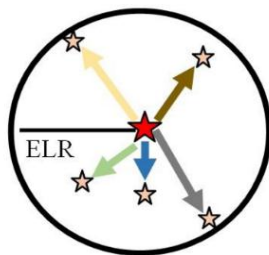
برای شروع الگوریتم بهینه‌سازی یک ماتریس $habitat$ به سائز $pop \times N_{var}$ تولید می‌شود، سپس برای هر کدام از این $habitat$ ها تعدادی تصادفی تخم تخصیص می‌یابد.

هر فاخته در یک دامنه مشخص تخم‌های خود را می‌گذارد.

(۳)

$$ELR = \alpha \times \frac{\text{Number of current cuckoo s eggs}}{\text{Total number of eggs}} \times (Var_{hi} - Var_{low})$$

در این فرمول ELR ^۱ شعاع تخم‌گذاری و α یک عدد صحیح است که قرار است حداکثر مقدار شعاع تخم‌گذاری را کنترل کند، بنابراین بعد از هر تخم‌گذاری $P\%$ از تمام تخم‌ها (معمولاً 10%) که مقدار تابع سود آنها کمتر است نابود می‌شوند. بقیه جوجه‌ها در لانه میزبان تغذیه شده و رشد می‌کنند [۱۸].



شکل ۳. روش تخم‌گذاری هر فاخته در یک دامنه مشخص [۱۸]

^۱ Egg laying Radius

۴-۵- Levy Flight، الگویی از حرکتهای تصادفی در

فضای مسئله

پرواز لوی یک مدل تصادفی برای حرکت ذرات است که در آن طول قدمها به دنباله تصادفی از مقادیر مثبت نمونه برداری شده از یک پرواز تصادفی پیروی می کند. در این روش، ذرات با طول قدمهای تصادفی به طور گسترده حرکت می کنند و در نتیجه به سرعت به سراسر فضا پخش می شوند.

$$Levy \sim U = t^{-\lambda} \quad (1 < \lambda \leq 3) \quad (11)$$

این روش می تواند به عنوان یک الگوریتم بهینه سازی در بسیاری از مسائل به کار گرفته شود. فرمول به صورت زیر است.

$$x_{n+1} = x_n + \sum_{i=1}^{\infty} \Delta x_i \quad (12)$$

$$P(\Delta x) = \frac{\alpha}{2|\Delta x|^{1+\alpha}} \quad (13)$$

برای استفاده از مدل پرواز لوی به عنوان یک الگوریتم بهینه سازی، می توان همان فرمول را به صورت زیر بازنویسی کرد.

$$x_{n+1} = x_n + \Delta x \quad (14)$$

در این فرمول، α یک پارامتر مثبت است که نشان دهنده شدت حرکت تصادفی است [۱۷]. ما با انجام پرواز لوی از فاخته های موجود (راه حل ها) فاخته های جدید (راه حل) تولید می کنیم. مطابق معادله ۱۵.

$$x_{new} = x_i + \beta \otimes Levy(\lambda) \quad (15)$$

قانون تکامل برای هر فاخته و در اینجا $\beta > 0$ نیز اندازه گام است. در بیشتر موارد، می توانیم از $\beta = 1$ استفاده کنیم. \otimes به معنای ضربهای ورودی است [۱۹].

۴-۶- خوشه بندی

الگوریتم K-Means یک الگوریتم بر پایه ی تکراری است که سعی می کند مجموعه داده ها را به زیر گروه های متمایز بدون همپوشانی تعریف کند، که به این زیر گروه ها خوشه گفته می شود. در این گروه ها هر نقطه داده فقط به یک گروه تعلق دارد. در این الگوریتم سعی می شود نقاط داده درون خوشه ای را تا حد ممکن شبیه به هم ساخت و در عین حال خوشه ها را تا حد امکان متفاوت (دور) از هم تعریف کرد. این الگوریتم داده ها را به یک خوشه اختصاص می دهد به طوری که مجموع فاصله مربع شده بین نقاط داده و مرکز گروه (میانگین محاسبه تمام نقاط داده ای که به آن خوشه تعلق دارند) در حداقل باشد. هدف از خوشه بندی طبقه بندی اطلاعات با میزان تشابه بین آنها در دو یا چند گروه است. مسئله خوشه بندی را

زمانی که فاخته های بالغ در تمام محیط، زندگی می کنند، تشخیص این که کدام فاخته متعلق به کدام گروه است دشوار است. برای حل این مشکل، گروه بندی فاخته ها با روش خوشه بندی K-Means انجام می شود. اکنون که گروه های فاخته تشکیل شده اند، میانگین ارزش سود آنها محاسبه می شود. سپس حداکثر مقدار این میانگین سود، گروه هدف را تعیین می کند و در نتیجه بهترین زیستگاه آن گروه، زیستگاه مقصد جدید فاخته های مهاجر است. هر فاخته تنها λ از مسیر را با درجه انحراف φ رادیان پرواز می کند پارامترهای λ و φ به فاخته ها کمک می کند تا مناطق بیشتری را برای یافتن زیستگاه بهینه با بیشترین منبع غذایی و بهترین شرایط زندگی جستجو کنند. وقتی تمام فاخته ها به سمت نقطه هدف مهاجرت کردند و نقاط سکونت جدید هر کدام مشخص شد، هر فاخته صاحب تعدادی تخم می شود. با توجه به تعداد تخم هر فاخته یک ELR برای آن مشخص می شود و سپس تخمگذاری شروع می گردد. فرمول عملگر مهاجرت در الگوریتم بهینه سازی فاخته به صورت رابطه زیر است.

$$X_{NextHabitat} = X_{CurrentHabitat} + F(X_{GoalPoint} - X_{CurrentHabitat}) \quad (6)$$

کاهش مرحله ای شعاع تخم گذاری می تواند منجر به بهبود جستجو شده و دقت جوابها را بالا ببرد. در الگوریتم COA این تغییر را می توان با کاهش تدریجی ضریب آلفا انجام داد. شعاع تخمگذاری به صورت زیر کاهش می یابد.

$$t = \frac{max-iter}{C} \quad (7)$$

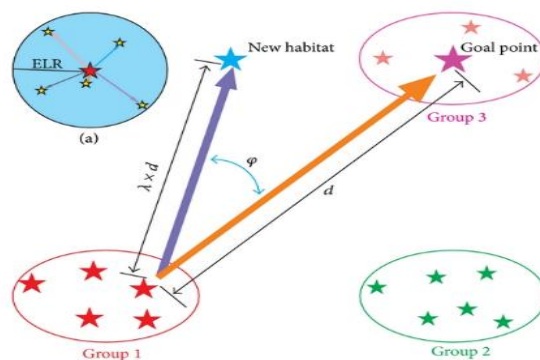
C یک مقدار ثابت از بازه (۰،۲۰) و $max-iter$ ماکزیمم تعداد تکرارها است. از این رو،

$$\alpha = \left\| \frac{iteration}{t} \right\| + 1 \quad (8)$$

$$e = Var_{hi} - Var_{low} \quad (9)$$

$$e_{new} = \frac{e_{old}}{\alpha} \quad (10)$$

$\| \|$ به معنای قسمت صحیح عدد خاص خواهد بود. بنابراین e در هر تکرار t با استفاده از معادله بالا اصلاح می شود.



شکل ۴. مهاجرت فاخته ها به سمت سکونتگاه جدید (۱۸).

۵- ارزیابی اعتماد

در محیط ابری، تامین کنندگان بسته به پارامترهای مختلف خدمات ابری به ارایه سرویس می پردازند. باید بهترین تامین کننده با سرویس بهتر را از لیست تامین کنندگان کاندید با ویژگی های مختلف کیفیت خدمات بسته به نیاز مصرف کنندگان انتخاب شود. مشکل انتخاب تامین کننده را با کاهش اندازه مسئله، همچنین کاهش فضای جستجو برای کاربران به صورت زیر مدل می شود. به طوری که،

$$p = \{p_1, p_2, \dots, p_n\} \quad (20)$$

که نشان دهنده n کاندید فراهم کننده با عملکرد یکسان است.

$$QOS(P_i) = \{qos_1, qos_2, \dots, qos_m\} \quad (21)$$

که بیانگر کیفیت تامین کننده p_i با m معیار است.

$$User Rq = \{CF; Const\} \quad (22)$$

فاکتور هزینه^۱ پارامتری است که مشتری برای هزینه اعمال محدودیت می کند و محدودیت^۲ نیز می تواند شامل هرگونه محدودیت همچون زمان پاسخدهی باشد، بعنوان مثال کاربر می تواند زمان پاسخدهی ۵ روزه را تعیین نماید، به طور کلی به محدودیت های مصرف کننده اشاره دارند و بر اساس نیاز مشتری قابل تغییر می باشند.

$$Const(p_i) = \{Const_1, Const_2, \dots, Const_d\} \quad (23)$$

مجموعه ای از محدودیت های تعریف شده توسط مصرف کننده که با $0 < d < m$ نشان داده می شود. وزن های ارائه شده توسط کاربر برای هر QOS را معادله (۲۴) نشان می دهد.

$$W = \{w_1, w_2, \dots, w_m\} \\ \sum_{j=1}^m W_j = 1 \quad (24)$$

برای مقادیر مثبت معادله (۲۵).

$$Qos^+(p) = \begin{cases} \frac{qi - q_{imin}}{q_{imax} - q_{imin}} & \text{if } q_{imax} \neq q_{imin} \\ 1 & \text{if } q_{imax} = q_{imin} \end{cases} \quad (25)$$

برای مقادیر منفی معادله (۲۶)

$$Qos^-(p) = \begin{cases} \frac{q_{imax} - qi}{q_{imax} - q_{imin}} & \text{if } q_{imax} \neq q_{imin} \\ 1 & \text{if } q_{imax} = q_{imin} \end{cases} \quad (26)$$

U_p نیز تابع هدف (مقدار تناسب) می باشد.

بصورت کلی می توان به شرح زیر بیان کرد.

در مجموع داده های $x = \{x_1, x_2, \dots, x_n\}$ شی داده شده به k خوشه (c_1, c_2, \dots, c_k) تقسیم می شود، بطوری که شباهت داده های درون هر خوشه حداکثر و شباهت بین داده های بین خوشه های متفاوت حداقل شود. این مسئله در معادله زیر بیان شده است.

$$\begin{cases} \bigcup_{i=1}^k c_i = x \\ c_i \cap c_j = \emptyset \quad i, j = 1, 2, 3, \dots, k \\ c_i \neq \emptyset \quad i = 1, 2, 3, \dots, k \quad i = 1, 2, \dots, k \end{cases} \quad (16)$$

k تعداد خوشه ها و c_i خوشه i ام می باشد از نظر ریاضی خوشه c_i بصورت معادله ۱۷ مشخص است، باید فاصله هر داده با مراکز خوشه محاسبه شود.

$$\begin{cases} c_i = \{x_j \mid \|x_j - \mu_i\| \leq \|x_j - \mu_p\|, x_j \in x\} \\ p \neq i, p = 1, 2, \dots, k \\ \mu_i = \frac{1}{|c_i|} \sum_{x \in c_i} x_j \quad i = 1, 2, 3, \dots, k \end{cases} \quad (17)$$

در این فرمول، x_j نماینده یک داده در مجموعه داده های x است. μ_i نماینده مرکز خوشه i است و k تعداد کل خوشه ها را نشان می دهد. بعنوان مثال در اینجا، $\|x_j - \mu_1\|$ نشان دهنده فاصله داده x_j تا مرکز خوشه ۱ است و $\|x_j - \mu_2\|$ و $\|x_j - \mu_3\|$ نشان دهنده فاصله داده x_j تا مراکز خوشه ۲ و ۳ به ترتیب هستند [۲۰].

$\| \|$ بیانگر فاصله اقلیدسی و یا کسینوسی بین هر دو نقطه است، μ_i مرکز خوشه c_i است که در واقع میانگین تمامی نقطه ها در خوشه را نشان می دهد. مجموع خطاها بصورت زیر محاسبه می شود.

$$SSE = \sum_{i=1}^k \sum_{x_j \in c_i} \|x_j - \mu_i\|^2 \quad (18)$$

مقدار تابع برازش در خوشه بندی داده ها برابر با میانگین فاصله داده ها از مراکز خوشه ها می باشد، که بصورت معادله (۱۹) است.

$$fitness(\mu_i) = \left(\frac{1}{n} \sum_{m=1}^k \sum_{x_j \in c_m} \|x_j - \mu_{im}\| \right)^{-1} \quad (19)$$

الگوریتم خوشه بندی K-means مرحله های توصیف شده در بالا (محاسبه مجدد نقطه اولیه و تخصیص خوشه) را تا زمانی که شرط همزمانی (به عنوان مثال، ثابت ماندن در دو تکرار متوالی) برآورده شود، تکرار می کند. این مرحله نقطه اوج نتایج خوشه بندی است. این الگوریتم دارای چند مزیت است که آن را معروف کرده و سادگی و سهولت پیاده سازی مهمترین فاکتورها به حساب می آیند. به دلیل پیچیدگی خطی، این الگوریتم بسیار سریع کار می کند [۲۱].

² Constraint

¹ Cost factor

$$U_p = \sum_{k=1}^m QOS_i \pm P * Wi \quad (27)$$

در این مورد، دیگر تمایزی بین تخم، لانه یا فاخته وجود ندارد و هر لانه مربوط به یک تخم است که همچنین نشان دهنده فاخته‌ای است که با یک تامین‌کننده مدل شده است. الگوریتم پیشنهادی با استفاده از پرواز لوی و خوشه‌بندی K-Means به منظور افزایش سرعت انتخاب و انتخاب دقیق‌تر قابل‌اعتمادترین تامین‌کننده و پرهیز از همگرایی زودرس به صورت زیر است.

الگوریتم ۲: شبه کد الگوریتم بهینه فاخته (COA)

Input: n (number of Cuckoos); Max It; QoS-/+ ; Const;
Output: Best provider

Data, $x = \{x_1, x_2, \dots, x_n\}$;
Cluster number, K ;
Clusters: $c = \{c_1, c_2, \dots, c_k\}$

- 1: Initialize: Generate initial population
Num Cuckoos: number of initial Population;
Min Number Of Eggs: min number of eggs for each cuckoo;
Max Number Of Eggs: maxi number of eggs for each cuckoo;
Max Iter: max iteration of the Cuckoo Clustering Algorithm;
Max Num Of Cuckoos: max number of cuckoos that can live at the same time;
Var low;
Var hi;
- 2: Generate the Position of Initialize Population randomly which each single $\mu = \{\mu_1, \mu_2, \dots, \mu_{numCuckoos}\}$; contains K randomly generated centroid vectors μ_i
 $\mu_i = \{\mu_{i1}, \mu_{i2}, \dots, \mu_{ik}\}$;
- 3: for $t=1$: max Iter do
- 4: for $i=1$: num Cuckoos do
- 5: Initialize number of eggs for each μ_i
- 6: end for
- 7: for $i=1$: num Cuckoos do
8. Calculate ELR (egg laying radius) for current μ_i
- 9: end for
- 10: Lay eggs in different nests
- 11.Remove the eggs on the same positions for all Cuckoos
- 12: for $i=1$: num Cuckoos do
- 13: Evaluate fitness function for each μ_i and its eggs and put them under each other
- 14: end for
- 15: if Population > max Num Of Cuckoos then
- 16: abandoned cuckoos in worst position
- 17: end if
18. Update the global best position μ_{best} : Select the best $\mu_{best i}$ from $\{\mu_1, \mu_2, \dots, \mu_{numCuckoos}\}$;
- 19: Move all cuckoos toward μ_{best}
- 20: end for
21. Result (best provider)

m تعداد ویژگی‌های یک تامین‌کننده است و مقادیر $+$ و $-$ میزان ارزش را نشان می‌دهد به نحوی که هر چه مقدار QoS مثبت‌تر باشد ارزش بیشتری دارد مثل قابلیت اعتماد و سرعت و مقدار $-$ نیز هر چه کمتر باشد ارزش بیشتری خواهد داشت همانند هزینه کمتر یا زمان پاسخگویی خدمات [۲۲]. البته برای تسهیل در محاسبه مقدار تابع هدف انتخاب تامین‌کننده، مقادیر ویژگی تامین‌کنندگان مختلف را به عددی بین صفر و ۱ نرمال می‌کنیم و شبه کد الگوریتم جستجوی فاخته بصورت زیر می‌باشد.

حال از یک فرآیند جستجوی تسریع شده برای بهبود نرخ همگرایی استفاده می‌شود. برای بهبود نرخ همگرایی، در الگوریتم پیشنهادی از پرواز لوی برای مزایای آن استفاده شده است. پروازهای لوی همانگونه که قبلاً توضیح داده شد برای کاوش در مناطق جستجو در مقیاس بزرگ کارآمدتر هستند. هر تخم در یک لانه نشان‌دهنده یک تامین‌کننده است و هر فاخته می‌تواند یک تخم بگذارد هدف استفاده از بهترین تامین‌کننده برای جایگزینی یک تامین‌کننده با ویژگی‌های کمتر در محیط ابری است (لانه).

الگوریتم ۱: شبه کد الگوریتم جستجوی فاخته (CSA)

Input: n (number of Cuckoos); Max It; QoS-/+; Const
Output: Best provider

Objective Function $f_x, x = \{x_1, x_2, \dots, x_n\}$

1. Generate initial population $x_i (1, 2, \dots, x)$
2. Evaluate fitness
- 3.while ($t \leq Max It$) do
 $It = It+1$
4. Get a Cuckoo randomly by Levy Flights
- 5.Evaluate its quality/Fitness f_i
6. Choose a random nest j ;
- 7.if ($f_i \geq f_j$) then
8. Replace j by the new solution ;
- 9.End if
- 10.Worst nest is abandoned with probability (pa) and new nest is built with Levy flights;
- 11.Keep the best solution (or cuckoo with quality solutions);
- 12.Rank the solutions and find the current best provider;
- 13.End while
- 14.Result (best provider)

۶- شبیه‌سازی و ارزیابی

است.

به طبع ضریب β نیز با بررسی و کاهش به مقدار ۰,۹۵ و خوشه‌بندی با ۱۰۰ گروه به فاخته‌ها کمک می‌کند تا محیط را به سرعت به چندین بخش تقسیم کرده و بهترین ناحیه را به صورت تخمینی مشخص نمایند. ناحیه‌ای که به احتمال زیاد شامل نقطه بهینه کلی می‌باشد. سپس تمامی فاخته‌ها به سمت این ناحیه مهاجرت می‌کنند و داخل آن ناحیه را به صورت بهتری جستجو می‌نمایند، این امر موجب همگرایی بسیار سریعتر الگوریتم می‌شود. خود فرایند تخمگذاری با جمعیت ۱۰۰ به تنهایی یک فرایند جستجوی محلی در مناطق بیشتر است. این در حالی است که سایر الگوریتم‌های بهینه‌سازی از وجود چنین عملگری بی‌بهره اند و نیاز دارند تا با الگوریتم‌های جستجویی ترکیب شوند. البته باید توجه داشت که می‌توان ضریب شعاع تخم‌گذاری را در هر تکرار با نزدیک شدن به جواب بهینه برای افزایش سرعت همگرایی الگوریتم کوچک‌تر نمود.

۶-۲- ارزیابی و رویکرد پیشنهادی

ارزیابی و رویکرد پیشنهادی شامل دو بخش زمان و دقت بهتر در پاسخ به وسیله الگوریتم پیشنهادی بود که در ادامه به تفکیک، مورد بررسی قرار می‌گیرد.

جدول ۱. پارامترهای عینی و درونی فراهم‌کنندگان کاندید و کاربران ابری

پارامترهای عینی و درونی سمت ارائه‌دهنده					پارامتر			
W	CU4	CU3	CU2	CU1	سطح خدمات	فراهم‌کننده		
						CS3	CS2	CS1
۰,۱	۲	۳	۶	۲	[۲:۱۰]	۸	۵	۳
۰,۱	۰,۶	۰,۷	۰,۵	۰,۹	[۵:۱۰]	۰,۶	۰,۸	۰,۶
						۰,۸	۱	۰,۹
۰,۳					[۱۰:۱۰۰]	۱۵	۲۰	۱۷
۰,۴	۲	۳	۳	۴	[۱:۵]	۴	۲	۳
۰,۱	۱۱	۱۳	۱۸	۱۷	[۱۰:۲۰]	۱۶	۱۸	۲۰
۰,۵	۰,۱	۰,۵	۰,۴	۰,۲	-	-	-	-

جدول ۲. مقادیر الگوریتم بهینه فاخته

پارامتر	مقدار
Population Size	۱۰۰
pa	۰,۹
β	۰,۹۵
Number of Clusters	۱۰۰

به منظور ارزیابی قابلیت الگوریتم پیشنهادی تمامی آزمایش‌ها توسط کامپیوتر با ۴ گیگابایت حافظه و پردازنده Core(TM) i3-1115G4، 3.00GHz صورت گرفته است. مراحل شبیه‌سازی در محیط نرم افزار متلب انجام شد و به منظور انجام مراحل شبیه‌سازی و ارزیابی راهکار پیشنهادی یک مجموعه داده بکار گرفته شده است.

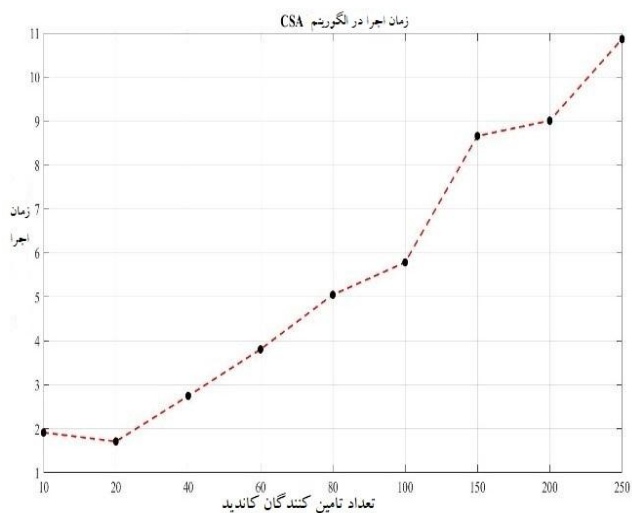
۶-۱- پارامترهای شبیه‌سازی

برای تخمین زمان پاسخگوی الگوریتم پیشنهادی تعداد کاربران و تامین‌کنندگان، در مراحل مختلف قابل تنظیم است چرا که این تغییر نمایانگر کاربرد الگوریتم COA با شرایط مختلف در فضای مسئله خواهد بود. از پارامترهای دیگر قابل تنظیم نیز مقادیر مولفه‌های اعتماد عینی و درونی بر مبنای اولویت کاربران می‌باشد. به عنوان نمونه در پارامترهایی همچون، دسترسی سریع و قابلیت اطمینان به عوامل زیادی همچون میزان پردازش سخت افزاری بستگی دارد، پس ممکن است کاربر بسته به نیاز خود به تعداد مختلف پردازشگر نیاز داشته باشد. این اولویت با توجه به خدمات سمت تامین‌کننده که دارای حداقل و حداکثر امکانات مطابق با قرارداد سطح سرویس با شخص ثالث است انتخاب می‌شود. در پارامترهای درونی نیز به عنوان نمونه امتیاز کاربران بر مبنای سن مجازی و میزان بازخوردهای ثبت شده، محاسبه و ثبت گردید. سن مجازی نیز اشاره به مدت زمان استفاده کاربر از سرویس‌های ابری است. به نحوی که کاربری که چندین سال متوالی از یک سرویس ابری بهره برده دارای امتیاز بیشتری نسبت به کاربرانی است که به تازگی دریافت سرویس کرده‌اند و به لحاظ اعتبار بازخورد از سطح بالاتری برخوردار است.

تمامی مقادیر پارامترهای موجود در جدول ۱ برای بررسی پاسخ در شرایط مختلف قابل تغییر و تنظیم بودند. سایر پارامترهای تنظیم شده در الگوریتم نیز در جدول ۲ نشان داده شده است. این موارد شامل، تعداد لانه‌های الگوریتم یا جمعیت اولیه است. مقادیری که به شعاع حرکت و ضریب الفا و بتا مربوط هستند ثابت و در طول نسل‌های آینده تغییر نمی‌کنند. داشتن پارامترهای کمتر جهت مقدار دهی، همگرایی سریعتر و دقت بالاتری به همراه دارد. مقادیر موجود در جدول ۲ به نحوی انتخاب شده‌اند که منجر به بهبود جستجو شده و دقت جواب‌ها را بالا ببرند. بعد از هر تخمگذاری $P\%$ از تمام تخم‌ها (معمولاً ۱۰٪) که مقدار تابع سود آنها کمتر است نابود می‌شوند. بقیه جوجه‌ها در لانه میزبان تغذیه شده و رشد می‌کنند. pa با توجه به پاسخ بهتر در اینجا ۰,۹ در نظر گرفته شده

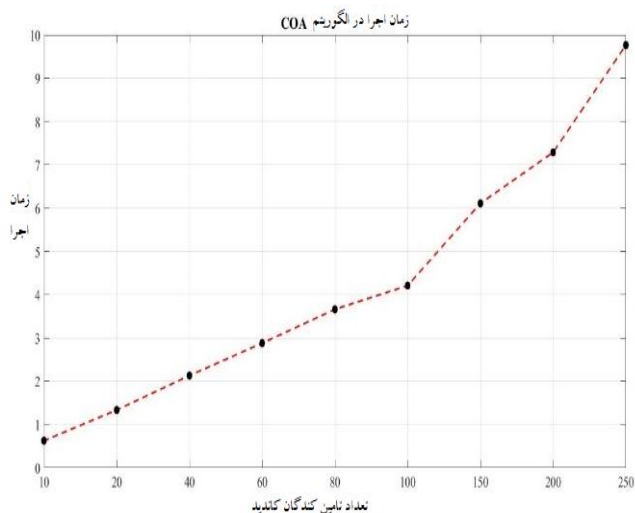
۳-۶- نتایج بررسی تخمین زمان اجرا

در مرحله اول، CSA با تغییر تعداد فراهم‌کنندگان برای تخمین زمان پاسخ شبیه‌سازی گردید، مشاهده شد که زمان اجرا با افزایش تعداد فراهم‌کننده افزایش می‌یابد (شکل ۵).



شکل ۵. عملکرد زمان الگوریتم نسبت به تعداد تامین‌کنندگان کانديد

سپس با همان شرایط COA شبیه‌سازی گردید (شکل ۶).



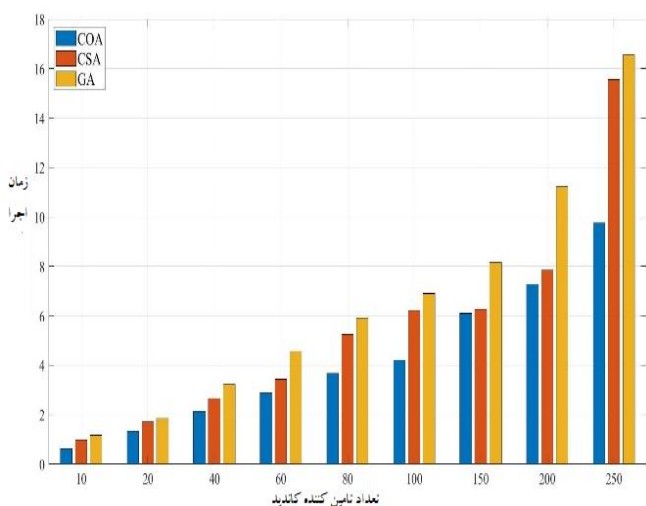
شکل ۶. عملکرد زمان الگوریتم COA نسبت به تعداد تامین‌کنندگان کانديد

همانگونه که در شکل ۶ مشاهده می‌شود با افزایش تعداد فراهم‌کنندگان، فرایند همچون الگوریتم CSA با افزایش زمان روبه‌رو می‌شود، این افزایش زمان به دلیل ایجاد لانه است، با وجود زمان صرف شده برای ایجاد لانه، پاسخ به درخواست مشتری به صورت بلادرنگ و به همان سرعت انجام می‌شود که باعث می‌شود زمان پاسخگویی این الگوریتم قابل قبول‌تر، کمتر و همچنین قابل اعتمادتر از الگوریتم CSA باشد.

۴-۶- تخمین زمان اجرای COA در مقایسه با سایر

الگوریتم‌ها

با توجه به کاربرد گسترده الگوریتم ژنتیک، با قابلیت بالا برای یافتن جواب بهینه از طریق جستجو، مقایسه الگوریتم پیشنهادی با الگوریتم ژنتیک صورت گرفته است. این الگوریتم در زیر مجموعه الگوریتم‌های جستجو قرار دارد بعلاوه در پژوهش‌هایی که تاکنون انجام شده بهترین پاسخ با استفاده از الگوریتم CSA بدست آمده است و هدف بهینه کردن سرعت و دقت جستجو نسبت به پژوهش‌های قبلی است. سناریوی اول در شبیه‌سازی دوم، بدین صورت بود که COA با CSA و GA برای همان تعداد از فراهم‌کنندگان ابری نامزد از نظر زمان پاسخگویی با تعداد در خواست‌دهنده مختلف مقایسه گردید. پس از پیاده‌سازی این الگوریتم‌ها با استفاده از نرم افزار متلب، الگوریتم‌ها زمانی که به تعداد تکرارهای مشخص شده می‌رسند متوقف می‌شوند. برای ارزیابی الگوریتم‌ها، تعداد فراهم‌کنندگان ابری کانديد در محدوده ۱۰ تا ۲۵۰ تغییر یافتند. شکل ۷ مقایسه زمان اجرای راه حل به‌دست‌آمده توسط سه الگوریتم را نشان می‌دهد. مشاهده شد که وقتی تعداد فراهم‌کنندگان ابری کانديد افزایش می‌یابد، زمان اجرای COA کمتر از CSA می‌باشد اما CSA نزدیک به GA است. زمان پاسخ COA به‌طور قابل توجهی کم است، این به دلیل زمان همگرایی آن است. مشاهدات نشان داد که COA در یافتن بهترین خدمات با نرخ موفقیت بالاتر در زمان کمتر بسیار کارآمدتر است.



شکل ۷. مقایسه زمان اجرای الگوریتم‌ها با ۲۵۰ فراهم‌کننده کانديد

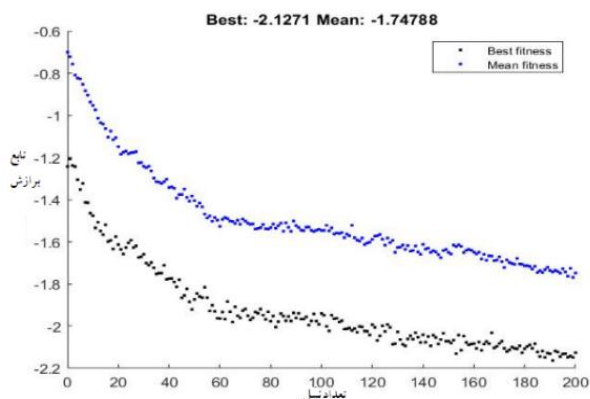
در جدول ۳، تعداد کاربران متقاضی، تعداد فراهم‌کنندگان و مقایسه زمان هر سه الگوریتم قابل مشاهده است. مثلاً در حالتی که تعداد فراهم‌کنندگان ۱۰ می‌باشد، الگوریتم پیشنهادی در ۰.۶۲۵ ثانیه

جدول ۴. مقایسه زمان اجرای الگوریتم‌ها با ۱۲۵۰ فراهم‌کننده کاندید

ردیف	تعداد تامین کننده کاندید	تعداد کاربران متقاضی	زمان (ثانیه)		
			GA	CSA	COA
۱	۱۰	۳	۱,۹۰۳۱	۱,۵۷۹۶	۱,۳۷۰۲
۲	۲۰	۷	۵,۵۰۷۰	۴,۰۲۰۱	۳,۵۷۹۵
۳	۴۰	۱۲	۷,۷۵۷۶	۶,۲۰۱۶	۴,۱۸۹۱
۴	۶۰	۲۰	۱۴,۰۱۶۷	۹,۸۱۱۷	۹,۳۹۱۲
۵	۸۰	۲۵	۱۹,۵۴۴۹	۱۷,۹۸۱۳	۱۰,۵۵۴۲
۶	۱۰۰	۳۰	۲۰,۷۵۷۰	۱۵,۵۶۷۸	۱۲,۰۳۹۱
۷	۱۵۰	۴۰	۳۰,۱۶۲۹	۲۲,۰۱۸۹	۲۱,۱۱۴۰
۸	۲۰۰	۵۰	۳۴,۰۳۳۸	۳۲,۳۳۲۱	۱۹,۳۹۹۳
۹	۲۵۰	۶۵	۴۳,۷۱۸۳	۳۸,۰۳۴۹	۲۶,۲۳۱۰
۱۰	۴۵۰	۸۰	۶۶,۳۴۱۱	۵۲,۴۰۹۵	۴۱,۷۹۴۹
۱۱	۸۵۰	۱۰۰	۶۹,۶۶۳۹	۶۴,۰۹۰۸	۳۷,۶۱۸۵
۱۲	۱۰۰۰	۱۵۰	۱۳۹,۶۴۱۴	۱۱۷,۲۹۸۸	۷۲,۶۱۳۵
۱۳	۱۲۵۰	۲۰۰	۲۰۷,۸۳۳۲	۱۵۷,۹۵۳۲	۱۱۶,۳۸۶۶

۶-۵- سرعت همگرایی الگوریتم برای کاهش زمان پاسخ

سرعت همگرایی در بهینه‌سازی به عواملی همچون نوع الگوریتم مورد استفاده در بهینه‌سازی، شرایط مسئله به این مفهوم که برخی مسائل بهینه‌سازی ممکن است به دلیل قیدها و محدودیت‌های تاخیر در همگرایی داشته باشند. اندازه و پیچیدگی مسئله (تعداد متغیرها و قیدها)، نقطه شروع اولیه الگوریتم، پارامترهای الگوریتم به عنوان مثال، انتخاب درست مقادیر برای پارامترهای مانند اندازه گام، تعداد تکرارها و سایر موارد می‌تواند به بهبود سرعت همگرایی کمک کند. عامل دیگر منطقه جواب است، در برخی مسائل، ناحیه جواب بهینه ممکن است حجم بزرگی داشته باشد و الگوریتم‌ها برای همگرایی به آن نیاز داشته باشند. این فاکتورها تنها برخی از عواملی هستند که ممکن است بر سرعت همگرایی نمودار در بهینه‌سازی تأثیر بگذارند. در شبیه‌سازی انجام شده سرعت همگرایی الگوریتم‌ها با ۲۰۰ تکرار به شرح ذیل می‌باشد.



شکل ۹. سرعت همگرایی GA

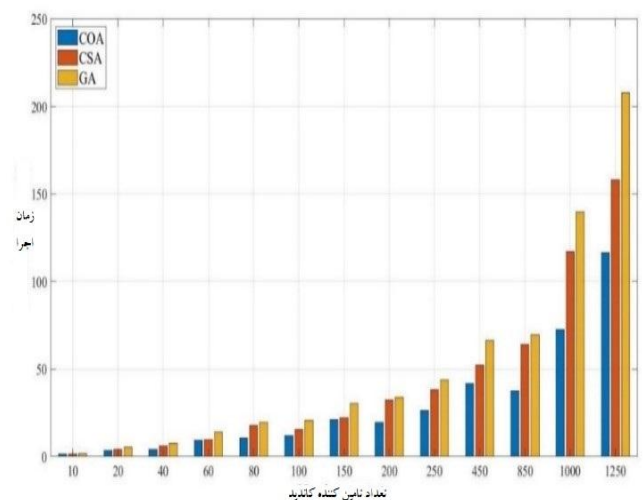
به پاسخ رسید، درحالی که با همان شرایط الگوریتم CSA در ۹,۹۸۴، ثانیه و الگوریتم GA در ۱,۱۵۷،۱۵۷ ثانیه به پاسخ می‌رسند. زمانی که تعداد فراهم‌کنندگان به ۲۵۰ می‌رسد، کمترین زمان به ۹,۷۶۵ ثانیه و بیشترین زمان به ۱۶,۵۵۱ ثانیه می‌رسد.

جدول ۳. مقایسه زمان اجرای الگوریتم‌ها با ۲۵۰ تامین‌کننده

ردیف	تعداد کاربران متقاضی	تعداد تامین کننده	زمان (ثانیه)		
			GA	COA	CSA
۱	۳	۱۰	۱,۱۵۷۸	۰,۶۲۵۲	۰,۹۸۴۱
۲	۷	۲۰	۱,۸۵۳۲	۱,۳۳۴۳	۱,۵۰۷۰
۳	۱۲	۴۰	۳,۲۲۹۳	۲,۱۳۱۳	۲,۶۴۸۰
۴	۲۰	۶۰	۴,۵۷۲۰	۲,۸۸۰۳	۳,۴۲۹۰
۵	۲۵	۸۰	۵,۹۰۶۰	۳,۶۶۱۷	۵,۲۵۶۴
۶	۳۰	۱۰۰	۶,۹۰۱۹	۴,۲۱۰۱	۶,۲۱۱۷
۷	۴۰	۱۵۰	۸,۱۴۶۵	۶,۱۰۹۸	۶,۲۷۲۸
۸	۵۰	۲۰	۱۱,۲۱۱۶	۷,۲۸۷۵	۷,۴۸۸۱
۹	۶۵	۲۵۰	۱۶,۵۵۱۴	۹,۷۶۵۳	۱۵,۵۵۸۴

در سناریوی دوم، COA با CSA و GA با تنظیم مقادیر در شرایط مختلف شامل تعداد فراهم‌کنندگان ابری نامزد و تعداد کاربران متقاضی، معیار کیفیت سرویس ارائه شده توسط ارائه‌دهنده و اولویت‌های مورد نیاز کاربر و از نظر زمان پاسخگویی با ۲۰۰ تکرار مقایسه گردید (شکل ۸).

همانگونه که در جدول ۴ مشاهده می‌شود، الگوریتم پیشنهادی در یافتن بهترین تامین‌کننده در زمان کمتر نسبت به سایر الگوریتم‌های بررسی شده بسیار کارآمدتر بود، چرا که این الگوریتم برای جستجو در فضای مسئله طراحی شده است و بصورت کلی و در فضای مسئله بزرگتر در زمان کمتر به نتایج بهتر می‌رسد.



شکل ۸. مقایسه زمان اجرای الگوریتم‌ها با ۱۲۵۰ ارائه‌دهنده سرویس کاندید

۶-۶- ارزیابی دستیابی به بهترین پاسخ با بیشترین

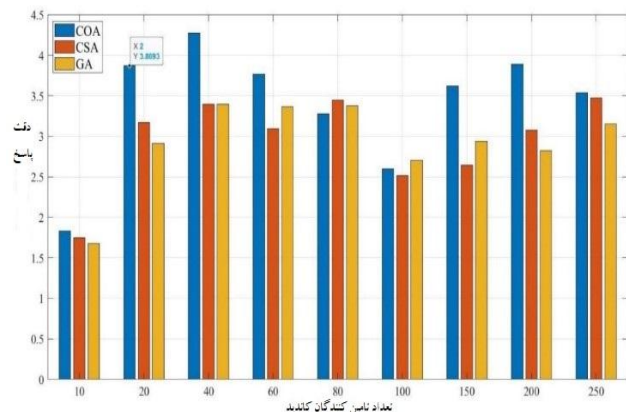
دقت

پیشنهاد تامین‌کننده به کاربر در کمترین زمان به تنهایی موثر نمی‌باشد مسئله مهم‌تر دقت در پاسخ و معرفی بهترین تامین‌کننده بر مبنای اولویت‌های کاربران حتی در زمان بیشتر است. در اینجا برای رفع این مشکل از الگوریتم پیشنهادی نیز بهره برده و مطابق سناریوهایی که در ادامه مطرح می‌شود این امر مرتفع گردیده است.

سناریوی اول در شبیه‌سازی سوم، تعداد کاربران ۷۰ و تامین‌کنندگان کاندید نیز ۲۵۰ در نظر گرفته شده است. بر مبنای پارامترهای مورد نظر کاربران نتیجه مطابق شکل ۱۲ بود.

همانگونه که مشاهده می‌شود با الگوریتم پیشنهادی به خروجی با میزان دقت و پاسخ بهتری دست یافته شده است (جدول ۶).

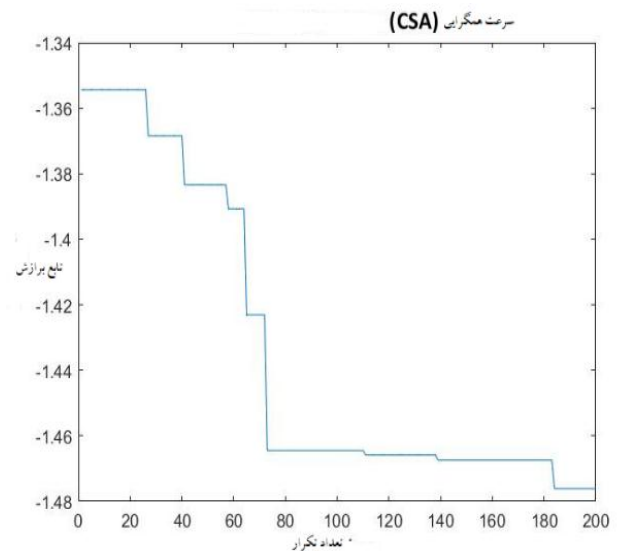
در سناریوی دوم، تعداد کاربران ۲۵۰ و تامین‌کنندگان کاندید نیز ۱۲۵۰ در نظر گرفته شد بر مبنای پارامترهای مورد نظر کاربران نتیجه مطابق شکل ۱۳ است.



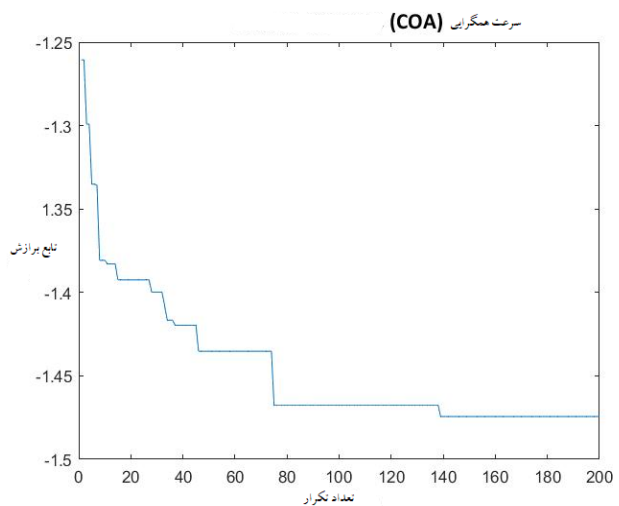
شکل ۱۲. مقایسه دقت پاسخ الگوریتم‌ها با ۲۵۰ ارائه‌دهنده سرویس کاندید

جدول ۶. مقایسه دقت پاسخ الگوریتم‌ها با ۲۵۰ فراهم‌کننده کاندید

ردیف	تعداد تامین‌کنندگان کاربران			زمان (ثانیه)
	کندید	متقاضی	کندید	
۱	۱۰	۳	۱,۶۷	۱,۸۳
۲	۲۰	۷	۲,۹۰	۳,۸۶
۳	۴۰	۱۲	۳,۳۹	۴,۲۷
۴	۶۰	۲۰	۳,۳۶	۳,۷۶
۵	۸۰	۲۵	۳,۳۷	۳,۲۷
۶	۱۰۰	۳۰	۲,۷۰	۲,۵۹
۷	۱۵۰	۴۰	۲,۹۴	۳,۶۱
۸	۲۰۰	۶۰	۲,۸۲	۳,۸۹
۹	۲۵۰	۷۰	۳,۱۵	۳,۵۳



شکل ۱۰. سرعت همگرایی CSA



شکل ۱۱. سرعت همگرایی COA

همانگونه که مشاهده می‌شود سرعت همگرایی در الگوریتم COA نسبت به دو الگوریتم دیگر بهتر بوده است، در جدول ۵ مقادیر سه الگوریتم ارائه شده است.

با به‌کارگیری چارچوب پیشنهادی و مقایسه با روش مرجع کاملا مشهود بود که روش ارائه شده نسبت به مرجع کارایی بهتری دارد و در زمان کمتر تامین‌کننده مورد نظر کاربر را پیشنهاد می‌دهد.

جدول ۵. اجرای الگوریتم COA با الگوریتم‌های GA و CSA

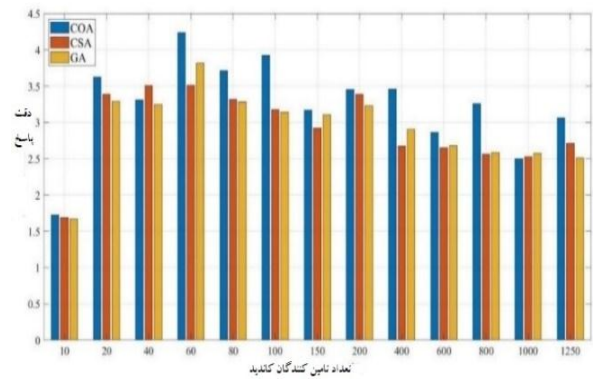
الگوریتم	تعداد کاربر	جمعیت اولیه	تعداد تکرار	همگرایی
COA	۲۰۰	۳۰	۲۰۰	۱,۴۷۴۴
CSA	۲۰۰	۳۰	۲۰۰	۱,۴۷۶۰
GA	۲۰۰	۳۰	۲۰۰	۲,۱۲۷۱

جدول ۷ بیانگر اطلاعات پارامترها بر مبنای امتیاز به فراهم‌کننده (اعتماد درونی) و اطلاعات پارامترها بر مبنای اولویت مورد نظر کاربران از لحاظ کیفیت سرویس است (اعتماد عینی). باید توجه داشت که امتیاز اعتبار کاربرانی که بازخورد خود را درخصوص تأمین‌کننده ثبت کرده بودند در میزان امتیاز فراهم‌کننده تأثیرگذار بوده است.

مطابق جدول ۸، پانزده فراهم‌کننده وجود دارد که براساس پارامترهای عینی و درونی ثبت شده و توافق نامه SLA در شخص ثالث امتیازات مذکور را در هر پارامتر دریافت کرده‌اند. با توجه به اولویت‌های ثبت شده ۳۰ کاربر، میزان دقت در پاسخ بر مبنای بازخوردهای عینی و درونی ارائه شده است.

جدول ۷. اطلاعات پارامترهای اعتماد عینی و درونی

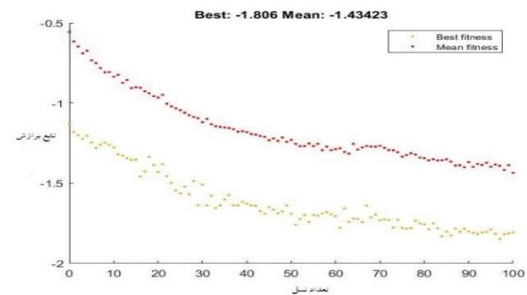
درصد شباهت	امتیاز اعتبار کاربر	امتیاز بازخورد	قابلیت اعتماد	زمان پاسخ‌دهی کاربر	کاربر
۰.۰۷	۱۲	۵	۰.۵	۱	CU1
۰.۱۵	۱۷	۲	۱	۳	CU2
۰.۰۸	۱۷	۵	۰.۹	۱	CU3
۰.۱۳	۱۴	۲	۰.۹	۱	CU4
۰.۱۳	۱۱	۵	۰.۸	۱	CU5
۰.۰۰	۱۲	۲	۰.۷	۱	CU6
۰.۱۱	۱۴	۱	۰.۸	۳	CU7
۰.۰۲	۱۰	۲	۰.۶	۳	CU8
۰.۱۹	۱۸	۵	۰.۸	۱	CU9
۰.۰۵	۱۱	۴	۰.۷	۳	CU10
۰.۰۰	۱۳	۲	۰.۶	۳	CU11
۰.۱۶	۱۶	۱	۰.۷	۱	CU12
۰.۰۰	۱۸	۱	۰.۶	۱	CU13
۸	۱۵	۴	۰.۷	۴	CU14
۰.۱۶	۱۴	۱	۰.۵	۳	CU15
۰.۱۴	۱۱	۱	۰.۵	۳	CU16
۰.۰۰	۱۰	۳	۰.۸	۲	CU17
۰.۱۹	۲۰	۵	۰.۸	۲	CU18
۰.۱۷	۱۲	۲	۰.۵	۳	CU19
۰.۲۰	۱۸	۱	۰.۸	۱	CU20
۰.۰۲	۱۸	۱	۰.۶	۲	CU21
۰.۰۴	۱۷	۳	۰.۶	۳	CU22
۰.۲۰	۱۰	۲	۰.۵	۴	CU23
۰.۰۴	۱۷	۵	۰.۹	۴	CU24
۰.۰۱	۱۰	۳	۰.۹	۳	CU25
۰.۰۰	۱۳	۳	۰.۶	۴	CU26
۰.۰۴	۲۰	۳	۰.۷	۳	CU27
۰.۱۳	۱۴	۴	۱	۴	CU28
۰.۱۸	۱۲	۳	۰.۶	۱	CU29
۰.۱۵	۱۳	۵	۰.۷	۴	CU30



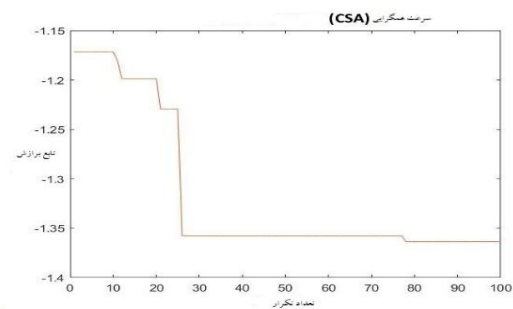
شکل ۱۳. مقایسه دقت پاسخ الگوریتم‌ها با ۱۲۵۰ فراهم‌کننده کاندید

۶-۷- سرعت همگرایی الگوریتم برای افزایش دقت پاسخ

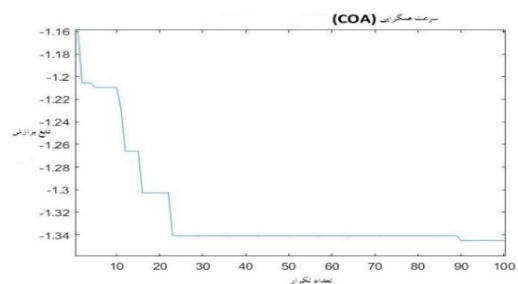
شکل ۱۶ نیز نمایانگر سرعت همگرایی بیشتر این الگوریتم نسبت به دو الگوریتم دیگر است شایان ذکر است پارامترهای تنظیم شده در این سناریو ۳۰ تأمین‌کننده کاندید، ۲۰۰ کاربر متقاضی و ۱۰۰ تکرار می‌باشد.



شکل ۱۴. سرعت همگرایی CSA



شکل ۱۵. سرعت همگرایی GA



شکل ۱۶. سرعت همگرایی COA

تأمین‌کننده سرویس ابری، لایه بخش ثالث مورد اعتماد، لایه کاربر ابری و لایه ثبت سرویس ارائه شده است. مدیریت اعتماد مبتنی بر چارچوب ارائه شده این امکان را فراهم می‌کند که کاربران ابری بدون نیاز به روشهای پیچیده بتوانند تأمین‌کننده مورد اعتماد خود را با بیشترین دقت و کمترین زمان انتخاب کنند. ارزیابی اعتماد کاربران و تأمین‌کنندگان سرویس ابری در چارچوب ارائه شده بر اساس بازخوردهای کاربران ابری و پارامترهای مولفه معیار انجام می‌شود. ارزیابی اعتماد تأمین‌کنندگان سرویس ابری مبتنی بر اطلاعات کیفیت سرویس و بازخوردهای ثبت شده کاربران در ابرهای مختلف انجام می‌شود. عامل TTP در چارچوب ارائه شده علاوه بر ارزیابی میزان اعتماد کاربر و تأمین‌کننده قادر است بهترین سرویس‌های ابری مورد اعتماد را بر اساس سلايق و اولویتهای شخصی کاربران در سریع‌ترین زمان ممکن ارائه کند. از مهمترین نوآوری‌های چارچوب مدیریت اعتماد ارائه شده در این پژوهش دسترسی در سریع‌ترین زمان ممکن به تأمین‌کننده بر مبنای اولویت و پارامترهای مورد نظر کاربر و همچنین دستیابی به بهترین نتیجه با بیشترین دقت با استفاده از الگوریتم بهینه فاخته است. در پژوهش‌هایی که تا کنون صورت گرفته بر چگونگی محاسبه اعتماد تمرکز شده است در حالی که سرعت و دقت در انجام این محاسبات از اهمیت زیادی برخوردار است. با بررسی مراجع موجود پژوهشی که به این موضوع پرداخته است راهکاری بر مبنای الگوریتم CSA ارائه شده است، ولی دقت پاسخ در آن در نظر گرفته نشده و این در حالی است که کاربران ابری برای انتخاب تأمین‌کننده علاوه بر انتخاب سرویس در اسرع وقت نیاز به دقت بالاتر در انتخاب خود دارند تا نتیجه با اولویتهای کاربر همسو و در نهایت اعتماد کاربر حاصل شود این امر خود از نقاط ضعفی بود که باید برای آن راهکاری پیشنهاد می‌شد. با توجه به نتایج بدست آمده در این پژوهش الگوریتم COA توانسته است این چالش را برطرف کند چرا که این الگوریتم با بهره بردن از پرواز لوی و خوشه‌بندی با جستجوی بیشتر در فضای مسئله و پرهیز از همگرایی زودرس پاسخی سریع‌تر نسبت به سایر الگوریتم‌ها داشته است ضمن اینکه در این پژوهش چالش دقت درانتخاب بهترین تأمین‌کننده مرتفع گردیده است. نتایج مقایسه زمان اجرا الگوریتم پیشنهادی و سرعت همگرایی آن با سایر الگوریتم‌های استفاده شده در پژوهش‌های قبلی در بخش ارزیابی عملکرد بیانگر قابلیت آن است ضمن این که دقت پاسخ الگوریتم‌ها در شرایط و تعداد فراهم‌کنندگان کاندید مختلف، با افزایش تکرار، افزایش محیط مسئله، تنوع پارامترها و اولویتهای کاربر به پاسخی بهتر نسبت به سایر الگوریتم‌های مذکور نائل گردیده است.

جدول ۸: امتیاز پارامترهای عینی و درونی تأمین‌کنندگان

تأمین‌کنندگان	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
زمان پاسخ	۳	۴	۲	۶	۹	۳	۱۰	۵	۷	۷	۱۰	۲	۵	۹	۱۰
قابلیت اعتماد (حداکثر)	۱۰	۶	۰	۵	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
قابلیت اعتماد (حداقل)	۱	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
هزینه	۶۶	۷۴	۹۹	۱۷	۵۴	۵۶	۶۲	۴۲	۲۶	۷۵	۵۶	۴۸	۶۴	۶۲	۳۶
امتیاز کاربر	۲	۴	۱	۲	۲	۴	۳	۵	۱	۱	۳	۳	۵	۴	۱
امتیاز اعتماد	۱۱	۱۷	۱۹	۱۷	۱۲	۱۰	۱۳	۱۳	۲۰	۱۵	۱۴	۱۰	۲۰	۱۹	۱۵

مقایسه پاسخ‌های سه الگوریتم بیان دقت و پاسخ بهتر الگوریتم COA نسبت به الگوریتم‌های GA و CSA به کاربران متقاضی بود. این برتری در جدول ۹ قابل مشاهده است.

به منظور بررسی صحت و کارایی چارچوب ارائه شده در دو مرحله بهینه‌سازی صورت پذیرفت. در اولین بخش با استفاده از پارامترهای عینی و درونی و بکارگیری چارچوب ارائه شده و مقایسه با روش مرجع، دستیابی به نتیجه در سریع‌ترین زمان ممکن نشان داده شد. در بخش دوم نیز با توجه به اهمیت کیفیت پاسخ با استفاده از راهکار پیشنهادی، دقت پاسخ نیز بهینه و بهترین پاسخ بر مبنای بازخوردهای عینی و درونی بدست آمد.

جدول ۹: مقایسه دقت در پاسخ به کاربران

کاربر	COA	CSA	GA
CU1	۱۳,۷۰۰۴	۵,۷۶۸۲	۲,۶۹۹۵
CU2	۱۴,۰۲۱۳	۵,۴۵۷۰	۹,۶۹۱۷
CU9	۱۳,۸۸۳۹	۹,۶۹۰۰	۱۳,۲۰۶۷
CU13	۱۰,۰۹۱۹	۹,۵۰۴۸	۳,۰۷۹۹
CU14	۱,۸۷۰۱۵	۱,۲۵۵۶	۱,۰۵۹۰
CU20	۱۴,۴۲۵۵	۱,۸۸۱۲	۳,۷۷۲۵
CU22	۱۵,۵۶۰۲	۱۴,۳۸۳۰	۹,۷۹۸۶
CU24	۱۲,۹۹۰۲	۱۲,۶۹۲۹	۲,۳۳۷۳
CU26	۱۲,۶۸۹۸	۰,۲۰۵۶	۴,۵۲۰۷
CU30	۷,۶۶۳۹	۱,۴۴۷۹	۱,۳۱۵۳

۷- بحث بر روی نتایج

در این پژوهش یک چارچوب مدیریت اعتماد با رویکرد بازخورد کاربران ابری در محیط محاسبات ابری مبتنی بر چهار لایه

۸- نتیجه گیری

آینده استفاده کند. بهینه‌سازی پیش بینی بازخوردهای مفقود در مولفه‌های شخص ثالث با استفاده از این الگوریتم امکان‌پذیر است.

مراجع

- [1] Jens Lansing, Ali Sunyaev, "Trust in Cloud Computing: Conceptual Typology and Trust Building Antecedents", ACM SIGMIS Database Systems Volume 47 Issue 2, pp 58-96, May 2016.
- [2] Huang J. Nicol, D.M., "Trust mechanisms for cloud computing" Journal of Cloud Computing Advances, System Applications volume 2, Article number: 9, pp 1-14, April 2013.
- [3] F. Azzedin and M. Maheswaran, "Towards Trust-Aware Resource Management in Grid Computing Systems," 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID02), Berlin, Germany, pp. 452-452, May 2002.
- [4] Abhishek Kesarwani, Pabitra Mohan Khilar, "Development of trust based access control models using fuzzy logic in cloud computing" Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 5, 2022, pp1958-1967, November 2022.
- [5] P. Gupta, M. K. Goyal and P. Kumar, "Trust and reliability based load balancing algorithm for cloud IaaS," 2013 3rd IEEE International Advance Computing Conference (IACC), Ghaziabad, India, pp. 65-6, 2013.
- [6] Mohammad Javad Salehi¹, Mehrdad Ashtiani² and Behrouz Minaei Bidgoli Cloud Service Selection based on the Credibility persistency of Users' Feedbacks, Volume 7, Issue 1 Pages 29-41, December 2018.
- [7] Tabassum N. Mujawar, Lokesh B. Bhajantri, "Behavior and feedback-based trust computation in cloud environment", Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 8, PartA, Pages 4956-4967, September 2022.
- [8] Aghaee Ghazvini, G. Mohsenzadeh, M., Nasiri, R. et al. "A new multi-level trust management framework (MLTM) for solving the invalidity and sparse problems of user feedback ratings in cloud environment". J Supercomputer 77, pp2326-2354, March 2021.
- [9] H. Hassan, A. I. El-Desouky, A. Ibrahim, E. -S. M. El-Kenawy and R. Arnous, "Enhanced QoS-Based Model for Trust Assessment in Cloud Computing Environment," in IEEE Access, vol. 8, pp. 43752-43763, 2020.
- [10] Aghaee Ghazvini, G. Mohsenzadeh, M., Nasiri, R., Masoud Rahmani, A. MMLT: A mutual multilevel trust framework based on trusted third parties in multi cloud environments. Software Pract. Exper, pp 125, January 2022.
- [11] Noor, T.H., Sheng, Q.Z. "Trust as a Service: A Framework for Trust Management in Cloud Environments" In: Bouguettaya, A., Hauswirth, M., Liu, L. (eds) Web Information System Engineering WISE 2011. Lecture Notes in Computer Science, vol 6997. Springer, Berlin, Heidelberg. pp315-317, 2011.
- [12] Shakeel Ahmad, Bashir Ahmad, Sheikh Muhammad Saqib and Rashid Muhammad Khattak, "Trust Model : Cloud's Provider and Cloud's User", International Journal of Advanced Science and Technology Vol.44, July 2012.
- [13] Saurabh Kumar Garg, Steve Versteeg, Rajkumar Buyya, A framework for ranking of cloud computing services, Future Generation Computer Systems, Volume 29, Issue 4, Pages 1012-1023, June 2013.
- [14] Muhammad Kashif Naseer, Sohail Jabbar, Dr. Irfan Zafar "ANovel Trust Model for Selection of Cloud Service Provider", (IEEE 2014), 2014.
- [15] Paul Manuel, 2015. "A trust model of cloud computing based on Quality of Service," Annals of Operations Research, Springer, vol. 233(1), pages 281-292, April 2013.

این مقاله، مسئله بهینه‌سازی انتخاب تأمین‌کننده ابری را به مدل ریاضی چندجمله‌ای برای یافتن راه‌حل بهینه تبدیل کرد. الگوریتم پیشنهادی اعتبار سنجی شده و با الگوریتم‌های دیگر مانند الگوریتم ژنتیک و جستجوی فاخته مقایسه شده است. بهینه‌سازی نتایج نشان داد که الگوریتم پیشنهادی از نظر زمان و دقت در پاسخ نسبت به سایر الگوریتم‌ها برتری دارد و نشان می‌دهد که COA می‌تواند با این نوع مشکل مقابله کند. بنابراین، COA برای مشکل انتخاب سرویس در یک محیط محاسبات ابری در مقایسه با بسیاری از الگوریتم‌های فراابتکاری دیگر، عمومی‌تر و قوی‌تر است.

۸-۱- پیشنهادات آتی

به عنوان بخشی از تحقیقات آینده، می‌توان برای توسعه چارچوب پیشنهادی بر روی موارد ذیل تمرکز کرد. با الهام از رفتار پرند فاخته در مخزن ذخیره بازخورد مدیریت اعتماد، برای بهینه‌سازی سیستم بازخورد و رتبه‌بندی می‌توان استفاده کرد. الگوریتم فاخته همچنین می‌تواند برای شناسایی و حذف رتبه‌های بازخورد جعلی استفاده شود. این را می‌توان با مقایسه رتبه‌های بازخورد هر کاربر با همسایگانش انجام داد. اگر رتبه‌بندی بازخورد یک کاربر به طور قابل توجهی با همسایگانش متفاوت باشد، ممکن است نشان دهد که کاربر بازخورد جعلی ارائه می‌کند. در این صورت می‌توان رتبه‌های بازخورد کاربر را از سیستم حذف کرد. به طور کلی، الگوریتم فاخته می‌تواند ابزار مفیدی برای بهینه‌سازی و بهبود سیستم پایه بازخورد مدیریت اعتماد با شناسایی و حذف بازخورد جعلی و بهبود دقت رتبه‌بندی بازخورد باشد. الگوریتم فاخته می‌تواند برای بهینه‌سازی امتیازات اعتماد و رتبه‌بندی کاربران در یک محیط ابری با به حداقل رساندن خطا بین امتیازات اعتماد و رتبه‌بندی‌ها واقعی و امتیازات و رتبه‌بندی‌های اعتماد پیش بینی شده استفاده شود. این می‌تواند به بهبود دقت سیستم مدیریت اعتماد و رتبه‌بندی کمک کند و منجر به تصمیم‌گیری بهتر و افزایش رضایت کاربر شود. از دیگر موارد قابل بهبود با این الگوریتم می‌توان به نحوه به روز رسانی اطلاعات پایگاه داده اعتماد TTP اشاره نمود و این به روز رسانی در صورتی انجام گردد که تناسب بهترین لانه فاخته بالاتر از یک آستانه خاص باشد. این کار برای اطمینان از اینکه فقط داده‌های قابل اعتماد به پایگاه داده TTP اضافه می‌شوند صورت می‌پذیرد که به نوبه خود باعث بهبود قابلیت اعتماد کلی سیستم می‌شود. TTP می‌تواند از امتیازات و رتبه‌بندی‌های اعتماد بهترین لانه فاخته برای به روز رسانی پایگاه داده اعتماد خود استفاده کند و از آن برای ارزیابی‌های

- [20] Rutravigneshwaran, P. and Anitha, G. Security Model to Mitigate Black Hole Attack on Internet of Battlefield Things (IoBT) Using Trust and K-Means Clustering Algorithm. *Int. J. Comput. Networks Appl*, 10(1), p.95, January – February 2023.
- [21] J.Gibson, R. Rondeau, D. Eveleigh and Q. Tan, "Benefits and challenges of three cloud computing service models," 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN), Sao Carlos, Brazil, 2012, pp. 198-205.
- [22] Sana, Kouchi & Nacer, Hassina.. Service Selection in Cloud Computing Environment by Using Cuckoo Search , pp219227, January 2022.
- [16] Talal H. Noor, Quan Z. Sheng, Athman Bouguettaya ,Trust Management in Cloud Services, Springer,pp1-119, September 2014.
- [17] Ricardo Perez-Truglia, Markets, trust and cultural biases: evidence from eBay, *Journal of Behavioral and Experimental Economics*, Volume 72, pp 17-27, February 2018.
- [18] Rajabioun, Ramin. (2011). Cuckoo Optimization Algorithm. *Applied Soft Computing*. 11. pp5508–5518, December 2011.
- [19] Elnahary, M., Hamed, A., El-Sayed1, H. (2022). 'Task Scheduling Optimization in cloud computing by Cuckoo Search Algorithm', *Sohag Journal of Sciences*, 7(3), pp. 29-37 ,September 2022.

Energy Efficient Target (Rhynchophorus Ferrugineus) Tracking in Wireless Sensor Network Using the Cat Swarm Optimization Algorithm and Fuzzy Logic

Shayesteh Tabatabaei^{1*}

¹Faculty of Engineering, Saravan University, Saravan, Iran

Received: 27 October 2023, Revised: 22 December 2023, Accepted: 28 February 2024

Paper type: Research

Abstract

The *Rhynchophorus ferrugineus* is a major pest that serves as a carrier for bacterial and fungal diseases, causing significant damage to palm plantations when observed on farms. Nowadays, advancements in wireless communication environments have made it possible to develop low-cost, energy-efficient, multi-functional, and short-range sensor nodes for tracking this pest in palm plantations. In existing target tracking algorithms, the probability of losing the target increases with its speed. Therefore, this paper proposes a new method for target tracking that reduces the likelihood of losing the target. Additionally, considering the energy constraints of battery-powered sensor nodes, we need a scheduling mechanism for their sleep and wake-up cycles to enhance the network's lifespan. To improve energy consumption, this paper utilizes a time scheduling approach to adjust the sleep and wakeup periods of nodes using the Cat Swarm Optimization algorithm and Fuzzy Logic optimization. By simulating the proposed method and comparing it with the Tracking-45-Degree-vectors method in the Opnet simulator, it can be observed that the proposed protocol performs significantly better. Specifically, the end-to-end delay rate improves by 27.02%, the media access delay rate improves by 2.01%, the throughput rate improves by 0.62%, the signal-to-noise ratio improves by 3.28%, and the average battery energy consumption improves by 8.77% compared to the Tracking-45-Degree-vectors protocol. It is worth mentioning that the proposed algorithm has been simulated and tested for a single target scenario.

Keywords: Clustering, Energy Consumption, WSN, Cat Swarm Optimization algorithm, Fuzzy Logic, Target tracking, *Rhynchophorus ferrugineus*

* Corresponding Author's email: shtabatabaei@yahoo.com

ردیابی هدف (سوسک سرخرطومی خرما) انرژی کارا در شبکه حسگر بی‌سیم با استفاده از الگوریتم ازدحام گربه و منطق فازی

شایسته طباطبائی*

^۱ دانشکده فنی مهندسی، دانشگاه سراوان، سراوان، ایران

تاریخ دریافت: ۱۴۰۲/۰۸/۰۵ تاریخ بازبینی: ۱۴۰۲/۱۰/۰۱ تاریخ پذیرش: ۱۴۰۲/۱۲/۰۹

نوع مقاله: پژوهشی

چکیده

آفت سوسک سرخرطومی حنایی عامل ورود بیماری‌های باکتریایی و قارچی به نخل می‌باشد که در صورت مشاهده در مزارع، خسارت سنگینی به نخلستان‌ها وارد می‌کند. امروزه تحول در محیط ارتباطات بی‌سیم امکان توسعه گره‌های حسگر کم هزینه، کم مصرف، چند عملکردی و کوتاه برد، جهت ردیابی این آفت در نخلستان را فراهم آورده است. در الگوریتم‌های ردیابی هدف موجود، با افزایش سرعت هدف، احتمال از دست دادن هدف نیز افزایش می‌یابد. بر این اساس در مقاله حاضر روش جدیدی برای ردیابی هدف پیشنهاد می‌شود که از دست دادن هدف را کاهش می‌دهد. از طرفی با توجه به محدودیت انرژی سطح باتری در گره‌های حسگر، نیازمند برنامه زمانی برای دوره خواب و بیداری حسگرها در راستای افزایش طول عمر شبکه هستیم. به منظور بهبود مصرف انرژی در این مقاله از رویکرد برنامه‌ریزی زمانی جهت تنظیم دوره خواب و بیداری گره‌ها با استفاده از الگوریتم بهینه‌سازی ازدحام گربه و منطق فازی استفاده شده است. از شبیه‌سازی روش پیشنهادی و مقایسه آن با روش Tracking-45-Degree-vectors در شبیه‌ساز Opnet می‌توان دریافت که پروتکل پیشنهادی عملکرد بسیار بهتری دارد، بطوریکه نرخ تأخیر انتها به انتها به میزان ۲۷/۰۲ درصد، نرخ تأخیر دسترسی به رسانه به میزان ۲/۰۱ درصد، نرخ گذردهی به میزان ۰/۶۲ درصد، نسبت سیگنال به نویز به میزان ۳/۲۸ درصد و میانگین انرژی مصرفی باتری به میزان ۸/۲۷۷ درصد نسبت به پروتکل Tracking-45-Degree-vectors بهبود یافته است. لازم به ذکر است الگوریتم پیشنهادی برای یک هدف شبیه‌سازی و تست شده است.

کلیدواژگان: خوشه‌بندی، مصرف انرژی، شبکه‌های حسگر بی‌سیم، الگوریتم بهینه‌سازی ازدحام گربه‌ها، منطق فازی، ردیابی هدف متحرک، سوسک سرخرطومی خرما.

* رایانامه نویسنده مسؤول: shtabatabaey@yahoo.com

۱- مقدمه

سرخرطومی خرما در محیط نخلستان است که توسط حسگرهای بی‌سیم با استفاده از الگوریتم ازدحام گریه‌ها و منطق فازی ردیابی می‌شود. از آنجا که مسأله انرژی برای بقا و افزایش طول عمر شبکه در یک شبکه حسگر بی‌سیم بسیار مهم است؛ بحث عمده در این زمینه بر روی ارائه الگوریتم‌هایی برای پیش‌بینی مکان بعدی هدف و ایجاد راهکاری برای خواب و بیدار کردن گره‌ها و مدیریت آنها می‌باشد. رویکرد اساسی در این موضوع، از دست ندادن هدف در خلال مسیر حرکتش خواهد بود.

اهداف اصلی مقاله شامل موارد زیر است:

- در مطالعات انجام شده قبلی، اندازه‌گیری‌ها در سرعت‌های حدود ۲۰ متر بر ثانیه انجام شده است. در این مطالعه، ردیابی هدف متحرک با سرعت آزمایش ۳۰ تا ۲۰۰ متر بر ثانیه با موفقیت انجام شد.
- یک الگوریتم جدید بر اساس الگوریتم ازدحام گریه و منطق فازی برای ردیابی هدف متحرک پیشنهاد شده است. ایده جدیدی که در الگوریتم ردیابی پیشنهادی مقاله حاضر تأکید شده و در سایر مقالات دیده نمی‌شود، به حداقل رساندن خطر از دست دادن ردیابی هدف با استفاده از الگوریتم هوشمند است.
- با مقایسه الگوریتم پیشنهادی و رویکرد موجود در مقاله [۲] (Tracking-45-Degree-vectors) از نتایج شبیه‌سازی، مشاهده شد که رویکرد پیشنهادی نتایج بسیار بهتری را در ردیابی هدف متحرک ارائه می‌دهد و احتمال از دست دادن ردیابی هدف را به میزان قابل توجهی کاهش می‌دهد.
- با کاهش میزان از دست دادن هدف با الگوریتمی که پیشنهاد شده، نیاز به مکانیسم بازیابی برای یافتن مجدد هدف نیز کاهش می‌یابد. هرچه از روشهای بازیابی کمتری استفاده شود، مصرف انرژی کمتر می‌شود که باعث افزایش طول عمر شبکه می‌شود.

ساختار مقاله حاضر به صورت زیر است:

در بخش دوم پیشینه تحقیق بیان می‌شود که به موضوع و مطالعات انجام یافته قبلی در زمینه بهبود مصرف انرژی برای کاربرد ردیابی اهداف متحرک در شبکه حسگر بی‌سیم می‌پردازد، در بخش سوم الگوریتم ازدحام گریه و منطق فازی که برای ردیابی هدف متحرک استفاده می‌شوند، بررسی خواهند شد، نتایج حاصل از شبیه‌سازی الگوریتم پیشنهادی در بخش چهارم مطرح می‌شود. در نهایت بخش پنجم به نتیجه‌گیری می‌پردازد.

سوسک سرخرطومی حنایی خرما از مهم‌ترین آفات نخلستان در قاره‌های آسیا، آفریقا و اروپا می‌باشد که برای اولین بار در سال ۱۳۶۹ در شهرستان سراوان خسارت آن بر روی درختان خرما گزارش شد. آفت سوسک سرخرطومی حنایی عامل ورود بیماری‌های باکتریایی و قارچی به نخل می‌باشد که در صورت مشاهده در مزارع، خسارت سنگینی به نخلستان‌ها وارد می‌کند. سوسک سرخرطومی حنایی برای زندگی، لانه‌سازی و پرورش لاروها معمولاً از تنه درخت نخل به عنوان یک بستر مناسب جهت رشد و تکثیر خود استفاده می‌کند و با ایجاد حفرات و راهروهای متعدد فضای کافی برای رشد خود به وجود می‌آورد، در این حال طی روند زندگی سوسک سرخرطومی در تنه درخت خرما به تدریج میزان رشد و باردهی درخت تحت تاثیر قرار می‌گیرد و نهایتاً در صورت عدم مبارزه و تیمار درخت منجر به مرگ نخل خرما می‌شود. معمولاً خسارت این حشره به سادگی قابل مشاهده نیست و تنها زمانی می‌توان متوجه وجود سوسک حنایی خرما شد که برگ‌ها پژمرده می‌شوند یا درخت می‌افتد. فعالیت این آفت در فصل گرم سال بوده که این مساله بر افزایش سختی، هزینه انجام عملیات ردیابی و حذف کانون‌های آلودگی می‌افزاید. ردیابی براساس پایش‌های چشمی، مشاهده علائم و استفاده از فرمون‌های جنسی مخصوص این سوسک صورت می‌گیرد. اگر چه نصب تله‌های فرمونی باعث جذب حشرات و کاهش جمعیت آفت‌ها می‌شود ولی هزینه‌بر است.

امروزه تحول در محیط ارتباطات بی‌سیم امکان توسعه گره‌های حسگر کم هزینه، کم مصرف، چند عملکردی و کوتاه برد را جهت ردیابی و تشخیص وجود لاروهای تغذیه‌کننده در تنه درخت فراهم آورده است. این گره‌های حسگر کوچک قادر به سنجش، پردازش داده‌ها و برقراری ارتباط هستند که در شبکه‌های حسگر با یکدیگر همکاری می‌کنند و می‌توانند مطابق با قابلیت‌های خود بر روی داده‌ها کار کنند. شبکه‌های حسگر دارای انواع مختلفی از حسگرها نظیر حسگرهای لرزه‌ای، حرارتی، مغناطیسی و بصری می‌باشند که می‌توانند تغییرات شرایط محیطی مانند دما، رطوبت، فشار، صدا، تحرک و نور در مناطق نظامی، محیطی، بهداشتی، خانگی و سایر مناطق تجاری را کنترل کنند. همچنین شبکه‌های حسگر در کاربردهای زیست محیطی برای نظارت بر تشخیص بیولوژیکی و حرکات حیوانات، نظارت بر آتش سوزی جنگل‌ها و سیل استفاده می‌شوند [۱]. در این مقاله، روی کاربرد ردیابی هدف در شبکه‌های حسگر بی‌سیم تمرکز شده است. که منظور از هدف سوسک

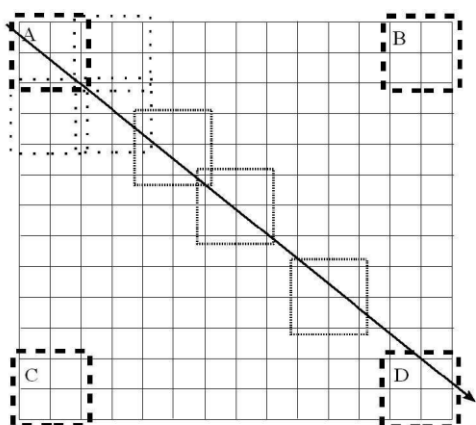
۲- کارهای مرتبط

برای مشارکت در فرآیند انتخاب مد نظر قرار می گیرند.

در [۶] یک روش پیش‌بینی هدف مبتنی بر خوشه پیشنهاد شده است. در این روش، همه‌ی گره‌های حسگر فعال اطلاعات خود را به رهبر خوشه ارسال می‌کنند. رهبر با استفاده از الگوریتم مکانیابی مرکز ثقل^۱، اطلاعات دریافتی را تجمیع کرده و با استفاده از الگوریتم پیش‌بینی حداقل مربعات بازگشتی در مورد موقعیت هدف تصمیم‌گیری می‌کند. نتایج پیش‌بینی برای انتخاب رهبر جدید خوشه و گره‌های فعال به کار گرفته می‌شود.

در [۷] یک الگوریتم جدید برای ردیابی هدف با توجه به همکاری و اطلاعات انرژی معرفی شده است که ترکیبی از فیلتر کالمن و الگوریتم مشارکتی مبتنی بر انرژی است. این الگوریتم فیلتر توزیع شده کالمن را فقط در گره‌های سرخوشه اجرا می‌کند. هنگام تغییر رهبری، گره سرخوشه بعدی از حالت پیش‌بینی شده گره سرخوشه قبلی برای مقدار اولیه فیلتر کالمن استفاده می‌کند. گره‌های جمع‌آوری داده‌ها با استفاده از الگوریتم مشارکتی، حسگرهای فعال را طوری انتخاب می‌کنند که انرژی کل شبکه کمینه شود.

در [۸] یک الگوریتم ردیابی هدف انرژی کارآمد مبتنی بر سلول پیشنهاد شده است. نویسندگان منطقه حسگری را به مکان‌های جغرافیایی، به نام سلول‌ها تقسیم نمودند و حرکت هدف از یک سلول به دیگر را به صورت شکل ۱ پیش‌بینی می‌کنند. آنها از یک رویه مبتنی بر انرژی استفاده می‌کنند که با اندازه‌گیری انرژی از چندین گره (۴ یا بیشتر) در یک لحظه خاص، مکان هدف را تخمین می‌زنند.



شکل ۱. نمایش شماتیک ردیابی تک هدف [۵]

ردیابی سوسک سرخرطومی حنایی خرما معمولاً با استفاده از پایش چشمی، مشاهده علائم ظاهری، فرمون‌های جنسی، کنترل شیمیایی و تجهیزات الکترونیکی انجام می‌شود. ردیابی مشاهده‌ای و بر اساس علائم ظاهری (خروج شیرابه قهوه‌ای، وجود مواد فیبری، بوی پوسیدگی و تخمیر، سوراخ‌های روی تنه و صدای خراپی لاروها در داخل تنه) تا مرحله خشک شدن درخت قابل تشخیص نیست و نیازمند حضور ردیابان ماهر در طول سال است. ردیابی با استفاده از تله‌های فرمونی منجر به جذب حشرات و کاهش جمعیت سوسک سرخرطومی حنایی خرما می‌شود، اما هزینه‌های آن بالاست. استفاده از روش‌های شیمیایی و مواد سمی مانند فسفید آلومینیوم در درختان که هنوز به طور کامل خشک نشده‌اند نیز به محیط زیست آسیب می‌رساند. به تازگی برای شناسایی لاروهای موجود در تنه درخت، از تجهیزات الکترونیکی یا سگهای آموزش دیده استفاده می‌کنند که صوت و بوی آنها را تشخیص می‌دهند [۳].

فناوری‌های مدرن به بررسی زمینه‌های مختلف آفات، مانند استفاده از برجسب‌های شیمیایی برای ردیابی حرکات حشرات و تشخیص اکو صدا از جنبش لارو متمرکز شده‌اند [۴]. با این حال، این تکنیک‌ها معمولاً هزینه‌های بالایی داشته و برای کشاورزان به عنوان یک راهکار مقرون به صرفه شناخته نمی‌شوند. یکی از فناوری‌های مدرن در شناسایی آفت سوسک سرخرطومی استفاده از حسگرهای بی سیم است. مشکل اصلی حسگرها منبع انرژی محدودشان است.

در [۲] به منظور بهبود مصرف انرژی در کاربردهای شناسایی هدف در شبکه‌های حسگر بی سیم، از داده کاوی و یادگیری Q به همراه پارامترهای میزان انرژی گره و تعداد همسایگان گره استفاده شده است که یک آتاماتای یادگیر سلولی را در نظر می‌گیرد. نتایج شبیه‌سازی نشان می‌دهد که استفاده از داده کاوی و یادگیری Q می‌تواند مصرف انرژی و تاخیر انتقال داده را بهبود بخشد و تعادل بار را در مقایسه با استاندارد IEEE 802.15.4 بهبود دهد.

در [۵]، یک رویکرد خوشه‌بندی مبتنی بر همکاری برای ردیابی هدف در شبکه‌های بی سیم متراکم پیشنهاد شده است. در روش پیشنهادی از یک الگوریتم مکان‌یابی مبتنی بر انرژی برای تخمین موقعیت هدف استفاده می‌شود. گره‌های حسگر که در ترکیب اطلاعات هدف شرکت می‌کنند، بر اساس سطح سیگنال دریافتی، میزان انرژی، محدوده حساسیت و اطلاعات شناسایی درباره هدف،

¹ Centroid

اساس الگوریتم ژنتیک پیشنهاد شده است که از روش‌های برش تطبیقی^۳ و انتخاب تورنومنت^۴ باینری به منظور افزایش طول عمر شبکه استفاده می‌کند. روش پیشنهادی، از ترکیب پارامترهای متعدد مانند انرژی باقیمانده، انرژی اولیه، فاصله تا سینک، تعداد همسایگان یک گره، پارامتر تعادل بار برای انتخاب سرخوشه در یک شبکه حسگر بی‌سیم ناهمگن استفاده می‌کند. نتایج شبیه‌سازی نشان می‌دهد که پروتکل پیشنهادی از نظر انرژی باقی‌مانده، طول عمر شبکه و توان عملیاتی بهتر عمل می‌کند.

در [۱۵] یک روش مسیریابی مبتنی بر خوشه‌بندی پیشنهاد شده است که از الگوریتم بهینه‌سازی گروه میگوها^۵ برای انتخاب گره-های سرخوشه بهره می‌برد. نتایج شبیه‌سازی حاکی از آن است که روش پیشنهادی نسبت به بهینه‌سازی ازدحام ذرات و جستجوی فاخته از نظر طول عمر شبکه عمل بهتر عمل می‌کند. ایراد این روش این است که فقط از معیار انرژی سطح باتری را برای انتخاب سرخوشه‌ها استفاده می‌کند.

در [۱۶] یک روش خوشه‌بندی برای شبکه حسگر بی‌سیم پیشنهاد شده است. روش پیشنهادی از ادغام مفاهیم محاسبات کوانتومی با TSA^۶ سنتی مشتق شده است. ارزیابی عملکرد روش پیشنهادی نشان می‌دهد که مدل ارائه‌شده، طول عمر را بهبود بخشیده، مصرف انرژی را کاهش داده و نسبت به روش‌های دیگر تأخیر کمتری دارد.

در [۱۷] از ترکیب الگوریتم‌های بهینه‌سازی کلونی پروانه^۷ و کلونی مورچه^۸ به همراه گره سینک ثابت و سینک متحرک، برای خوشه‌بندی در شبکه‌های حسگر استفاده شده است. الگوریتم بهینه‌سازی کلونی پروانه سرخوشه‌های بهینه را تعیین می‌کند و الگوریتم بهینه‌سازی کلونی مورچه مسیریابی آگاه از انرژی را انجام می‌دهد، که منجر به کاهش مصرف انرژی و به حداکثر رساندن طول عمر شبکه می‌شود. همچنین در روش پیشنهادی از سینک متحرک برای حل مشکل مسئله نقطه داغ و افزایش طول عمر شبکه استفاده می‌شود. نتایج شبیه‌سازی نشان می‌دهد که الگوریتم ترکیبی کلونی پروانه و کلونی مورچه به همراه گره سینک ساکن عملکردی بهتری از نظر انرژی باقیمانده گره‌ها نسبت به الگوریتم‌های CRWO، ERP و IHSbeer دارد.

در [۹] از روش فیلتر ذرات توزیع شده برای ردیابی هدف در شبکه‌های حسگر استفاده شده است. در روش پیشنهادی با استفاده از ماشین بردار پشتیبان، ذرات فشرده می‌شوند و از تکنیک‌های میانگین‌گیری توزیع شده برای ترکیب اطلاعات حسگرها استفاده می‌شود. این روش مقیاس پذیر است چرا که داده‌ها به سادگی بین حسگرهای مجاور به اشتراک گذاشته می‌شوند.

در [۱۰] از ترکیب الگوریتم فیلتر ذرات (PF) و ماشین بردار (SVM) به منظور ردیابی دقیق‌تر هدف و کاهش مصرف انرژی در حسگرها استفاده می‌کند. بکارگیری PF، امکان پیگیری دقیق‌تر موقعیت هدف را فراهم می‌کند و استفاده از SVM، قابلیت دسته‌بندی دقیق‌تر داده‌ها را بهبود می‌بخشد. نتایج شبیه‌سازی نشان می‌دهد که روش پیشنهادی دقت و کارایی پیگیری هدف را افزایش و مصرف انرژی را کاهش می‌دهد.

در [۱۱] یک روش استقرار تطبیقی برای بهینه‌سازی پویای عملکرد ردیابی ارائه شده است. همچنین، یک مدل شبکه واقعی ایجاد شده که ویژگی‌های تصادفی مرتبط با سنجش، تصادفی بودن قدرت سیگنال دریافتی و همبستگی بین قدرت سیگنال دریافتی را نشان می‌دهد. نتایج ارزیابی نشان می‌دهد که روش پیشنهادی بازدهی بالایی دارد.

در [۱۲] یک رویکرد زمان‌بندی تحمل‌پذیر خطا به نام FTSS^۱ با در نظر گرفتن معیارهای انرژی باقیمانده، پوشش حسگر و پوشش همپوشانی ارائه شده است تا احتمال از دست دادن هدف را برای کاربردهای ردیابی هدف به حداقل برساند. علاوه بر این، از یک الگوریتم گرگ خاکستری باینری^۲ بهبود یافته در فرآیند بهینه‌سازی استفاده شده تا به سرعت به همگرایی برسد. نتایج شبیه‌سازی نشان داده‌اند که FTSS نسبت به الگوریتم‌های موجود احتمال از دست دادن هدف را کاهش می‌دهد و طول عمر شبکه را افزایش می‌دهد.

در [۱۳] یک طرح ردیابی مبتنی بر شبکه ساده، که از یک مدل رگرسیون برای یادگیری همبستگی زمانی بین نمونه‌های متوالی یک مسیر هدف استفاده می‌کند پیشنهاد شده است. نتایج شبیه‌سازی نشان می‌دهد که الگوریتم پیشنهادی می‌تواند تا ۷۰ درصد انرژی شبکه را ذخیره کند.

در [۱۴]، یک روش خوشه‌بندی در شبکه‌های حسگر بی‌سیم بر

³ adaptive crossover

⁴ Tournament selection

⁵ Krill Herd Algorithm

⁶ Tunicate Swarm Algorithm

⁷ Butterfly Optimization Algorithm

⁸ Ant Colony Optimization Algorithm

¹ Fault-Tolerant Sensor Scheduling

² Binary Grey Wolf Optimizer

شود هر گربه دارای موقعیتی است که دارای M بعد است در کنار این موقعیت هر گربه دارای یک سرعت برای هر بعد می باشد. در این روش یک مقدار برازندگی که نشان دهنده میزان برازندگی آن گربه است تعریف می شود که توسط منطق فازی بدست می آید. علاوه بر موارد ذکر شده هر گربه دارای یک نشانه پرچم است که برای شناسایی اینکه گربه در حالت ردیابی و یا جستجو است بکار می رود.

همانطور که عنوان شد جهت محاسبه برازندگی هر گربه از منطق فازی استفاده می شود. منطق فازی [۲۲] استفاده شده سعی در مدل سازی تصمیم گیری ها دارد و قادر است نتایج دقیقی بر اساس اطلاعات نامشخص و مبهم ارائه دهد. محبوب ترین مدل های استنتاج فازی، مدانی و سوگنو هستند. در این مقاله، از مدل فازی سوگنو برای انتخاب برازندگی هر گربه استفاده شده است.

در طراحی مدل فازی، دو متغیر به عنوان ورودی فازی در نظر گرفته می شود: انرژی باقیمانده، فاصله تا سینک. انرژی باقیمانده گره (اولین ورودی) مهمترین متغیری است که مستقیماً بر طول عمر شبکه تأثیر می گذارد. هر چه انرژی باقیمانده بیشتر باشد، گره برازندگی بالاتری دارد. فاصله تا سینک (ورودی دوم) فاصله هر گربه را تعیین می کند هر چقدر فاصله تا سینک کمتر باشد میزان برازندگی افزایش می یابد. متغیرهای ورودی فازی بمنظور اجتناب از انتخاب گره های کم انرژی (ورودی اول)، به حداقل رساندن کل انرژی مصرفی گره های سرخوشه (ورودی دوم) با انتخاب گره های نزدیک به سینک در نظر گرفته می شوند. یک قانون فازی سوگنو با دو ورودی x_1 و x_2 ، و یک خروجی y را می توان به صورت «اگر $x_1=a$ و $x_2=b$ ، آنگاه $y=f(x_1, x_2)$ » نشان داد، که در آن a و b تابع عضویت ورودی ها و y مقدار برازندگی یا خروجی هستند.

۳- روش پیشنهادی

در روش پیشنهادی از الگوریتم ازدحام گربه ها به همراه منطق فازی به منظور ردیابی هدف در شبکه های حسگر بی سیم استفاده می شود. چهار عامل اصلی در حالت جستجوی گربه تعریف می شود که شامل:

۱. جستجوی منبع حافظه (SMP^4): این عامل برای تعریف منبع حافظه جستجو برای هر گربه تعریف می شود که موقعیت های جستجو شده توسط گربه را نشان می دهد. گربه یکی از موقعیت هایی را که در حافظه اش می باشد طبق قواعدی

در [۱۸] به منظور کاهش مصرف انرژی در شبکه های حسگر از یک روش خوشه بندی تحمل پذیر خطا با قابلیت مسیریابی آگاه از انرژی استفاده شده است. روش پیشنهادی شامل سه فاز اصلی یعنی خوشه بندی مبتنی بر شعله (MFO) برای انتخاب سرخوشه و ایجاد خوشه ها، فاز تحمل پذیری خطا برای افزایش طول عمر گره های شبکه و فاز مسیریابی مبتنی بر عنکبوت اجتماعی (SSO) برای انتخاب بهینه مسیره می باشد. نتایج شبیه سازی نشان دهنده افزایش طول عمر بیشتر شبکه در روش پیشنهادی نسبت به روش های EAFTC-RIS، FUCCAR-GWSO، FBECs، MLSLEEP و خوشه بندی مبتنی بر کلونی زنبور است.

در [۱۹] یک روش جدید خوشه بندی و مسیریابی چند گامه بهینه پیشنهاد شده است. در این روش در سطح اول، سرخوشه ها با استفاده از الگوریتم ژنتیک انتخاب می شوند و در سطح دوم مسیریابی چند گامه در بین سرخوشه ها در نظر گرفته می شود. نتایج شبیه سازی حاکی از عملکرد بهتر طرح پیشنهادی از نظر افزایش طول عمر شبکه است.

در [۲۰] یک روش مسیریابی مبتنی بر خوشه بندی با هدف افزایش طول عمر شبکه پیشنهاد شده است. این تکنیک مشتمل بر دو بخش، انتخاب سرخوشه بهینه که از طریق الگوریتم جدید میدان الکتریکی مصنوعی ML-AEFA^۱ و با در نظر گرفتن معیارهای انرژی، درجه گره، فاصله بین گره های حسگر، فاصله بین CH و ایستگاه پایه (BS) و زمان مرگ گره انجام می شود و مرحله انتقال داده ها که توسط الگوریتم بهینه سازی گرگ خاکستری سفارشی (CGWO)^۲ انجام می شود. نتایج شبیه سازی نشان داد که طول عمر شبکه روش پیشنهادی نسبت به MSA، GWO، AEFA، BOA + ACO افزایش یافته است.

در مقاله حاضر از الگوریتم ازدحام گربه ها به همراه منطق فازی به منظور ردیابی هدف در شبکه های حسگر بی سیم استفاده می شود تا مصرف انرژی را بهبود بخشد.

الگوریتم بهینه سازی گربه ها^۳ توسط شوچوان چو و همکارانش [۲۱] ارائه شده است. در بهینه سازی ازدحام گربه ها دو رفتار اصلی گربه ها با دو زیر مدل به نام حالت ردیابی و جستجو مدل می شوند. در این الگوریتم مکان گربه ها گویای یک پاسخ است. در این بهینه سازی ابتدا تصمیم گرفته می شود که از چند گربه استفاده

¹ Moth Levy adopted Artificial Electric Field Algorithm

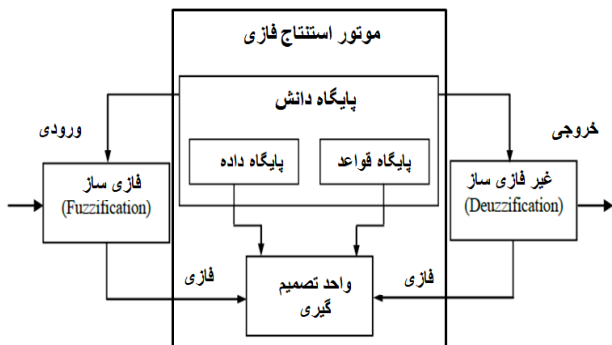
² Customized Grey Wolf Optimization

³ Cat Swarm Optimization

⁴ Seeking Memory Pool(SMP)

انتخاب می‌کند.

n بر اساس دو متغیر ورودی است.



شکل ۲. بلوک دیاگرام سیستم منطق فازی

فازی‌ساز: فازی‌ساز با استفاده از توابع عضویت، ورودی‌ها را به متغیرهای فازی زبانی تبدیل می‌کند. توابع عضویت برای متغیرهای ورودی در شکل ۳ و ۴ نمایش داده شده است. مجموعه‌های فازی استفاده شده برای فازی‌سازی هر متغیر ورودی را به سه تابع عضویت کم، متوسط و زیاد تقسیم می‌کند. در این فاز، مجموعه‌های فازی برای متغیرهای ورودی که شامل فاصله هر گره حسگر تا سینک و انرژی سطح باتری گره حسگر و خروجی فازی که شامل نرخ برازندگی هر حسگر است تعریف می‌شوند. توابع عضویت برای خروجی در شکل ۵ نمایش داده شده است. پس از فازی‌سازی، موتور استنتاج^۵ فازی سوگنو قوانین فازی از پیش تعیین شده را پردازش می‌کند.

در مرحله استنتاج، نرخ برازندگی هر گره با استفاده از قواعد فازی^۶ و با توجه به مقدار پارامترهای در نظر گرفته شده محاسبه می‌شود. هر قاعده فازی شامل دو قسمت است: یک قسمت مقدمه به صورت "اگر فاصله تا سینک زیاد و انرژی باتری نیز کم شود" و یک قسمت نتیجه به صورت "آن‌گاه نرخ برازندگی خیلی کم است". برای هر یک از پارامترهای ورودی سه مجموعه فازی تعریف شده است که به نه قاعده فازی منتهی خواهد شد، این قواعد در جدول ۱ تعریف شده‌اند.

سپس، خروجی فازی برای هر قانون فعال شده محاسبه می‌شود. در نهایت غیرفازی‌ساز خروجی‌های فازی را تجمیع کرده و به یک مقدار واحد $Fs(i)$ یا نرخ برازندگی تبدیل می‌نماید.

غیرفازی‌ساز: برای ترجمه خروجی فازی به مقدار عددی، از غیرفازی‌ساز استفاده می‌شود. در این مقاله از غیرفازی‌ساز میانگین مراکز استفاده شده است که با استفاده از رابطه ۱ محاسبه می‌شود.

۲. تعداد ابعادی که تغییر می‌کند (CDC^1): این عامل ابعادی را که تغییر خواهد کرد تعیین می‌کند.

۳. جستجو در محدوده تعریف شده ابعاد (SRD^2): این عامل نرخ تغییرات برای بعد انتخاب شده را نشان می‌دهد. در حالت جستجو اگر بعدی برای تغییر انتخاب شود تفاوت بین مقدار جدید و قدیم آن خارج از محدوده‌ای که توسط SRD تعریف شده است نخواهد بود.

۴. توجه به موقعیتی که گره در آن است (SPC^3): این عامل یک متغیر بولین می‌باشد که تعیین می‌کند که آیا این نقطه‌ای که گره در حال حاضر در آن ایستاده یکی از کاندیدهای حرکت خواهد بود و یا خیر. درست و یا غلط بودن ارزش SPC بر روی ارزش SMP تاثیری نخواهد داشت.

در ادامه این بخش به توضیح روش پیشنهادی برای بهبود مصرف انرژی در شبکه حسگر در کاربرد ردیابی هدف متحرک می‌پردازیم. در روش پیشنهادی هر گره حسگر بعنوان یک گره در نظر گرفته می‌شود. الگوریتم پیشنهادی شامل دو فاز خوشه‌بندی و ردیابی هدف است. در فاز خوشه‌بندی مراحل زیر اجرا می‌شود.

مرحله ۱: به تعداد NP تا گره حسگر یا گره بصورت تصادفی ایجاد می‌شود. برای هر گره مقدار موقعیت فیزیکی و سرعت بصورت تصادفی مقداردهی اولیه می‌شود. مقدار $NP=60$ در نظر گرفته شده است.

مرحله ۲: برای هر گره (گره حسگر) تابع برازندگی طبق سیستم منطق فازی محاسبه می‌شود، این فرآیند توسط یک کنترل کننده تطبیقی صورت می‌پذیرد سپس بهترین موقعیت گره در x_{best} ذخیره می‌شود. در ادامه به تشریح عملکرد سیستم فازی استفاده شده برای محاسبه برازندگی هر گره می‌پردازیم:

۳-۱- فازی

عملکرد سیستم فازی جهت محاسبه پارامتر برازندگی شامل سه بخش است (شکل ۲): فازی‌سازی^۴، موتور استنتاج فازی سوگنو، و غیرفازی‌سازی. همچنین، همانطور که ذکر شد، دو متغیر ورودی، انرژی باقیمانده: $E(n)$ ، فاصله از سینک: $D(n)$ برای هر گره n در نظر گرفته می‌شود. هدف روش پیشنهادی محاسبه برازندگی گره

¹ Counts of Dimension to Change(CDC)

² Seeking Range of the selected Dimension(SRD)

³ Self -Position Considering(SPC)

⁴ Fuzzification

⁵ Inference

⁶ Fuzzy Rule

$$Fs(i) = \frac{\sum_{l=1}^m y^{-l} \prod_{i=1}^n \mu A_i^l(X_i)}{\sum_{l=1}^m \prod_{i=1}^n \mu A_i^l(X_i)} \quad (1)$$

پارامترهای این رابطه عبارتند از، $Fs(i)$ تابع برازندگی حسگر i ام، i : اندیس گره، m : تعداد قواعد فازی (در اینجا ۹ می باشد)، n : تعداد توابع عضویت متغیرهای ورودی (در اینجا ۳ می باشد)، $\mu A_i^l(X_i)$: مقدار فازی توابع عضویت و y^{-l} نیز مراکز خروجی می باشد. بهترین پاسخ که نرخ برازندگی بالاتری دارد توسط سیستم فازی تعیین می شود و بهترین پاسخ در x_{best} ذخیره می شود.

مرحله ۳: بصورت تصادفی $NP*MR$ تا از گربهها انتخاب و آنها را در حالت ردیابی قرار داده می شوند و مقدار پرچم آنها به یک تنظیم می شود، سایر گربهها در حالت جستجو قرار داده شده و مقدار فیلد پرچم آنها به صفر تنظیم می شود. بطوریکه MR تعداد گربههایی که وارد مرحله تعقیب و گریز می شوند را نسبت به کل گربهها تعیین می کند. در این مقاله $MR = \frac{1}{3}$ در نظر گرفته شده است لذا چهل گربه در وضعیت جستجو و بیست گربه در وضعیت ردیابی قرار می گیرند.

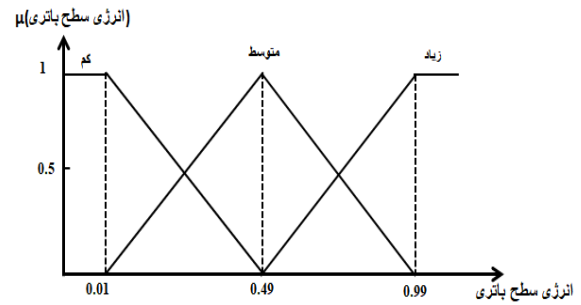
مرحله ۴: اگر گربه i ام در مرحله جستجو باشد (مقدار پرچم آن برابر صفر باشد) فاز خوشه بندی آغاز می شود و مراحل ۴-۱ تا ۴-۶ که طی آن سرخوشهها تعیین می گردد اجرا می شود.

مرحله ۴-۱: هر گره حسگر مانند i ، $timer$ خودش را قبل از شروع انتخاب سرخوشه طبق رابطه ۲ تنظیم می کند.

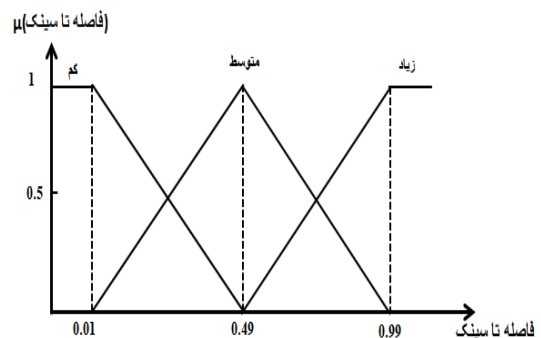
$$timer(i) = \frac{E_m(i) - E_r(i)}{E_m(i)} * T_{CH} \quad (2)$$

که در آن T_{CH} حداکثر زمان اختصاص داده شده برای انتخاب سرخوشه است و $E_m(i)$ و $E_r(i)$ ، حداکثر انرژی و انرژی باقی مانده برای گره حسگر i است. هنگامی که $timer$ تمام می شود گره i خودش را بعنوان سرخوشه موقت معرفی و $Advertisment$ را برای اعلام سرخوشه شدنش که حاوی Ip و موقعیت فیزیکی و مقدار انرژی باقیماندهاش است در رنج ارتباطی خود همه پخش می کند. هر گره ای که این پیام را دریافت می کند $timer$ خود را متوقف می کند در صورتیکه یک گره چندین پیام اعلامیه را دریافت کند، سپس چنین گرههایی که سرخوشه نیستند، عملیات جستجو را طبق الگوریتم ازدحام گربه انجام می دهد. در مرحله جستجو هر گربه به جستجو در اطراف موقعیت خود می پردازد و به سرخوشه با موقعیت بهتر (فاصله کمتر تا سرخوشه و انرژی بیشتر) می پیوندد.

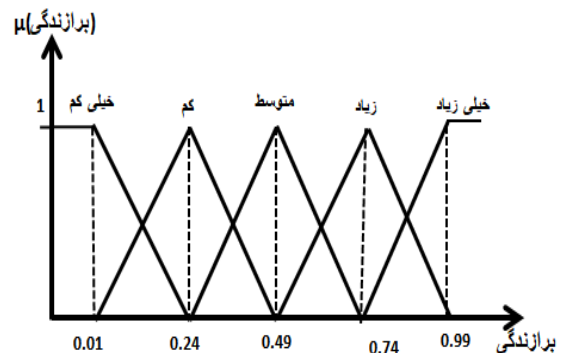
مرحله ۴-۲: در این مرحله (مرحله جستجو) هر گربه در حال استراحت به جستجوی موقعیت بهتر در اطراف موقعیت خود



شکل ۳. توابع عضویت برای متغیر ورودی انرژی باقیمانده سطح باتری



شکل ۴. توابع عضویت برای متغیر ورودی فاصله تا سینک



شکل ۵. توابع عضویت برای متغیر خروجی نرخ برازندگی هر گربه

جدول ۱. پایگاه قواعد فازی

انرژی سطح باتری / فاصله تا سینک	کم	متوسط	زیاد
کم	متوسط	زیاد	خیلی زیاد
متوسط	کم	متوسط	زیاد
زیاد	خیلی کم	کم	متوسط

می‌گردد. وقتی گربه وارد مرحله ردیابی می‌شود برحسب سرعت تحرک هدف، در هر بعد شعاع حسی خود را تنظیم می‌کند و با دور شدن هدف از محدوده رادیوی‌اش به خواب می‌رود و سرخوشه به گره‌های عضو خوشه‌اش اعلام خواب می‌دهد.

مرحله ۵-۱: در روش پیشنهادی فرض شده محدوده تعریف شده برای ابعاد ۳۶۰ درجه باشد. برای هر بعد $(v_{k,d})$ سرعت حرکت گربه با توجه به رابطه (۴) به‌روز می‌شود. منظور از سرعت حرکت حسگر یا گربه، همان سرعت افزایش شعاع حسی حسگر است.

$$v_{k,d} = v_{k,d} + r * c * (x_{best,d} - x_{k,d}), \quad (4)$$

$$d = 1,2,3,4$$

بطوریکه d تعداد ابعاد است که برابر چهار می‌باشد، $x_{best,d}$ موقعیت گربه‌ای که بیشترین مقدار برازندگی را دارد، $x_{k,d}$ موقعیت گربه k ام است. c عددی ثابت که $c=0.9$ فرض شده و r عددی تصادفی در بازه $[0,1]$ است.

مرحله ۵-۲: در این مرحله بررسی می‌شود که آیا سرعت‌ها در محدوده تعریف شده هستند یا خیر؟ در صورتیکه سرعتی بیشتر از محدوده تعریف شده بود با حداکثر مقدار ممکن در محدوده جایگزین می‌شود. در روش پیشنهادی رنج شعاع حسی هر گره حسگر، به چهار قسمت تقسیم می‌شود که هر قسمت یک بعد در نظر گرفته شده است (بعد اول تا ۴۵ درجه می‌باشد، بعد دوم ۹۰ درجه و بعد سوم ۱۸۰ و بعد چهارم ۳۶۰ می‌باشد). هر گره حسگر می‌تواند رنج شعاع حسی خود را از ۴۵ تا ۳۶۰ درجه افزایش یا کاهش دهد. اگر شعاع حسی حسگر تا ۵۰ متر را پوشش دهد هدف تا ۵۰ متری برای حسگر قابل شناسایی خواهد بود. بنابراین در صورتیکه رنج شعاع حسی حسگر روی ۴۵ درجه تنظیم شود حسگر قادر خواهد بود تا ۶ متری خود را حس کند وقتی هدف در ۳ متری حسگر است نیازی نیست حسگر شعاع حداکثری خود یعنی ۳۶۰ درجه را استفاده کند چرا که استفاده از شعاع حداکثری نیاز به صرف انرژی بالایی دارد لذا بهتر است در مواقع نزدیکی هدف به حسگر رنج شعاع حسی کاهش یابد. لذا بسته به نزدیک یا دور شدن هدف، حسگر رنج شعاع حسی خود را تنظیم می‌کند برای مثال در صورتیکه هدف در ۱۶ متری حسگر باشد اگر رنج شعاع حسی در ۱۸۰ درجه تنظیم شود می‌تواند هدف را تشخیص دهد و دیگر نیازی به افزایش رنج شعاع حسی تا ۵۰ متر یا ۳۶۰ درجه که باعث صرف انرژی زیاد باتری می‌شود نیست. حال اگر حسگر در بعد دوم (۹۰ درجه) باشد و هدف در فاصله ۶۰ متری حسگر واقع شده باشد حسگر برای ردیابی هدف، حداکثر قادر است شعاع حسی خود را تا ۲۸۰ درجه دیگر افزایش دهد تا

می‌پردازد. هر گره حسگر یک پیغام hello را برای شناسایی همسایه‌ها در رنج خود همه‌پخش می‌کند به ازای هر پاسخ lack از همسایه‌هایش دریافت می‌کند. موقعیت همسایه بعنوان کپی جدید از موقعیت گربه در نظر گرفته می‌شود و تعداد کپی‌ها در متغیر SMP ثبت می‌شود. بنابراین در این مرحله به تعداد z کپی از هر گربه ایجاد می‌شود. در روش پیشنهادی مقدار متغیر $SPC=TRUE$ فرض شده لذا مقدار $SMP=1-z$ خواهد بود و بدین ترتیب موقعیت اولیه گره به عنوان یکی از کپی‌ها در نظر گرفته می‌شود.

مرحله ۴-۳: برای هر کدام از کپی‌ها (گره‌های همسایه گره سرخوشه موقت) میزان تابع برازندگی را مطابق رابطه (۱) با استفاده از منطق فازی محاسبه می‌کند.

مرحله ۴-۴: اگر تمام FSها برابر نبود احتمال انتخاب هر گره کاندید با توجه به برازش نرمال شده آن موقعیت محاسبه می‌شود در غیر این صورت احتمال انتخاب همه موقعیت‌های کاندید برابر قرار داده می‌شود. رابطه (۳) نحوه محاسبه برازش نرمال شده یک گربه کاندید را با توجه به میزان برازندگی آن گربه نشان می‌دهد.

$$p_i = \frac{|FS_i - FS_b|}{FS_{max} - FS_{min}} \quad (3)$$

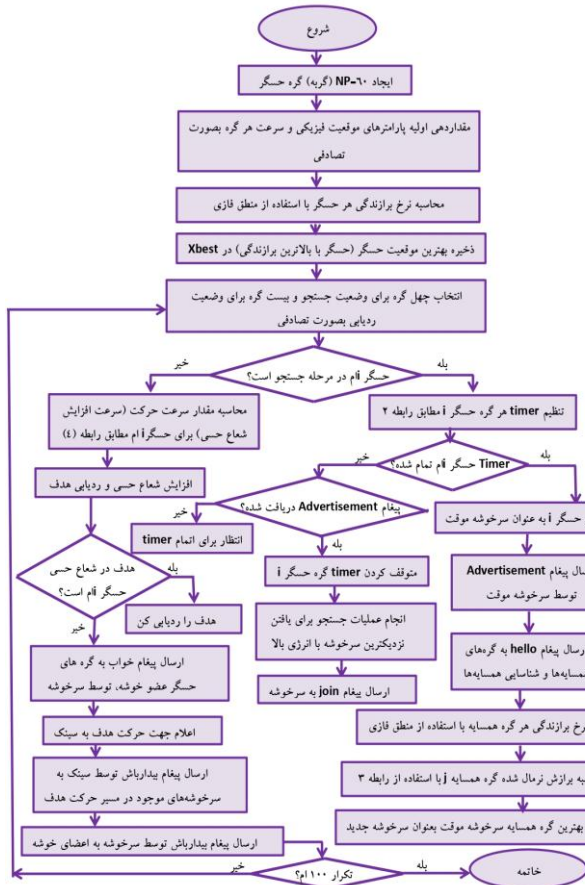
بطوریکه p_i احتمال انتخاب موقعیت گربه i ام و FS_i تابع برازندگی گربه i ام، FS_{max} حداکثر فاصله گربه تا هدف، FS_{min} کمتری فاصله گربه تا هدف است. با توجه به اینکه هدف تابع برازندگی در روش پیشنهادی پیدا کردن راه حل منیمم است لذا $FS_b = FS_{max}$ فرض می‌شود.

مرحله ۴-۵: پس از محاسبه احتمال انتخاب موقعیت گربه i ام طبق رابطه (۳) از بین کپی‌ها یک موقعیت بعنوان موقعیت جدید گربه انتخاب می‌شود بدین صورت که گره با مقدار p_i بالا، که بهترین موقعیت را از بین گره‌های همسایه سرخوشه موقت دارد بعنوان سرخوشه جدید در نظر گرفته می‌شود.

مرحله ۴-۶: تمام گره‌های همسایه که در رنج رادیویی سرخوشه جدید هستند پیغام Join را که حاوی مقدار انرژی باقی مانده، IP گره و موقعیت فیزیکی‌شان است را به سمت گربه با بهترین موقعیت در رنج خود می‌فرستند بدین ترتیب خوشه‌ها تشکیل می‌شوند و فاز خوشه‌بندی در این دور خاتمه می‌یابد.

مرحله ۵: اگر گربه i ام در مرحله تعقیب و گریز یا ردیابی باشد (مقدار پرچم آن برابر یک باشد) فاز ردیابی آغاز می‌شود و مراحل ۱-۵ تا ۵-۵ که طی آن ردیابی هدف متحرک انجام می‌شود اجرا

الگوریتم پایان می یابد در غیر این صورت مراحل ۳ تا ۵ اجرا می شود در روش پیشنهادی اگر انرژی سرخوشه ها از مقدار آستانه (۵۰ ژول) کمتر شد الگوریتم مراحل جستجو را برای خوشه بندی مجدد انجام می دهد در غیر این صورت با رسیدن به تکرار ۱۰۰ ام الگوریتم خاتمه می یابد. شکل ۶ فلوچارت روش پیشنهادی را نمایش می دهد.



شکل ۶ فلوچارت روش پیشنهادی

لازم به ذکر است پیچیدگی زمانی الگوریتم ازدحام گره، اغلب با تعداد تکرارها و اندازه جمعیت یا تعداد گره های حسگر مرتبط است که با افزایش تعداد گره ها و تعداد تکرارها افزایش می یابد. از طرفی پیچیدگی زمانی سیستم استنتاج فازی به تعداد قوانین، و تعداد توابع عضویت در مجموعه های فازی بستگی دارد. ترکیب این دو الگوریتم از نظر بار محاسباتی، در دوره های ابتدای اجرا (قبل از یافتن سرخوشه های مناسب و انجام عمل خوشه بندی) موجب افزایش سربار پردازشی می شود. افزایش بار محاسباتی موجب افت کارایی و بالا رفتن مصرف انرژی در گره های حسگر می شود. از طرفی در دوره های ابتدای شبیه سازی، تشخیص سرخوشه های مناسب و تشکیل خوشه ها نیاز به آزمون و خطا دارد لذا این فرآیند معمولاً باعث افزایش تعداد محاسبات و افزایش انرژی مصرفی

به ۳۶۰ درجه (۵۰ متر) برسد افزایش بیش از آن برای این بعد از حسگر عملی نیست که در این صورت قادر به ردیابی این هدف نخواهد بود.

مرحله ۳-۵: به هنگام سازی موقعیت گره k، تا شعاع ۳۶۰ درجه؛ هر حسگر به اندازه ۴۵ درجه شعاع حسی خود را زیاد می کند تا بتواند هدف را ردیابی کند.

مرحله ۴-۵: اگر هدف از محدوده حسی گره های حسگر یا دید گره ها دور شد بایستی گره های خوشه بخواب بروند. لذا با خروج هدف از دید اعضای خوشه گره سرخوشه به اعضای داخل خوشه پیغام به خواب رفتن را صادر می کنند. پس از اتمام ردیابی حسگرهای داخل خوشه بسته ای را به سرخوشه می فرستد که شامل اطلاعات هدف ردیابی شده مانند <شماره حسگر، شماره بعد، محل ورود هدف، محل خروج هدف، زمان ورود، زمان خروج> می باشد. سرخوشه پس از دریافت بسته از حسگرهای خوشه اش، بسته ها را به سینک می فرستد و سینک پس از مقایسه زمان و محل ورود - خروج هدف می تواند مسیر را شناسایی کند. حسگر برای ارسال بسته به حسگرهایی که در جهت حرکت هدف هستند از سرخوشه استفاده می کند. گره حسگر، جهت حرکت هدف را

اطلاع می دهد و سرخوشه نیز به سینک اطلاع می دهد. سپس سینک با توجه به اینکه آیا حسگرها در زمان عبور اهداف فعال بوده اند یا خیر زمان خواب آنها را می تواند تغییر دهد. مثلاً هنگام عبور هدف از منطقه شناسایی گره حسگر، اگر قبل از رسیدن هدف به منطقه مجاور سرخوشه منطقه مجاور خواب باشد ولی حسگری در همسایگی سرخوشه مذکور هدف را شناسایی و ردیابی کرده است، سینک، سرخوشه گره مجاور را بیدار می کند تا دقت

ردیابی افزایش یابد. بر این اساس سینک یک پیغام بیدارباش به سرخوشه مجاور که در لحظه بعدی هدف را مشاهده خواهد نمود ارسال می کند و سرخوشه نیز گره های داخل خوشه خود را بیدار می کند و همچنین سینک به سرخوشه ای که هدف را رویت کرده و احتمال رویت آن در زمان بعدی کم است پیغام به خواب رفتن را ارسال تا سرخوشه به گره های عضو خوشه خود اطلاع دهد که انتقال دهنده ایشان را خاموش و به خواب بروند. بدین ترتیب در مصرف انرژی صرفه جویی شده و طول عمر شبکه افزایش می یابد.

مرحله ۵-۵: هر گره سرخوشه که پیغام بیدارباش را از سینک دریافت نموده به گره های عضو خود حالت بیدارباش داده و آنها را از خواب بیدار می کند.

مرحله ۶: اگر شرط خاتمه (رسیدن به تکرار ۱۰۰ ام) برقرار بود

دلیل مقایسه سناریوهای الگوریتم پیشنهادی با سناریوی Tracking-45-Degree-vectors این بود که هر دو روش توپولوژی یکسانی دارند و بر اساس صدا کار می‌کنند همچنین در هر دو روش هدف قابل ردیابی سوسک سرخرطومی خرما می‌باشد. در ادامه به نتایج شبیه‌سازی پروتکل پیشنهادی بر اساس سناریوها می‌پردازیم.

لازم به ذکر است، در هر دو سناریو توپولوژی شبکه بصورت ۶۰ گره حسگر در نظر گرفته شده است (شکل ۷). گره‌های حسگر استفاده شده در شبیه‌سازی قابلیت تشخیص صدا سوسک سرخرطومی را دارند و از این طریق هدف را تشخیص و شناسایی می‌کنند. مشخصات سخت‌افزاری گره‌های حسگر استفاده شده برای ردیابی سوسک سرخرطومی در شکل ۸ نمایش داده شده است.

۴-۲- معیارهای کارایی در روش پیشنهادی

به منظور بررسی کارایی روش پیشنهادی از معیارهای زیر استفاده می‌شود.

۴-۲-۱- تأخیر انتها به انتها: میانگین زمان تحویل بسته از مبدا به مقصد، میانگین تأخیر انتها به انتها نامیده می‌شود. شکل ۹ به مقایسه تأخیر انتها به انتها برای سناریوهای الگوریتم پیشنهادی و سناریوی Tracking-45-Degree-vectors می‌پردازد. محور عمودی تأخیر انتها به انتها و محور افقی زمان شبیه‌سازی است. همانگونه که مشاهده می‌شود، با بکارگیری پروتکل پیشنهادی، نرخ تأخیر انتها به انتها برای توپولوژی شبکه با ۶۰ گره به میزان ۲۷/۰۲ درصد نسبت به سناریوی Tracking-45-Degree-vectors بهبود یافته است. در سناریوی Tracking-45-Degree-vectors بدلیل عدم پایداری لینک و در نتیجه انتخاب گره‌های نامناسب بعنوان سرخوشه عمل انتقال اطلاعات تکمیل نمی‌شود، لذا تأخیر انتها به انتها افزایش می‌یابد، در حالیکه در پروتکل پیشنهادی بدلیل استفاده از الگوریتم ازدحام گره و سیستم کنترل فازی مسیرها با پایداری بالاتری (انتخاب سرخوشه‌های مناسب بر حسب دو پارامتر) در قیاس با سناریوی مقاله Tracking-45-Degree-vectors انتخاب می‌شوند که کاهش تأخیر انتها به انتها را در شبکه به همراه خواهد داشت.

گره‌های حسگر می‌شود. پس از تشکیل خوشه‌ها در دوره‌های آخر (دور صدم به بعد) اگر انرژی سرخوشه‌ها از مقدار آستانه تعیین شده (۵۰ ژول) کمتر نباشد مراحل جستجو برای خوشه‌بندی مجدد انجام نمی‌شود لذا هزینه پردازش و به تبع آن مصرف انرژی گره‌های حسگر کاهش می‌یابد.

۴- شبیه‌سازی روش پیشنهادی

۴-۱- محیط شبیه‌سازی

در این مقاله برای شبیه‌سازی روش پیشنهادی و مقایسه آن با روش پیشنهادی مقاله [۲] از نرم افزار OpnetModeler^۱ استفاده شده است. لازم به ذکر است در این مقاله، دو الگوریتم ازدحام گربه به همراه منطق فازی و روش پیشنهادی مقاله [۲] جهت مقایسه و شبیه‌سازی انتخاب شده‌اند. این انتخاب از آنجا نشأت می‌گیرد که دو الگوریتم، از روشهای هوشمند برای ردیابی سوسک سرخرطومی خرما استفاده می‌کنند.

جدول ۲ پارامترهای شبیه‌سازی بکار گرفته شده توسط شبیه‌ساز را نشان می‌دهد.

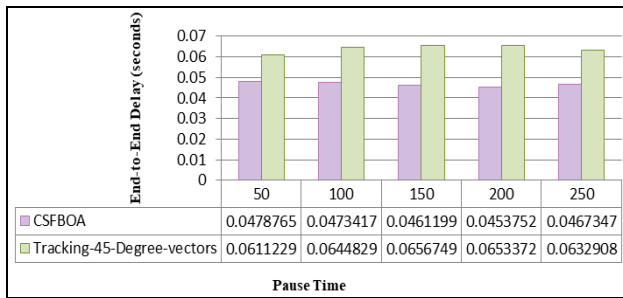
همان گونه که مطرح شد، در سناریو اول گره‌ها بر اساس پروتکل پیشنهادی مقاله [۲] (یک پروتکل خوشه‌بندی مبتنی بر آتاماتای یادگیر در شبکه‌های حسگر بی‌سیم) در محیط پراکنده شده‌اند و در سناریو دوم (روش پیشنهادی) که با نام CSFBOA^۲ (الگوریتم ردیابی مبتنی بر الگوریتم ازدحام گربه و منطق فازی) نامگذاری شده گره‌ها بصورت تصادفی در محیط پخش می‌شوند.

جدول ۲. پارامترهای شبیه‌سازی

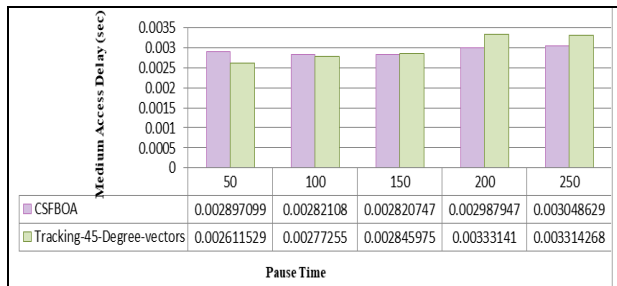
پارامتر	مقدار
روش پخش گره‌ها	تصادفی
اندازه محیط شبیه‌سازی	۱۰۰۰m × ۱۰۰۰m × ۱۰۰۰m
تعداد گره‌ها	۶۰
اندازه بسته	۱۰۲۴byte
مدل باتری	Constant
زمان شبیه‌سازی	۲۵۰ ثانیه
پروتکل لایه mac	IEEE802.15.4
مقدار اولیه انرژی	۲۰۰ تا ۴۰۰ ژول
تعداد هدف	۱

^۱ Optimized Network Engineering Tool

^۲ Cat Sarwm Fuzzy Based Optimization Algorithm



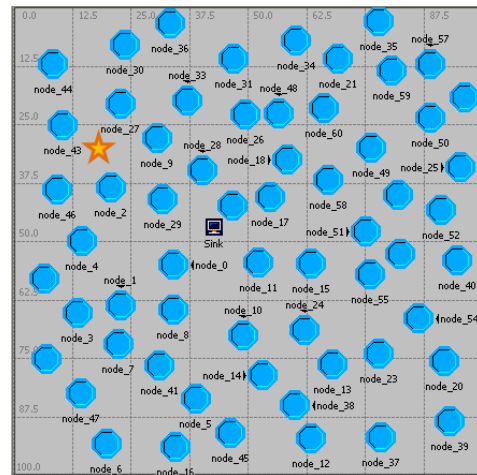
شکل ۹. تأخیر انتها به انتها



شکل ۱۰. تأخیر دسترسی به رسانه

این مساله از آنجا نشات می‌گیرد که در شبکه حسگر بی سیم زیرلایه MAC مکانیسم‌های آدرس‌دهی و کنترل دسترسی کانال را فراهم می‌کند که به گره‌های حسگر اجازه می‌دهد در شبکه‌ای که دارای یک رسانه مشترک است، ارتباط برقرار کنند. برای انجام این کار، MAC با استفاده از یک پروتکل دسترسی به کانال، تعیین می‌کند که یک گره حسگر چه زمانی می‌تواند رسانه را در اختیار بگیرد و داده خود را ارسال کند یا در حال گوش کردن به رسانه باشد. به طور معمول، بسته به نوع پیوند فیزیکی، کانال خدمات ارتباطی تک پخشی، چندپخشی (برای پیوندهای سیمی) یا همه پخشی (یعنی برای پیوندهای بی سیم) را فراهم می‌کند. هنگام در نظر گرفتن پیوندهای بی سیم، نیاز به یک گره هماهنگ کننده برای جلوگیری از تداخل برخورد (انتقال همزمان روی رسانه، که منجر به پارازیت سیگنال می‌شود ضروری است.

در پروتکل پیشنهادی جهت دسترسی سریع به رسانه حسگرها به طور مداوم به کانال گوش می‌دهند، تا در صورت آزاد بودن کانال سریع رسانه را برای ارسال داده در اختیار بگیرند که این امر موجب کاهش تأخیر دسترسی به کانال شده است. در صورتیکه در روش Tracking-45-Degree-vectors دسترسی به رسانه به صورت تصادفی می‌باشد که در آن انتقال‌ها به صورت محلی و مستقل در هر گره سرخوشه تصمیم‌گیری می‌شوند که این امر تأخیر دسترسی به رسانه را در روش را افزایش می‌دهد.

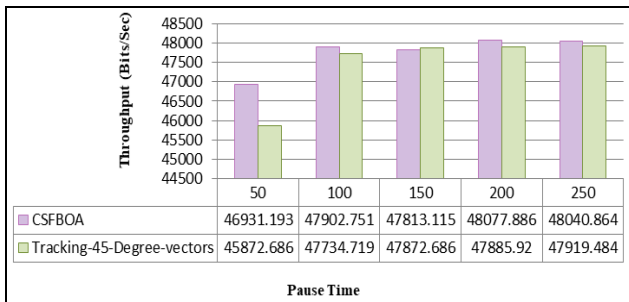


شکل ۷. توپولوژی شبکه با تعداد ۶۰ گره حسگر و یک هدف

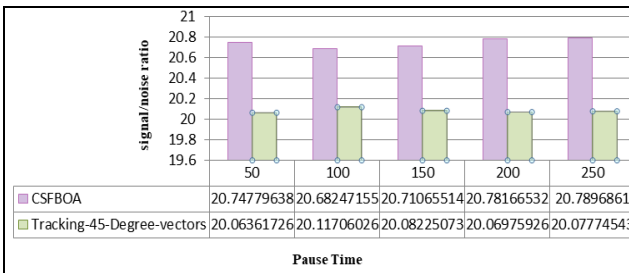
name	node_24
model	wpan_sensor_node_v1_0
x position	80.4
y position	20.6
threshold	0.0
icon name	Micaz
creation source	Object copy
creation timestamp	09:34:39 ÅæÑíá 16 2015
creation data	Copy of node_7
CSMA/CA Parameters	(...)
Battery	(...)
Current Draw	(...)
Initial Energy	200
Power Supply	2 AA Batteries (3V)
IEEE 802-15-4	(...)
Device Mode	Slave
MAC Address	Auto Assigned
MAC Attributes	(...)
Batterie Life Extension	disabled
WPAN Settings	(...)
Beacon Order	5
Superframe Order	1
PAN ID	1
Enable Log File	enabled
Traffic Source	(...)
Sensory Data	(...)
Packet Interarrival Time (seconds)	exponential (1.0)
Packet Size (bits)	constant (900)
Start Time (secs)	1.0
Stop Time (secs)	Infinity

شکل ۸. مشخصات سخت‌افزاری گره‌های حسگر استفاده شده در روش پیشنهادی

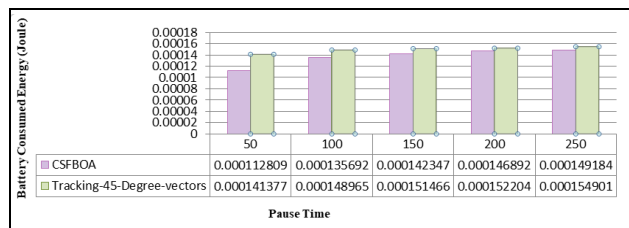
۴-۲-۲- تأخیر دسترسی به رسانه: برابر مدت زمان بین دریافت بسته توسط لایه MAC تا زمان بارگیری کامل بسته بر روی رسانه تعریف می‌شود. شکل ۱۰ به مقایسه تأخیر دسترسی به رسانه برای سناریوهای الگوریتم پیشنهادی و سناریوی Tracking-45-Degree-vectors می‌پردازد. همانگونه که مشاهده می‌شود، در پروتکل پیشنهادی، نرخ تأخیر دسترسی به رسانه برای توپولوژی شبکه با ۶۰ گره به میزان ۲/۰۱ درصد نسبت به سناریوی Tracking-45-Degree-vectors، بهتر بوده است.



شکل ۱۱. نرخ گذردهی



شکل ۱۲. نسبت سیگنال به نویز



شکل ۱۳. انرژی مصرفی

در دوره‌های اولیه اجرا (قبل از تکرار ۱۰۰ ام و قبل از ۵۰ ثانیه اول اجرا)، گره‌های حسگر نیاز به پردازش و انتقال داده‌ها به سینک را دارند که برای انجام پردازش‌های مورد نیاز (الگوریتم ازدحام گره‌ها و منطق فازی)، گره‌های حسگر باید از منابع پردازشی و حافظه‌ای خود استفاده کنند لذا بار محاسباتی الگوریتم افزایش می‌یابد. بارهای محاسباتی بالاتر منجر به افزایش مصرف انرژی گره‌های حسگر می‌شود. از طرفی در مراحل اولیه اجرا به دلیل نیاز به راه‌اندازی ارتباطات و تبادل پیام‌های کنترلی بار محاسباتی بیشتر و در نتیجه مصرف انرژی گره‌های حسگر نیز بیشتر می‌شود. پس از تشکیل خوشه‌بندی، تا زمانیکه گره‌های سرخوشه، انرژی بالا داشته باشند نیاز به اجرای الگوریتم نیست لذا در مراحل بعدی مصرف انرژی کاهش می‌یابد. همانگونه که مشاهده می‌شود، با بکارگیری پروتکل پیشنهادی، بعد از ثانیه ۵۰ام، میانگین انرژی مصرفی باتری برای توپولوژی شبکه با ۶۰ گره به میزان ۸/۲۷۷ درصد بهبود یافته است. در پروتکل Tracking-45-Degree-vectors در انتخاب سرخوشه گره‌ای که تعداد همسایه بیشتر دارد و انرژی باقیمانده‌اش کمتر از حد آستانه نباشد سرخوشه می‌شود. اگر گره‌ای با این ویژگی موجود نبود، این الگوریتم، خوشه‌بندی

۴-۲-۳- نرخ گذردهی: به عنوان کل بسته‌های دریافت شده توسط گیرنده‌ها، تقسیم بر زمان بین دریافت اولین بسته و آخرین بسته تعریف می‌شود. نرخ گذردهی سناریوهای مطرح شده را می‌توان در شکل ۱۱ مشاهده کرد. محور افقی زمان شبیه‌سازی و محور عمودی نرخ گذردهی می‌باشد. همانگونه که مشاهده می‌شود با بکارگیری پروتکل پیشنهادی، نرخ گذردهی برای توپولوژی شبکه با ۶۰ گره به میزان ۰/۶۲ درصد نسبت به سناریو پروتکل سناریوی Tracking-45-Degree-vectors بالاتر است. این مساله را می‌توان ناشی از هوشمندتر بودن پروتکل پیشنهادی در انتخاب مسیر دانست، زیرا این پروتکل با بکارگیری الگوریتم ازدحام گره‌ها و منطق فازی گره‌های مناسب را بعنوان سرخوشه تعیین کرده و داده را از طریق سرخوشه‌های مناسب ارسال می‌کند که موجبات بهبود نرخ گذردهی را فراهم می‌آورد. سناریوی مقاله Tracking-45-Degree-vectors نسبت به روش پیشنهادی بدلیل خاموش شدن احتمالی گره و عدم خوشه‌بندی مجدد در دوره‌های بعدی نرخ گذردهی کمتری دارد.

۴-۲-۴- نسبت سیگنال به نویز: میزان قدرت نویز تحمیل شده به یک سیگنال در مقابل قدرت خود سیگنال را نشان می‌دهد. نسبت سیگنال به نویز در ساختارهای شبیه‌سازی شده در شکل ۱۲ نشان داده شده است. با بکارگیری پروتکل پیشنهادی، نسبت سیگنال به نویز به میزان ۳/۲۸ درصد نسبت به سناریوی Tracking-45-Degree-vectors عملکرد بهتری دارد. همانگونه که مشاهده می‌شود، سناریوی Tracking-45-Degree-vectors نسبت به روش پیشنهادی نسبت سیگنال به نویز کمتری دارد، زیرا ممکن است در روش Tracking-45-Degree-vectors حین ارسال داده تعداد بیت‌هایی که دچار خطا شده‌اند (بدلیل خاموش شدن احتمالی سرخوشه‌ها در اثر کاهش انرژی باتری) زیاد شده و نسبت سیگنال به نویز کاهش یابد. در روش پیشنهادی بدلیل انتخاب گره‌های مناسب به عنوان سرخوشه توسط الگوریتم ازدحام گره‌ها و منطق فازی، احتمال وقوع اغتشاش کاهش می‌یابد که افزایش نسبت سیگنال به نویز را بدنبال خواهد داشت.

۴-۲-۵- میانگین انرژی مصرفی باتری گره‌های شبکه: مجموع انرژی مصرف شده توسط گره‌های درون شبکه برای ارتباطات، که شامل انتقال و انتظار می‌باشد. شکل ۱۳ به مقایسه میانگین انرژی مصرفی باتری برای سناریوهای الگوریتم پیشنهادی و سناریوی Tracking-45-Degree-vectors می‌پردازد.

[۳] عباسی ج.، دبیری ح.، امیری ع. " آفت قرنطینه‌ای سرخرطومی حنایی خرما "، ج ۱، انتشارات مدیریت هماهنگی ترویج کشاورزی استان فارس، ص ۱۴، بهار ۱۳۹۶. https://agrilib.areeo.ac.ir/book_3292.pdf

- [4] M. C. F. Lima, M. E. D. de Almeida Leandro, C. Valero, L. C. P. Coronel, and C. O. G. Bazzo, "Automatic detection and monitoring of insect pests—A review," *Agriculture*, vol. 10, no. 5, p. 161, 2020.
- [5] S. Suganya, "A cluster-based approach for collaborative target tracking in wireless sensor networks," in 2008 First International Conference on Emerging Trends in Engineering and Technology, 2008, pp. 276-281: IEEE.
- [6] Z. Wang, H. Li, X. Shen, X. Sun, and Z. Wang, "Tracking and predicting moving targets in hierarchical sensor networks," in 2008 IEEE International Conference on Networking, Sensing and Control, 2008, pp. 1169-1173: IEEE.
- [7] S. Balasubramanian, S. K. Jayaweera, and K. R. Namuduri, "Energy-aware, collaborative tracking with ad-hoc wireless sensor networks," in IEEE Wireless Communications and Networking Conference, 2005, 2005, vol. 3, pp. 1878-1883: IEEE.
- [8] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE signal processing magazine*, vol. 19, no. 2, pp. 17-29, 2002.
- [9] H.-Q. Liu, H.-C. So, F. K. W. Chan, and K. W. K. Lui, "Distributed particle filter for target tracking in sensor networks," *Progress In Electromagnetics Research C*, vol. 11, pp. 171-182, 2009.
- [10] K. R. Madhavi, M. N. M. Nawi, B. B. Reddy, K. Baboji, K. H. Kishore, and S. Manikanthan, "Energy efficient target tracking in wireless sensor network using PF-SVM (particle filter-support vector machine) technique," *Measurement: Sensors*, vol. 26, p. 100667, 2023.
- [11] S. Xiang and J. Yang, "A novel adaptive deployment method for the single-target tracking of mobile wireless sensor networks," *Reliability Engineering & System Safety*, vol. 234, p. 109135, 2023.
- [12] Z. Qu, H. Xu, X. Zhao, H. Tang, J. Wang, and B. Li, "A fault-tolerant sensor scheduling approach for target tracking in wireless sensor networks," *Alexandria engineering journal*, vol. 61, no. 12, pp. 13001-13010, 2022.
- [13] J. Munjani and M. Joshi, "A non-conventional lightweight Auto Regressive Neural Network for accurate and energy efficient target tracking in Wireless Sensor Network," *ISA transactions*, vol. 115, pp. 12-31, 2021.
- [14] B. M. Sahoo, H. M. Pandey, and T. Amgoth, "A genetic algorithm inspired optimized cluster head selection method in wireless sensor networks," *Swarm and Evolutionary Computation*, vol. 75, p. 101151, 2022.
- [15] M. Sadrishojaei, N. J. Navimipour, M. Reshadi, and M. Hosseinzadeh, "A new clustering-based routing method in the mobile internet of things using a krill herd algorithm," *Cluster Computing*, pp. 1-11, 2022.
- [16] P. Srinivas and P. Swapna, "Quantum tunicate swarm algorithm based energy aware clustering scheme for wireless sensor networks," *Microprocessors and Microsystems*, vol. 94, p. 104653, 2022.
- [17] J. Amutha, S. Sharma, and S. K. Sharma, "An energy efficient cluster based hybrid optimization algorithm with static sink and mobile sink node for Wireless Sensor Networks," *Expert Systems with Applications*, vol. 203, p. 117334, 2022.
- [18] R. F. Mansour et al., "Energy aware fault tolerant clustering with routing protocol for improved survivability in wireless sensor networks," *Computer Networks*, vol. 212, p. 109049, 2022.
- [19] M. Kaedi, A. Bohlooli, and R. Pakrooh, "Simultaneous optimization of cluster head selection and inter-cluster routing in

مجدد را انجام نمی‌دهد. لذا ممکن است سرخوشه‌های انتخابی در دوره‌های بعدی انرژی کمتری داشته باشند و انرژی خود را سریع از دست بدهند که این مساله بهم ریختگی توپولوژی شبکه را به دنبال خواهد داشت. همچنین در این پروتکل معیار فاصله تا سینک در نظر گرفته نشده است، در نتیجه گره‌هایی که فاصله زیاد با سینک دارند انرژی بیشتری جهت ارسال داده صرف خواهند نمود. در حالیکه در پروتکل پیشنهادی بعد از انجام عمل خوشه‌بندی، با استفاده از الگوریتم ازدحام گره و منطق فازی گرهی به‌عنوان سرخوشه استفاده می‌شود که انرژی بیشتر، فاصله تا سینک کمتری را داشته باشند. همچنین گره‌های عضو نیز با توجه به فاصله‌شان به گره سرخوشه می‌پیوندند، لذا برای ارسال داده از گره عضو به سرخوشه نیز نیازی نیست انرژی مصرفی زیادی صرف شود.

۵- نتیجه‌گیری

در این مقاله، مصرف انرژی به عنوان یکی از چالش برانگیزترین مشکلات موجود در شبکه‌های حسگر، بویژه برای کاربردهای ردیابی اهداف متحرک مورد مطالعه قرار گرفت. سپس به منظور ارایه یک روش مناسب برای خوشه‌بندی و ردیابی هدف متحرک یک روش پیش بینی به منظور پیش بینی مسیر حرکت هدف با استفاده از الگوریتم ازدحام گره و منطق فازی ارائه شد روش پیشنهادی با روش Tracking-45-Degree-vectors (یک روش جدید خوشه‌بندی بر اساس آتاماتای یادگیر بر پایه معیارهای میزان انرژی سطح باتری و تعداد همسایگی) در شبیه‌ساز OPNET شبیه‌سازی شد و نتایج شبیه‌سازی نظیر انرژی مصرفی، تأخیر دسترسی به رسانه‌ای، تأخیر انتها به انتها، نرخ سیگنال به نویز و نرخ گذردهی به منظور بررسی چگونگی عملکرد روش پیشنهادی استخراج شدند. به طور کلی مشاهده شد که روش پیشنهادی، رفتار بهتری نسبت به روش Tracking-45-Degree-vectors دارد و به دلیل انتخاب مسیرهای پایدارتر با گره‌هایی با انرژی بالا کارایی کلی شبکه را بهبود بخشیده و قابلیت ردیابی هدف را افزایش داده است.

مراجع

- [1] G. Simon et al., "Sensor network-based countersniper system," in Proceedings of the 2nd international conference on Embedded networked sensor systems, 2004, pp. 1-12.
- [2] S. Tabatabaei, "A Novel Method for Optimizing Energy Consumption in Applications for Detecting Palm Rhynchophorus Ferrugineus in WSNs Using Data mining and Q-Learning," *Wireless Personal Communications*, vol. 121, no. 1, pp. 1-17, 2021.

- [21] S.-C. Chu, P.-W. Tsai, and J.-S. Pan, "Cat swarm optimization," in PRICAI 2006: Trends in Artificial Intelligence: 9th Pacific Rim International Conference on Artificial Intelligence Guilin, China, August 7-11, 2006 Proceedings 9, 2006, pp. 854-858: Springer.
- [22] L. A. Zadeh, "Fuzzy sets," Information and control, vol. 8, no. 3, pp. 338-353, 1965.
- wireless sensor networks using a 2-level genetic algorithm," Applied Soft Computing, vol. 128, p. 109444, 2022.
- [20] N. Malisetti and V. K. Pamula, "Energy efficient cluster based routing for wireless sensor networks using moth levy adopted artificial electric field algorithm and customized grey wolf optimization algorithm," Microprocessors and Microsystems, vol. 93, p. 104593, 2022.

Investigating an Approach to Identify and Classify Mutants Based on the Characteristics of Mutants with Machine Learning Algorithms

Zeinab Asghari¹, Bahman Arasteh^{2*}, Abbas Koochari³

¹ Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

² Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

³ Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Received: 16 March 2024, Revised: 30 March 2024, Accepted: 02 May 2024

Paper type: Research

Abstract

Software mutation testing is one of the effective methods to evaluate code quality and detect hidden errors. However, this method faces challenges such as producing equivalent mutants and the process being time-consuming. The test cases designed for software testing must have the necessary sufficiency. For this purpose, the criterion of mutation test score is used. One of the main issues related to software mutation testing is the generation of equivalent mutants. Equivalent mutants are mutants that do not change the behavior of the program and have the same output as the original program. Identifying these mutants can make the mutation testing process time-consuming and costly. It is also possible that these mutants are mistakenly classified as hard to kill mutants. While hard to kill mutants can be rejected by changing the test items and strengthening them. In this paper, we present an efficient method for classifying program mutants to identify and separate equivalent mutations from hard to kill mutants. Using machine learning algorithms, we intend to optimize this process and improve the efficiency and effectiveness of the method. This is done by extracting different features from the mutants and using them to train machine learning models.

Keywords: Software Testing, Mutation Testing, Equivalent Mutants, Test Adequacy, Mutation Score.

* Corresponding Author's email: bahman.arasteh@istinye.edu.tr

ارائه رویکردی برای شناسایی و طبقه‌بندی جهش‌ها براساس ویژگی‌های جهش‌ها با الگوریتم‌های یادگیری ماشین

زینب اصغری^۱، بهمن آراسته^{۲*}، عباس کوچاری^۳

^۱دانشجوی دکترای دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران

^۲دانشیار دانشگاه آزاد اسلامی واحد تبریز، تبریز، ایران

^۳استادیار دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۱۲/۲۶ تاریخ بازبینی: ۱۴۰۳/۰۱/۱۱ تاریخ پذیرش: ۱۴۰۳/۰۲/۱۳

نوع مقاله: پژوهشی

چکیده

آزمون جهش نرم‌افزار، یکی از روش‌های موثر برای ارزیابی کیفیت کد و تشخیص خطاهای پنهان است. با این حال، این روش با چالش‌هایی مانند تولید جهش‌های معادل و زمان‌بر بودن فرایند مواجه است. موردآزمون‌های طراحی شده برای آزمون نرم‌افزار باید از کفایت لازم برخوردار باشند. برای این کار از معیار امتیاز آزمون جهش استفاده می‌شود. یکی از مسائل اصلی مرتبط با آزمون جهش نرم‌افزار، تولید جهش‌های معادل است. جهش‌های معادل، جهش‌هایی هستند که باعث تغییر در رفتار برنامه نمی‌شوند و خروجی یکسانی با برنامه اصلی دارند. شناسایی این جهش‌ها می‌تواند فرایند آزمون جهش را زمان‌بر و هزینه‌بر کند. همچنین امکان دارد این جهش‌ها به اشتباه در دسته جهش‌های سرسخت قرار بگیرند. درحالی که جهش‌های سرسخت را می‌توان با تغییر موردآزمون‌ها و تقویت آنها پیدا کرد. در این مقاله، ما به بررسی روشی برای طبقه‌بندی جهش‌های برنامه برای شناسایی و جداسازی جهش‌های معادل از جهش‌های سرسخت می‌پردازیم. با استفاده از الگوریتم‌های یادگیری ماشین، ما قصد داریم که این فرایند را بهینه‌سازی کنیم و کارآمدی و کارایی روش را بهبود دهیم. این کار با استخراج ویژگی‌های مختلف از جهش‌ها و استفاده از آنها برای آموزش مدل‌های یادگیری ماشین انجام می‌شود.

کلیدواژگان: آزمون نرم‌افزار، آزمون جهش، جهش‌های معادل، کفایت آزمون، امتیاز جهش.

* رایانامه نویسنده مسئول: bahman.arasteh@istinye.edu.tr

۱- مقدمه

عملگر جایگزین حسابی (AORB)، عملگری است که جهش‌های زیر را می‌تواند تولید کند: $x+y \rightarrow x-y$, $x+y \rightarrow x*y$, $x+y \rightarrow x/y$, $x+y \rightarrow x\%y$.

مثال دیگر برای عملگرهای جهشی، عملگری بنام VDL است که در عبارت موجود، یک عملگر را حذف می‌کند. نمونه‌ای از عملیات این عملگر به شرح زیر است: $x+y \rightarrow x$, $x+y \rightarrow y$. اگر مورد آزمون‌های طراحی شده، توانایی شناسایی این نسخه‌های جهش یافته را نداشته باشند، به این معنی است که آن آزمون‌ها ناکارآمد هستند و نیاز دارند تا بهبود داده شده و قوی‌تر شوند. از این رو، تولید جهش‌های معادل (جهش‌هایی که عملکرد برنامه را تغییر نمی‌دهند اما توسط آزمون‌ها تشخیص داده می‌شوند) و جداسازی آنها از جهش‌های سرسخت، یکی از چالش‌های اساسی در این زمینه محسوب می‌شود [۴۳].

در این مقاله، به دنبال ارائه رویکردی نوین برای شناسایی و جداسازی جهش‌های معادل از جهش‌های سرسخت هستیم. برای این منظور، به کمک الگوریتم‌های یادگیری ماشین و با استفاده از ویژگی‌های اصلی و تاثیرگذار جهش‌ها، سعی در توسعه یک مدل دقیق و کارآمد برای طبقه‌بندی جهش‌ها خواهیم داشت. این رویکرد، بهبود قابل توجهی در کارآمدی و کارایی فرآیند آزمون جهش نرم‌افزار دارد.

۲- کارهای پیشین

در این بخش، به بررسی جدیدترین روش‌های شناسایی جهش‌های معادل در آزمون جهش نرم‌افزار پرداخته شده است. این روش‌ها به شش دسته کلی تقسیم شده‌اند که هر کدام به اختصار توضیح داده شده است.

۲-۱- روش مبتنی بر تحلیل پویا

در این روش، نرم‌افزار به صورت پویا اجرا می‌شود و رفتار آن در طول اجرا مشاهده می‌شود. تحلیل پویا از جریان کنترل، استفاده از حافظه، ورودی‌ها و خروجی‌ها و... برای شناسایی جهش‌های معادل استفاده می‌کند [۵]. به عنوان مثال، اگر یک جهش معادل توسط یک آزمون تشخیص داده نشود و در زمان اجرا و تحلیل پویا، اثرات مشابهی با جهش اصلی داشته باشد، ممکن است نشان‌دهنده معادل بودن آن جهش باشد. این روش برای شناسایی جهش‌های معادل از تحلیل جریان کنترل، استفاده از حافظه، ورودی‌ها و خروجی‌ها و دیگر جنبه‌های پویای نرم‌افزار استفاده می‌کند [۶]. با تحلیل پویای رفتار نرم‌افزار در زمان اجرا، می‌توان اثرات جهش‌های معادل را با جهش‌های اصلی مقایسه کرد. به عنوان مثال، اگر یک جهش معادل

در دنیای امروزی پر از نرم‌افزارهای پیچیده، اطمینان از کیفیت و عملکرد صحیح آنها امری بسیار حیاتی است. هر گونه خطا یا نقص در یک نرم‌افزار می‌تواند عواقب جبران‌ناپذیری را به دنبال داشته باشد، از دست دادن اعتماد کاربران و مشتریان گرفته تا خسارات مالی و حتی اجتماعی. در این راستا، آزمون جهش به عنوان یکی از روش‌های موثر برای ارزیابی کیفیت آزمون‌های نرم‌افزاری به ویژه برنامه‌های حیاتی و حساس مورد توجه قرار گرفته است [۲۱]. آزمون جهش در اصل با تولید نسخه‌های متغیر از کد منبع (جهش‌های مصنوعی) و سپس اجرای آزمون‌ها روی این نسخه‌های جهش یافته، سعی در بررسی توانایی آزمون‌ها برای شناسایی خطاهای موجود دارد. برای اعمال این خطاهای هوشمند، یک سری عملگر بنام عملگرهای جهشی بکار گرفته می‌شوند تا با ایجاد تغییرات نحوی کوچک، سعی در تغییر برنامه اصلی داشته باشند. هر نسخه جدیدی از برنامه اصلی که ایجاد می‌شود، برنامه جهش نام دارد. به عنوان مثال، عبارت $x+y$ را در نظر بگیرید. این عبارت می‌تواند به عنوان یک عبارت جهشی مورد آزمون قرار بگیرد. جدول ۱، چند نمونه عملگر جهشی را نشان می‌دهد.

جدول ۱. عملگرهای جهشی پرکاربرد [۷]

تعریف عملگر	نام عملگر
$\text{remove}(x) \mid x \in \{++, --\} \{ \{x, \dots\} \}$	AODS: Shortcut Arithmetic Operator Deletion
$\{(-v, v)\}$	AODU: Unary Arithmetic Operator Deletion
$\{(v, --v), (v, v--), (v, ++v), (v, v++)\}$	AOIS: Shortcut Arithmetic Operator Insertion
$\{(v, -v)\}$	AOIU: Unary Arithmetic Operator Insertion
$\{(x,y) \mid x,y \in \{+, -, *, /, \% \} \wedge x \neq y\}$	AORB: Binary Arithmetic Operator Replacement
$\{(x,y) \mid x,y \in \{++, --\} \wedge x \neq y\}$	AORS: Shortcut Arithmetic Operator Replacement
$\{(x,y) \mid x,y \in \{+,, -, =, *, /, \% \} \wedge x \neq y\}$	ASRS: Shortcut Assignment Operator Replacement
$\{(op\ c, \text{remove}(op\ c)) \mid op \in \{+, -, *, /, \% \}, >, >=, <, <= \}$	CDL: Constant Deletion
$\{(!e), e \mid e \in \{\text{if}(e), \text{while}(e), \text{for}(s; e; s)\}\}$	COD: Conditional Operator Deletion
$\{(e, !e) \mid e \in \{\text{if}(e), \text{while}(e), \text{for}(s; e; s)\}\}$	COI: Conditional Operator Insertion
$\{(x,y) \mid x,y \in \{\&\&, \ \, \wedge\} \wedge x \neq y\}$	COR: Conditional Operator Replacement
$\{(v, \sim v)\}$	LOI: Logical Operator Insertion
$\{(v\ op, \text{remove}(v\ op)), (op\ v, \text{remove}(op\ v)) \mid op \in \{+, -, *, /, \% \}, <, <=, >, >= \}, \{(v++, v), (v--, v), (-v, v), (++, v)\} \mid op \in \{++, --\}\}$	ODL: Operator Deletion
$\{(x,y) \mid x,y \in \{>, >=, <, <=, ==, !=\} \wedge x \neq y\}$	ROR: Relational Operator Replacement
$\{(s, \text{remove}(s))\}$	SDL: Statement Deletion
$\{(v\ [op], \text{remove}(v\ [op])) \mid op \in \{+, -, *, /, \% \}, ++, --, <, <=, >, >= \}$	VDL: Variable Deletion

تحلیل‌ها می‌تواند به شناسایی جهش‌های معادل کمک کند و از طریق آن‌ها می‌توان به بهبود آزمون جهش و افزایش صحت و کارایی آن دست یافت [۱۳، ۱۴].

۲-۴- روش مبتنی بر تحلیل کد

این روش بر اساس تحلیل ساختار و کد منبع نرم‌افزار استوار است [۱۵]. با تحلیل و مقایسه کدهای مربوط به جهش‌های معادل و جهش‌های اصلی، می‌توان جهش‌های معادل را شناسایی کرد. به عنوان مثال، اگر کدهای منبع مربوط به دو جهش با یکدیگر تطابق داشته باشند، این ممکن است نشان‌دهنده معادل بودن آنها باشد. این تطابق ممکن است به مواردی اشاره کند که علاوه بر تغییراتی که جهش‌ها ایجاد کرده‌اند، تفاوتی در کد اصلی و کد معادل وجود ندارد [۱۶]. علاوه بر تطابق کدها، در این روش می‌توان به دنبال تحلیل تفاوت‌های جزئی در کدها بود که ممکن است نشان‌دهنده جهش‌های معادل باشد. این تفاوت‌ها می‌توانند شامل اختلاف در نحوه پیاده‌سازی یک قسمت از کد، محل یا روش فراخوانی یک تابع، یا سایر جزئیات فنی باشند [۱۷، ۱۸].

۲-۵- روش مبتنی بر تحلیل تاریخچه اجرا

این روش بر اساس تحلیل تاریخچه اجرای برنامه و نتایج آزمون‌ها استوار است. با مشاهده و تحلیل تاریخچه اجرای برنامه و تفاوت‌های آن بین جهش‌های معادل و جهش‌های اصلی، می‌توان این جهش‌ها را شناسایی کرد. به طور کلی، این روش شامل مراحل زیر می‌شود:

- تهیه آزمون‌ها: ابتدا، آزمون‌هایی طراحی می‌شود که برای بررسی عملکرد و صحت برنامه مورد استفاده قرار می‌گیرند. این آزمون‌ها می‌توانند از انواع مختلفی باشند که بسته به نیازهای آزمون معین می‌شوند.
- اجرای آزمون‌ها: در این مرحله، آزمون‌ها روی برنامه اجرا می‌شوند و نتایج اجرا ثبت می‌شود. این نتایج می‌توانند شامل خروجی‌های برنامه، اطلاعات لاگ، متریک‌های عملکرد، و غیره باشند.
- تحلیل نتایج: در این مرحله، نتایج اجرا و آزمون‌ها مورد بررسی قرار می‌گیرند. با مقایسه نتایج مربوط به جهش‌های اصلی با نتایج مربوط به جهش‌های معادل، تفاوت‌ها و الگوهای خاصی که ممکن است به نشان دادن وجود جهش‌های معادل بیانجامد، شناسایی می‌شود.
- شناسایی جهش‌های معادل: با توجه به تفاوت‌ها و الگوهای که در نتایج مشاهده می‌شود، جهش‌های معادل ممکن است شناسایی شوند. این تفاوت‌ها ممکن است شامل تغییرات در

توسط یک آزمون شناسایی نشود، اما در زمان اجرا و تحلیل پویا اثرات مشابهی با جهش اصلی داشته باشد، این ممکن است نشان‌دهنده معادل بودن آن باشد. به عبارت دیگر، اگر جهش معادل در زمان اجرا همان رفتار یا تأثیرات مشابهی که جهش اصلی ایجاد می‌کند، داشته باشد، می‌توان آن را به عنوان جهش معادل شناسایی کرد. میزان انتشار خطا توسط دستورالعمل‌ها نیز برای شناسایی جهش‌های معادل، موثر است [۱۷].

۲-۲- روش مبتنی بر مدل‌سازی فرآیند

این روش بر اساس مدل‌سازی دقیق فرآیند تولید و اجرای جهش‌ها استوار است [۸]. با تحلیل این فرآیند، جهش‌های معادل را می‌توان با دقت شناسایی کرد. به عنوان مثال، با مدل‌سازی جزئیاتی از فرآیند تولید جهش‌ها و شناسایی الگوهای خاص، می‌توان جهش‌های معادل را پیدا کرد این مدل‌سازی ممکن است شامل مراحل مختلف فرآیند تولید جهش‌ها، از جمله ایجاد جهش‌ها، اعمال جهش‌ها بر روی کد، اجرای آزمون‌ها، و تحلیل نتایج آزمون‌ها باشد [۹]. با تحلیل این فرآیند به دقت، الگوهایی که به وجود جهش‌های معادل اشاره می‌کنند، شناسایی می‌شوند. به عنوان مثال، اگر در مدل‌سازی فرآیند تولید جهش‌ها الگوهایی مشخص برای جهش‌های معادل شناسایی شوند، می‌توان از این الگوها برای شناسایی جهش‌های معادل استفاده کرد [۱۰]. این الگوها ممکن است شامل ترتیب اجرای جهش‌ها، نحوه تولید و اعمال جهش‌ها، و تأثیر آنها بر رفتار برنامه باشد. با استفاده از این روش، می‌توان به طور دقیق و کارآمد جهش‌های معادل را شناسایی کرده و از این طریق به بهبود آزمون جهش و افزایش کیفیت و قابلیت اطمینان برنامه کمک کرد [۱۱].

۲-۳- روش مبتنی بر تحلیل خروجی آزمون

این روش بر اساس تحلیل و مقایسه نتایج آزمون‌ها برای جهش‌های معادل و جهش‌های اصلی تمرکز دارد. با تحلیل این تفاوت‌ها، می‌توان جهش‌های معادل را شناسایی کرد [۱۲]. برای مثال، اگر خروجی آزمون برای یک جهش معادل با خروجی آزمون برای جهش اصلی تفاوتی نداشته باشد، ممکن است این جهش معادل باشد. به عنوان مثال، اگر خروجی یک آزمون برای یک جهش معادل با خروجی آزمون مربوط به جهش اصلی تفاوتی نداشته باشد، این ممکن است نشان‌دهنده جهش معادل باشد. در این صورت، معادل بودن دو جهش با توجه به نتایج آزمون مشخص می‌شود. در این روش، ممکن است از معیارهای مختلفی برای تحلیل نتایج آزمون استفاده شود، از جمله مقایسه خروجی‌های آزمون، مقایسه زمان اجرا، تحلیل تغییرات در رفتار برنامه و سایر معیارهای مشابه. این

خروجی برنامه، الگوهای خاص در تاریخچه اجرا، یا میزان تاثیر جهش بر عملکرد برنامه باشد [۱۹،۲۰].

۲-۶- روش مبتنی بر تحلیل ویژگی‌های جهش

روش‌های مبتنی بر تحلیل ویژگی‌های جهش به منظور شناسایی جهش‌های معادل، از تحلیل و مقایسه ویژگی‌های خاصی که جهش‌ها دارند، استفاده می‌کنند [۲۱]. این ویژگی‌ها می‌توانند ویژگی‌های کد، ساختار، یا رفتاری از جهش‌ها باشند. این روش‌ها بر اساس فرضیه‌ای عمل می‌کنند که جهش‌های معادل باید ویژگی‌های خاصی را با جهش‌های اصلی به اشتراک داشته باشند. این روش بر اساس تحلیل ویژگی‌های جهش و تفاوت‌های آنها با جهش‌های اصلی تمرکز دارد [۲۲]. با مشاهده و تحلیل ویژگی‌های خاصی که جهش‌های معادل از جهش‌های اصلی دارند، می‌توان این جهش‌ها را شناسایی کرد [۲۳]. به عنوان مثال، اگر یک جهش معادل ویژگی‌های خاصی مانند تغییرات محلی در کد را نشان دهد، ممکن است معادل بودن آن را نشان دهد [۱۵].

۳- روش پیشنهادی

در بخش پیشین به روش‌های مختلف شناسایی و حذف جهش‌های معادل پرداخته شده است. در این بخش، براساس تحقیق انجام گرفته در [۷]، تحلیل ایستا و پویای کد برنامه انجام گرفته و در نتیجه آن، دستورالعمل‌هایی که نسبت به نرخ انتشار خطا، اثربخشی کمتری دارند شناسایی شده است. دستورالعمل‌هایی که نرخ توزیع خطای کمتری دارند، با احتمال بسیار بالایی جزء جهش‌های معادل طبقه‌بندی شده‌اند. یعنی هرچقدر تاثیرپذیری دستورالعمل کمتر باشد، جهش‌های تزریق شده به آن دستورالعمل‌ها، معادل و بلااثر خواهند بود. برطبق این نتیجه گیری، ما در این مقاله به بررسی جهش‌های ایجاد شده در دستورالعمل‌های سطح پایین پرداخته‌ایم. به این صورت که دستورالعمل‌های سطح پایین، احتمال اینکه جهش تزریق شده به آن نوع بلااثر باشد مورد هدف این مقاله قرار گرفته است.

در این روش، ابتدا تمامی دستورالعمل‌های سطح پایین موجود در برنامه‌های محک شناسایی شده و سپس تمامی جهش‌هایی که بر روی آن دستورالعمل‌ها اعمال شده‌اند، مورد بررسی قرار گرفته‌اند. شناساس این نوع دستورالعمل‌ها با کمک الگوریتم‌های یادگیری ماشین انجام شده است. هنگامی که جهشی بر روی دستورالعملی اعمال می‌شود چند حالت ممکن است اتفاق بیفتد: اول این که ممکن است آن جهش توسط موردآزمون‌های طراحی شده شناسایی شود که به این حالت می‌گوییم جهش کشته شده است. حالت دوم

زمانی رخ می‌دهد که، جهش تزریق شده توسط هیچکدام از مورد آزمون‌های موجود در بستر آزمون طراحی شده شناسایی نشود. به این حالت گوییم جهش زنده مانده است. بر طبق رابطه ۱ که امتیاز جهش را نشان می‌دهد، هدف از آزمون جهش، بیشینه کردن مقدار امتیاز جهش با شناسایی تمام جهش‌های قابل کشتن و به حداقل رساندن جهش‌های معادل است.

$$Mutation\ Score = \frac{Kill\ Mutants}{All\ Mutants - Equivalent\ Mutants} \quad (1)$$

در ابتدا لازم است یک سری تعاریف از مشخصه‌های جهش‌ها را توضیح دهیم.

Mutant Severity: شدت جهش به میزان پتانسیل تأثیر یا جدیت یک جهش بر رفتار یا عملکرد یک نرم‌افزار اشاره دارد. این موضوع، درجه اهمیت یک جهش را از نظر قابلیت ایجاد خطاها یا اشتباهات در کد بررسی می‌کند. شدت یک جهش می‌تواند بر اساس عواملی مانند نوع جهش، مکان جهش و زمینه کد مورد نظر متغیر باشد. در زیر چند نکته اساسی در تعریف شدت جهش آمده است:

نوع جهش: اپراتورهای جهش مختلف درجات متفاوتی از شدت دارند. به عنوان مثال، جهش‌هایی که شامل تغییر اپراتورهای حسابی هستند، احتمالاً شدت کمتری نسبت به جهش‌هایی دارند که شامل تغییرات در بیانیه‌های شرطی یا ساختارهای کنترل حلقه می‌شوند. **تأثیر بر رفتار برنامه:** جهش‌هایی که بر رفتار برنامه تأثیر معنی‌داری دارند، مانند تغییر جریان منطقی یا ایجاد خطاهای معنی‌دار، نسبت به جهش‌هایی با تأثیر کمتر، شدت بیشتری دارند. **پتانسیل ایجاد خطا:** جهش‌هایی که احتمال بالاتری برای وارد کردن خطاها یا اشتباهات به کد دارند، معمولاً به عنوان جهش‌هایی با شدت بیشتر محسوب می‌شوند. این شامل جهش‌هایی است که نقاط تصمیم‌گیری اصلی را تغییر می‌دهند یا سازوکارهای دستورات اشتباه را تغییر می‌دهند.

سهولت شناسایی: جهش‌هایی که با استفاده از آزمون‌ها یا مرور کد به راحتی قابل شناسایی نیستند، ممکن است به عنوان جهش‌هایی با شدت بالاتر محسوب شوند، زیرا احتمالاً در طول اجرا یا برخورد با ورودی‌های غیرمنتظره به شکل اشکال به نمایش خواهند گذاشت.

از آنجایی که رابطه استاندارد برای محاسبه شدت جهش وجود ندارد، مفهوم شدت جهش اغلب وابسته به زمینه است. این امر، شامل ارزیابی عوامل مختلف مرتبط با تأثیر بالقوه یا جدیت یک جهش بر روی سیستم نرم‌افزاری است. با این حال، ما می‌توانیم یک رابطه مفهومی را تعریف کنیم که این عوامل را در خود جای دهد. شایان ذکر است که این رابطه ممکن است براساس زمینه و الزامات

شدت کلی جهش بر اساس ترکیب وزنی عوامل است. نمرات شدت بالاتر نشان دهنده جهش‌هایی است که شدیدتر هستند و به طور بالقوه تأثیر بیشتری بر روی برنامه دارند.

Mutant Diversity تفاوت جهش، شامل ارزیابی تفاوت آن با سایر جهش‌های مجموعه با استفاده از فاکتور فاصله است. رابطه ۳، رابطه تفاوت جهش را نشان می‌دهد. در اینجا نحوه محاسبه تنوع برای هر جهش آمده است:

متریک فاصله: متریک فاصله، تفاوت بین جهش‌ها را کمیت می‌کند و می‌تواند بر اساس عوامل مختلفی مانند تعداد عبارات جهش یافته، نوع عملگر جهش اعمال شده یا هر عامل مرتبط دیگری باشد. متریک فاصله به صورت $d(m_i, m_j)$ نشان داده می‌شود که در آن، m_i و m_j دو جهش هستند و d ، فاصله بین این دو جهش را نشان می‌دهد. برای هر جهش m_i تفاوت جهش با تمامی جهش‌های دیگر موجود را با کمک متریک فاصله، محاسبه می‌کنیم. تفاوت جهش m_i را با $div(m_i)$ نشان می‌دهیم و به صورت زیر محاسبه می‌شود:

$$div(m_i) = \sum_{j \neq i} (m_i, m_j) \quad (3)$$

رابطه ۳، مجموع فواصل بین جهش m_i و تمام جهش‌های موجود در مجموعه جهش‌ها را محاسبه می‌کند.

با یک مثال، فاکتور تفاوت جهش را محاسبه می‌کنیم. فرض کنید یک مجموعه جهش برای یک دستورالعمل مفروض به صورت $\{m_1, m_2, m_3\}$ داریم. می‌خواهیم، معیار تفاوت جهش را برای هر جهش بر اساس فاکتور فاصله بین جهش‌ها محاسبه کنیم. در اینجا مجموعه‌ای فرضی از جهش‌ها به همراه تعداد عبارات جهشی برای هر جهش آمده است:

در ۵ عبارت جهشی ظاهر شده است: m_1

در ۳ عبارت جهشی ظاهر شده است: m_2

در ۷ عبارت جهشی ظاهر شده است: m_3

اکنون، تفاوت جهشی را برای هر جهش محاسبه می‌کنیم:

For m_1 :

Diversity of m_1 with respect to m_2 : $|5-3|=|5-3|=2$

Diversity of m_1 with respect to m_3 : $|5-7|=|5-7|=2$

Total diversity for m_1 : $2+2=4$

For m_2 :

Diversity of m_2 with respect to m_1 : $|3-5|=|3-5|=2$

Diversity of m_2 with respect to m_3 : $|3-7|=4$

Total diversity for m_2 : $2+4=6$

For m_3 :

Diversity of m_3 with respect to m_1 : $|7-5|=2$

Diversity of m_3 with respect to m_2 : $|7-3|=4$

Total diversity for m_3 : $2+4=6$

بنابراین، بر اساس متریک فاصله، تفاوت جهش، برای هر جهش به این صورت خواهد بود:

$Diversity(m_1) = 4$

خاص نرم‌افزار مورد آزمایش، نیاز به تغییراتی داشته باشد. رابطه پیشنهادی برای شدت جهش در اینجا آمده است:

$$MS = \sum_{i=1}^n (w_i \times f_i) \quad (2)$$

در رابطه ۲، n تعداد عوامل در نظر گرفته شده در ارزیابی شدت هر جهش است. w_i ، وزن‌های اختصاص داده شده به هر عامل است تا اهمیت نسبی آنها را منعکس کند. مجموع این وزن‌ها باید در نهایت، ۱ باشد. f_i امتیازاتی هستند که به هر عامل بر اساس سطح شدت آن اختصاص داده می‌شود. فاکتور f_i در رابطه در نظر گرفته شده، شامل موارد زیر است:

- Functional Impact: امتیازی که میزان تأثیر جهش بر عملکرد یا رفتار مورد نظر برنامه را نشان می‌دهد.
 - Fault Detection Difficulty: امتیازی که منعکس کننده دشواری تشخیص خطای معرفی شده توسط جهش با استفاده از مجموعه آزمایشی است.
 - Potential Consequences: امتیازی که نشان‌دهنده پیامدهای بالقوه خطا، مانند خرابی سیستم، خرابی داده‌ها یا آسیب‌پذیری‌های امنیتی است.
 - Frequency of Execution: امتیازی که تعداد دفعات اجرای کد جهش را در طول اجرای برنامه معمولی نشان می‌دهد.
- وزن‌های اختصاص داده شده به هر عامل باید اهمیت نسبی آنها را در فرآیند ارزیابی شدت جهش منعکس کند. این وزن‌ها را می‌توان بر اساس تحلیل تجربی تعیین کرد. با جمع‌بندی امتیازهای وزنی این عوامل، رابطه یک نمایش کمی از شدت جهش ارائه می‌کند و به ما این امکان را می‌دهد تا جهش‌های تولید شده را بر اساس سطح شدت جهش‌ها اولویت‌بندی کنیم.

در اینجا یک مثال عددی برای رابطه پیشنهادی برای شدت جهش نشان می‌دهیم. برای سادگی، اجازه دهید سناریویی را با سه عامل در نظر بگیریم: تأثیر عملکردی (FI)، مشکل تشخیص خطا (FDD) و پیامدهای بالقوه (PC). فرض کنید برای هر عامل به صورت زیر وزن تعیین می‌کنیم:

$$w_{FI}=0.4 (40\%), w_{FDD}=0.3 (30\%), w_{PC}=0.3 (30\%)$$

همچنین فرض کنیم که هر عامل را در مقیاس ۱ تا ۱۰ درجه‌بندی می‌کنیم که ۱۰ شدیدترین آنها است:

Functional Impact (FI): 8

Fault Detection Difficulty (FDD): 5

Potential Consequences (PC): 7

حال شدت جهش (MS) را با استفاده از رابطه ۲ محاسبه کنیم:

$$MS = (w_{FI} \times f_{FI}) + (w_{FDD} \times f_{FDD}) + (w_{PC} \times f_{PC})$$

$$MS = (0.4 \times 8) + (0.3 \times 5) + (0.3 \times 7)$$

$$MS = (3.2) + (1.5) + (2.1)$$

$$MS = 6.8$$

در این مثال، امتیاز شدت جهش ۶٫۸ است. این امتیاز نشان‌دهنده

محاسبه کنیم:

$$MV = (W_{Scope} \times S_{Scope}) + (W_{Change} \times S_{Change}) + (W_{Readability} \times S_{Readability})$$

$$= (0.33 \times 5) + (0.33 \times 7) + (0.33 \times 8) = (0.33 \times 5) + (0.33 \times 7) + (0.33 \times 8)$$

$$= (1.65) + (2.31) + (2.64) = (1.65) + (2.31) + (2.64)$$

$$\approx 6.60$$

Mutant Complexity

در آزمون جهش، پیچیدگی جهش به سطح دشواری یا پیچیدگی مرتبط با ایجاد جهش در کد برنامه اشاره دارد. آزمون جهش، تکنیکی است که برای ارزیابی کیفیت مجموعه‌های آزمایشی با وارد کردن تغییرات کوچک به نام جهش در کد منبع و سپس اجرای مجموعه آزمایشی برای مشاهده اینکه آیا هر یک از جهش‌ها شناسایی شده‌اند، استفاده می‌شود. تکه کد محاسبه مساحت زیر را در نظر بگیرید.

```
public static double calculateArea (double length,
double width) {
    return length * width;
}
```

حال، فرض کنید می‌خواهیم جهش‌ها را برای اهداف آزمایش جهش وارد این تابع کنیم. جهش‌ها با ایجاد تغییرات کوچک در کد اصلی ایجاد می‌شوند. در مثال خود، جهش‌ها را در نقاط مختلف تابع محاسبه مساحت معرفی می‌کنیم و پیچیدگی آن‌ها را با استفاده از معادله ارائه شده در رابطه ۶ مطرح می‌کنیم:

$$MCom = w_1 \times I + w_2 \times C + w_3 \times D \quad (6)$$

برای سادگی، سه عامل را در نظر بگیریم که در پیچیدگی جهش نقش دارند:

- *Importance of the instruction (I)* دستورالعمل برای عملکرد کلی برنامه چقدر مهم است؟
- *Complexity of the instruction (C)*: دستورالعمل چقدر پیچیده است؟
- *Dependency level of the instruction (D)* دستورالعمل چقدر به قسمت‌های دیگر کد وابسته است؟ حال به این عوامل وزن اختصاص می‌دهیم:

$$w_1 = 0.5 \text{ for importance,}$$

$$w_2 = 0.3 \text{ for complexity,}$$

$$w_3 = 0.2 \text{ for dependency level}$$

در اینجا نحوه محاسبه پیچیدگی جهش ($MCom$) برای هر جهش آمده است:

برای تابع اصلی $I=1$. (تابع اصلی برای محاسبه ضروری است). $C=1$, (عملیات ضرب ساده) و $D=1$ (بدون وابستگی به قسمت‌های دیگر کد). تابع اصلی دارای پیچیدگی جهش ۱ است. جهش شماره ۱: تغییر عملیات ضرب به جمع

$$Diversity(m_2) = 6$$

$$Diversity(m_3) = 6$$

Mutant Coverage پوشش جهش، براساس رابطه ۴ محاسبه می‌شود.

$$Mutant Coverage \text{ for One Mutant} = (\text{Total Number of Test cases in the Test Suite} / \text{Number of Test Cases that Detect the Mutant}) \times 100 \quad (4)$$

به عنوان مثال، اگر مجموعه آزمایشی شما از ۵۰ مورد آزمون تشکیل شده باشد و ۱۰ مورد از این موارد آزمایش، جهش خاصی مانند m_i را که مدنظر دارید شناسایی کند، پوشش جهش برای آن جهش خواهد بود:

$$Mutant Coverage \text{ for } m_i = (10/50) / 100\%$$

مقدار ۲۰٪ به این معنی است که ۲۰٪ از موارد آزمون در مجموعه آزمایشی، جهش خاص مدنظر را تشخیص می‌دهند.

Mutant Visibility: ارزیابی نمایان بودن یک جهش، به معنای این است که چقدر یک جهش قابل مشاهده یا قابل تشخیص توسط مجموعه آزمون است. در واقع، نمایان بودن یک جهش نشان دهنده این است که آیا تغییر معنایی که توسط جهش ایجاد شده است، توسط مجموعه آزمون تشخیص داده می‌شود یا خیر. رابطه ۵، نمایان بودن جهش را نشان می‌دهد.

$$MV = (W_{Scope} \times S_{Scope}) + (W_{Change} \times S_{Change}) + (W_{Readability} \times S_{Readability}) \quad (5)$$

در رابطه ۵، میزان رویت یک جهش را با استفاده از یک مثال ساده نشان می‌دهیم. فرض کنید ما در حال ارزیابی نمایان بودن یک جهش هستیم که شامل تغییر نام متغیر در یک تابع است. ما سه عامل را در نظر خواهیم گرفت: محدوده جهش، ماهیت تغییر و خوانایی کد. محدوده جهش (*Scope*)، دامنه جهش را ۵ از ۱۰ رتبه‌بندی می‌کنیم. این تغییر تنها بر یک متغیر منفرد در یک تابع تأثیر می‌گذارد که در مقایسه با تغییر کل تابع یا کلاس، دامنه نسبتاً کوچکی است. ماهیت تغییر (*change*) را ۷ از ۱۰ ارزیابی می‌کنیم. تغییر نام متغیر یک تغییر نسبتاً قابل توجه است، زیرا نحو کد برنامه را بدون تغییر اساسی منطق یا رفتار آن تغییر می‌دهد. خوانایی کد (*Readability*) را ۸ از ۱۰ رتبه‌بندی می‌کنیم. کد جهش یافته واضح و قابل درک باقی می‌ماند، زیرا تغییر فقط شامل یک نام متغیر است و پیچیدگی یا سردرگمی ایجاد نمی‌کند. در مرحله بعد، به هر عامل وزنی اختصاص می‌دهیم تا اهمیت نسبی آنها را منعکس کند. بیا بید وزن‌های مساوی را برای سادگی فرض کنیم:

$$W_{Scope} = 0.33 \text{ (33\%)}$$

$$W_{Change} = 0.33 \text{ (33\%)}$$

$$W_{Readability} = 0.33 \text{ (33\%)}$$

اکنون می‌توانیم میزان نمایان بودن جهش را با استفاده از رابطه ۵

I=0.8: (هنوز مهم است اما کمتر از عملکرد اصلی مهم است).

C=1: عملیات حسابی ساده

D=1: بدون وابستگی

$$MCom = w_I \times I + w_C \times C + w_D \times D$$

$$MCom = 0.5 \times 0.8 + 0.3 \times 1 + 0.2 \times 1 = 0.8 \times 0.5 + 1 \times 0.3 + 1 \times 0.2 = 0.5 + 0.3 + 0.2 = 1$$

جهش شماره ۱، دارای پیچیدگی جهش ۱ است، زیرا تغییری را معرفی می‌کند که بر عملکرد حیاتی تابع تأثیر می‌گذارد.

جهش شماره ۲: اضافه کردن یک عبارت شرطی

I=0.7: (بسیار مهم، اما کمتر از نسخه اصلی)

C=2: (افزایش پیچیدگی به دلیل بیان شرطی)

D=1: بدون وابستگی

$$MCom = 0.5 \times 0.7 + 0.3 \times 2 + 0.2 \times 1 = 0.7 \times 0.5 + 2 \times 0.3 + 1 \times 0.2 = 0.35 + 0.6 + 0.2 = 1.15$$

جهش شماره ۲ دارای پیچیدگی جهش یافته ۱،۱۵ است که نشان دهنده افزایش پیچیدگی معرفی شده توسط عبارت شرطی است.

Mutant Impact

تأثیر جهش به میزان اثرگذاری جهش در کد برنامه بر روی رفتار برنامه اشاره دارد. هر جهش نشان دهنده تغییر کوچکی است که در کد منبع ایجاد شده است و تأثیر این تغییر با مشاهده چگونگی تأثیر آن بر نتیجه مجموعه مورد آزمون‌ها ارزیابی می‌شود. در برنامه زیر، یک جهش با تغییر عملگر ضرب به جمع ایجاد می‌کنیم. حال میزان تأثیر جهش در رابطه ۷ رل محاسبه می‌کنیم.

```
public class FactorialCalculator {
    public int calculateFactorial(int n) {
        if (n == 0)
            return 1;
        else
            return n * calculateFactorial(n - 1);
    }
}
```

$$MI \text{ (Mutant Impact)} = FI \times (1 - CD) \times CC \quad (7)$$

به طوری که:

MI is the mutation impact.

FI is the functionality impact

CD is the fault detection difficulty.

CC is the code coverage affected by the mutation.

حال برای مثال مفروض، مقادیری را به این فاکتورها اختصاص دهیم:

$$FI = 1$$

به این معناست که جهش به‌طور کامل، عملکرد روش را تغییر می‌دهد

$$CD = 0.8$$

به این معنا که تشخیص خطای ایجاد شده توسط جهش به راحتی

انجام نمی‌پذیرد.

$$CC = 20\%$$

به این معناست که جهش تنها بخش کوچکی از کد را تحت تأثیر قرار می‌دهد، زیرا محدود به فراخوانی بازگشتی است.

حالا این مقادیر را به فرمول اضافه می‌کنیم:

$$MI = 1 \times (1 - 0.8) \times 0.2 = 0.2 \times 0.2 = 0.04$$

در این مثال، تأثیر جهش ۰،۰۴ است. با توجه به اینکه جهش به طور قابل توجهی رفتار روش را تغییر می‌دهد، نشان‌دهنده تأثیر متوسط است، اما تنها بخش کوچکی از کد را تحت تأثیر قرار می‌دهد و ممکن است شناسایی از طریق آزمون تا حدودی چالش برانگیز باشد.

Mutant Coupling

به میزان وابستگی متقابل یا تعامل بین جهش‌ها در یک برنامه نرم‌افزاری در طول آزمون جهش اشاره دارد. این فاکتور، تغییرات ایجاد شده در یک جهش که بر رفتار، تشخیص یا اثربخشی سایر جهش یافته‌ها تأثیر می‌گذارد را اندازه می‌گیرد. به عبارت دیگر، میزان همبستگی جهش‌ها با یکدیگر را از نظر تأثیر آنها در آزمایش ارزیابی می‌کند. بیابید Mutant coupling را با یک مثال جاوا نشان دهیم و سپس یک فرمول مفهومی برای محاسبه آن ارائه کنیم.

یک کلاس جاوا ساده را در نظر بگیرید که روشی را برای محاسبه مربع یک عدد پیاده سازی می‌کند:

```
public class SquareCalculator {
    public int square(int x) {
        return x * x;
    }
}
```

حال، فرض کنید دو جهش را به این روش معرفی می‌کنیم:

Mutant 1: عمل ضرب (×) را به جمع (+) تغییر می‌دهد.

Mutant 2: به جای ضرب، یک عمل تفریق (-) را معرفی می‌کند. در

اینجا نحوه اعمال این جهش‌ها آمده است:

Mutants 1

```
public int square(int x) {
    return x + x;
}
```

Mutant 2:

```
public int square(int x) {
    return x - x;
}
```

حال، بیابید تأثیر تشخیص یک جهش بر دیگری را تجزیه و تحلیل کنیم. فرض کنید مجموعه آزمایشی ما برای تشخیص انحرافات از رفتار صحیح مربع کردن یک عدد طراحی شده است. اگر مجموعه آزمایشی به طور موثر جهش ۱ (عملیات جمع) را تشخیص دهد اما نتواند جهش ۲ (عملیات تفریق) را شناسایی کند، ما جفت جهش را مشاهده می‌کنیم. در این مورد، تشخیص جهش ۱ بر تشخیص یا

سطح دستورالعمل‌های بلاثر پیشنهاد شده است تا بتوان بر این اساس، جهش‌های معادل را از جهش‌های سرسخت جدا نمود. همانطور که در بخش قبل، درمورد ویژگی‌های جهش‌ها توضیح داده شده است، هر کدام از این مشخصه‌ها به نوعی، تعیین کننده میزان اهمیت، درجه سختی، جفت شدن و سایر مشخصه‌های مرتبط با جهش است. با توجه به اینکه، جهش‌های تزریق شده به دستورالعمل‌های سطح D، با احتمال بالاتری، جهش معادل هستند، در این مقاله تنها به بررسی دستورالعمل‌های سطح D شناسایی شده می‌پردازیم. درواقع دستورالعمل‌های سطوح بالاتر در این بخش مورد تحلیل قرار نمی‌گیرند. بدین ترتیب که برای هر برنامه محک که شامل تعدادی دستورالعمل سطح پایین است، تمام جهش‌های تزریق شده به آنها را استخراج نموده و در جدولی، مشخصه‌های مربوط به آن جهش‌ها را محاسبه کرده و به صورت یک مجموعه داده آماده کرده‌ایم. بخشی از این مجموعه داده در جدول ۲ نشان داده شده است.

شکل ۱، ماتریس همبستگی ویژگی‌های جهش‌ها را نشان می‌دهد. طبق شکل، همبستگی بین ویژگی‌ها در حد قابل قبولی است.

جدول ۲. بخشی از مجموعه داده ایجاد شده

Rank	Location	Diversity	Coverage	Visibility	Severity	Type	Coupling	Complexity	ID	Instruction
1	0	4	80%	9	2	2	50%	0.4	M1	Inst1
1	0	5	70%	5	1	4	20%	0.6	M2	Inst1
0	1	3	90%	7	4	2	10%	0.8	M3	Inst1
1	0	8	80%	2	3	1	70%	0.2	M1	Inst2
0	-1	2	80%	4	4	4	60%	0.4	M2	Inst2
0	-1	4	70%	1	2	3	20%	0.7	M1	Inst3
0	0	5	60%	5	5	1	70%	0.5	M2	Inst3
1	0	6	40%	6	3	3	20%	0.3	M3	Inst3
1	0	4	50%	3	2	3	40%	0.6	M1	Inst4
0	1	2	70%	7	4	2	50%	0.2	M2	Inst4

اثربخشی جهش ۲ تأثیر می‌گذارد، که نشان دهنده درجه‌ای از وابستگی متقابل بین جهش‌ها است. از آنجایی که فرمول استاندارد برای محاسبه Mutant Coupling وجود ندارد، می‌توانیم آن را به صورت مفهومی به صورت رابطه ۸ نمایش دهید:

$$MC = \frac{N_{\text{detected_together}}}{N_{\text{detected_alone}}} \quad (8)$$

بطوری که:

$N_{\text{detected_together}}$ ، تعداد جهش‌هایی است که با هم شناسایی شده‌اند (یعنی هر دو جهش توسط یک مورد آزمایشی یا مجموعه‌ای از موارد آزمایشی شناسایی شده‌اند).

$N_{\text{detected_alone}}$ ، تعداد جهش‌هایی است که به تنهایی شناسایی می‌شوند (یعنی هر جهش توسط موارد آزمایش جداگانه شناسایی می‌شود). فرض کنید ما مجموعه‌ای از ۱۰ جهش (M1 تا M10) داریم و علاقه مند به محاسبه مقدار فاکتور Coupling برای M1 هستیم. جهش M1 توسط مورد آزمون A شناسایی می‌شود. مورد آزمون A همچنین جهش M2، M3 و M4 را علاوه بر M1 شناسایی می‌کند. مجموعه جهش‌ها در کل از ۱۰ جهش تشکیل شده است. حالا بیا باید مقدار جفت شدن جهش M1 را محاسبه کنیم:

تعداد جهش‌های شناسایی شده توسط مورد آزمون A شامل M1:

۴

تعداد کل جهش یافته‌ها در مجموعه جهش: ۱۰

$$\text{Amount of Coupling} = \frac{\text{Total number of mutants}}{\text{Number of mutants detected by Test Case A}} \times 100\%$$

$$\text{Amount of Coupling} = \frac{4}{10} \times 100\% = 40\%$$

بنابراین، مقدار coupling برای جهش M1 برابر ۴۰٪ است. این نشان می‌دهد که ۴۰٪ از جهش‌ها در مجموعه جهش توسط همان مورد آزمایشی شناسایی می‌شوند که جهش M1 را شناسایی می‌کند.

۴- آزمایش‌ها

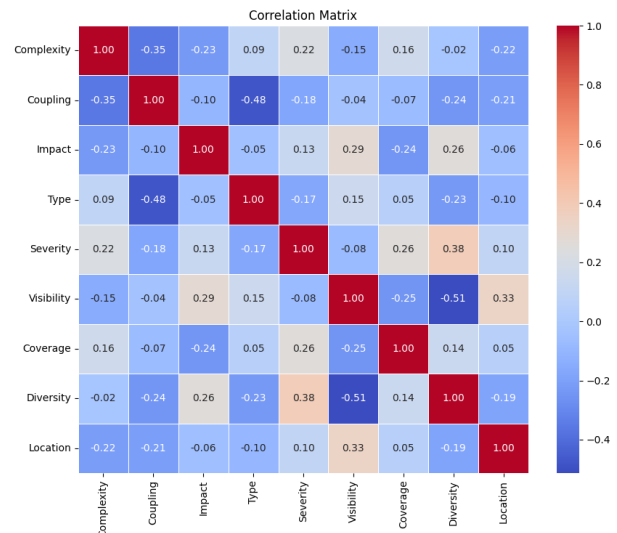
هدف از روش پیشنهادی، جداسازی جهش‌های معادل از جهش‌های سرسخت با توجه به ویژگی‌های جهش‌ها است. به دلیل اینکه نتیجه اجرای برخی از مورد آزمون‌ها با نتیجه اجرای برنامه اصلی یکسان می‌باشد، آن جهش‌ها را در دسته جهش‌های معادل قرار می‌دهیم. اما در این بین، برخی از جهش‌های دیگری نیز به صورت اشتباهی به عنوان جهش‌های معادل ارزیابی می‌شوند. به آن دسته از جهش‌ها، جهش‌های سرسخت گفته می‌شود. شناسایی و جداسازی جهش‌های سرسخت از گروه جهش‌های معادل، نیاز به تلاش دستی زیاد و صرف هزینه و زمان زیادی است. برای همین منظور، روش پیشنهادی با استفاده از بررسی ویژگی‌های جهش‌های ایجاد شده در

پس از بررسی شرایط تشکیل مثلث، نوع مثلث بر اساس مقادیر ورودی تعیین می‌شود. یکی از انواع مثلث متساوی الاضلاع، متساوی الساقین و قائم الزاویه به‌عنوان خروجی نشان داده می‌شود. اگر شرط تشکیل مثلث برآورده نشود، خروجی یک پیغام خطایی با عنوان عدم تشکیل مثلث چاپ می‌کند. برای محاسبه فاکتوریل یک عدد صحیح از برنامه Factorial استفاده شده است. پس از بررسی شرایط محاسبه فاکتوریل عدد وارد شده، خروجی برنامه، نتیجه محاسبه فاکتوریل عدد وارد شده خواهد بود. برنامه Prime، با دریافت یک عدد از کاربر، اول بودن عدد را تعیین می‌کند. از برنامه Middle، برای محاسبه و نمایش میانه سه عدد استفاده می‌شود. در این برنامه، ورودی شامل سه عدد است که مقدار وسط سه عدد وارد شده تعیین شده و به‌عنوان نتیجه برگردانده می‌شود. ورودی برنامه FindMax، شامل مجموعه‌ای از اعداد است. تعداد این اعداد در ابتدا توسط کاربر مشخص می‌شود. خروجی این برنامه، عددی است که بیشینه مقدار اعداد ورودی را مشخص می‌کند. برنامه Bubblesort، لیستی از اعداد صحیح را می‌گیرد و آنها را به ترتیب صعودی با روش مرتب‌سازی حبابی مرتب می‌کند و در خروجی نمایش می‌دهد. در برنامه BubbleSort، مرتب‌سازی با ورودی انجام می‌شود و رشته مرتب شده در خروجی چاپ می‌شود. برنامه DOW شامل سه مقدار ورودی سال، ماه و روز بر اساس تاریخ میلادی است. بعد از دریافت این سه ورودی، روز هفته معادل با تاریخ شده را در خروجی چاپ می‌کند. به‌عنوان مثال، سال ۱۹۸۶ ماه ۴ روز ۱۸، معادل روز جمعه است. جمعه به‌عنوان خروجی چاپ می‌شود. جدول ۳، مشخصه‌های کلی این برنامه‌ها را نشان می‌دهد.

جدول ۳. مشخصات برنامه‌های محک

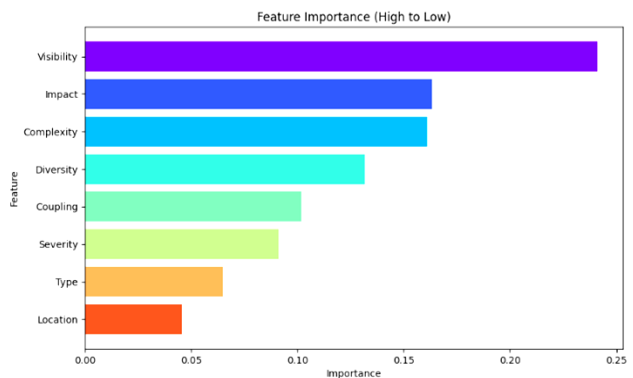
نام برنامه	اندازه خطوط برنامه	توصیف برنامه
Triangle	۲۸	تعیین نوع مثلث
Factorial	۲۴	تعیین مقدار فاکتوریل عدد ورودی
Mid	۱۴	تعیین عدد میانه سه عدد ورودی
FindMax	۳۰	یافتن بزرگترین مقدار
Prime	۳۳	تعیین اینکه آیا عدد ورودی، عدد اول است یا نه
Bubblesort	۲۱	مرتب‌سازی لیست ورودی به‌صورت حبابی
DOW	۵۶	تعیین روز هفته معادل با تاریخ ورودی

انتخاب مجموعه مورد آزمون‌ها یکی دیگر از عناصر اصلی تعیین‌کننده در افزایش کارایی روش پیشنهادی محسوب می‌شود. تولید تصادفی داده‌های آزمون، پوشش تمام دستورالعمل‌های برنامه را تضمین نمی‌کند. از این‌رو در آزمون‌های انجام شده باید پوشش تمامی دستورالعمل‌ها در نظر گرفته شود. در این روش، برای هر برنامه محک، مجموعه‌های مختلفی از داده‌های آزمون بر اساس معیارهای پوشش تهیه شده است. برای این منظور، گراف جریان



شکل ۱. ماتریس همبستگی ویژگی‌های جهش

شکل ۲، درجه اهمیت ویژگی‌های جهش‌ها را نشان می‌دهد. همانطور که در شکل دیده می‌شود، ویژگی Visibility بالاترین اهمیت را دارد.



شکل ۲. مقایسه درجه اهمیت ویژگی‌های جهش‌ها

۵- برنامه‌های محک

در این مقاله، هفت برنامه محک مختلف مورد بررسی و آزمایش قرار گرفته است. هرکدام از این برنامه‌ها در سطح واحد و یا تابع بررسی شده‌اند. چون اکثر برنامه‌های کاربردی دنیای واقعی، از توابع مختلف استفاده می‌کنند که همگی آنها در کنار هم، برنامه کاربردی دنیای واقعی را می‌سازد. بنابراین بررسی برنامه‌های کوچک در سطح واحد با همان پیچیدگی برنامه‌های واقعی، نتیجه معتبری را خواهد داشت. در واقع، توابع در میلیون‌ها خط‌کد وجود دارند که برنامه‌های کاربردی دنیای واقعی را تشکیل می‌دهند. اندازه استاندارد یک تابع در دنیای واقعی بین ۵ تا ۵۰ خط‌کد است. بر همین اساس، برنامه‌های محک انتخاب شده برای روش پیشنهادی عبارتند از: Triangle، Factorial، FindMax، Prime، Middle، Bubblesort و DOW. برای تعیین نوع مثلث از برنامه محک Triangle استفاده شده است. ورودی‌های این برنامه شامل سه عدد صحیح است که

```

public static int Numbers(int[]x)
{
    int n ;
    n=x.length;
    System.out.println(n+" is number of integers");
    int max=x[0];
    System.out.println(max+" is the first number ");
    for(int i = 1;i<n;i+=1)
    {
        if(max<x[i])
            max = x[i];
    }
    System.out.println("maximum of " +n + "
numbers is " +max);
    return max;
}

```

شکل ۳. برنامه FindMax در بستر Muclipse

شکل ۴ یکی از موردآزمون‌های طراحی شده برای اجرا بر روی برنامه محک FindMax را نشان می‌دهد. همان‌طور که مشاهده می‌کنید، مقادیر ورودی آزمون شامل مقادیر {753,564,87,235,123,78} است. مقدار مورد انتظار خروجی عدد ۷۵۳ است. مورد آزمون با مقادیر ورودی فوق، خروجی مد نظر را تولید کرده است.

دسته دوم از برنامه‌های طراحی شده با پسوند M در بخش Muclipse:Mutants قرار می‌گیرند و برنامه‌های جهشی تولید شده در این بخش قرار می‌گیرند. یک نمونه از نتیجه اجرای برنامه‌هایی که حاوی جهش‌های مرتبه اول هستند (هر برنامه تنها یک خطای جهشی دارد)، در شکل ۵ نشان داده شده است.

```

package maxx;
import static org.junit.Assert.*;
import org.junit.Test;
import org.junit.After;
import org.junit.Before;
import org.junit.Test;
import static org.junit.Assert.assertEquals;
import static org.junit.Assert.*;
import javax.lang.model.type.ArrayType;
import org.junit.Test;
import junit.framework.TestCase;
public class Max2Test extends TestCase{
    private Max2 c2;
    public void setUp()
    {
        c2 = new Max2();
    }
    @Test
    public void test1() {
        int value [] = {753,564,87,235,123,78};
        int expectedOrder =753 ;
        assertEquals(expectedOrder,c2.Numbers(value));
    }
}

```

شکل ۴. یک مورد آزمون طراحی شده برای برنامه FindMax

کنترلی هر برنامه محک ایجاد می‌شود. پوشش لبه به‌عنوان یک معیار مبتنی بر نمودار گراف جریان کنترلی برای تولید داده‌های آزمون استفاده می‌شود. پوشش تمام لبه‌ها در این گراف، پوشش کد منبع کل برنامه را تضمین می‌کند. مجموعه داده ایجاد شده در قالب یک فایل اکسل یا csv، توسط الگوریتم یادگیری ماشین در مجموعه ابزار RapidMiner استفاده شده است. بخش اصلی روش پیشنهادی، یک طبقه‌بندی‌کننده جهش است که با آموزش الگوریتم یادگیری ماشین ایجاد شده است. در این مطالعه، برای تولید برنامه‌های جهش (برنامه‌های خطادار)، از ابزاری بنام Muclipse استفاده شده است. Mujava، افزونه‌ای است که برای اجرای آزمون و ارزیابی امتیاز جهش در ابزار Eclipse گنجانده شده است. کار با ابزار Muclipse برای آزمایش روش پیشنهادی، شامل سه بخش اصلی است. بخش اول آزمایش، مربوط به اجرای مورد آزمون‌های طراحی شده برای هر برنامه محک است. در واقع از بین مورد آزمون‌های طراحی شده، هر بار یکی از موارد آزمون برای اجرا انتخاب خواهد شد و نتیجه آن ذخیره می‌شود. مجری این بخش از ابزار که همان Test runner است، مربوط به Junit است. دو نتیجه مختلف از اجرای مورد آزمون‌ها به دست می‌آید. در حالت اول، اجرای آزمون با موفقیت انجام می‌شود؛ یعنی مقدار مورد انتظار خروجی با نتیجه مورد آزمون تطابق دارد. حالت دوم این است که اجرای آزمون با شکست مواجه شود. در این حالت نسبت به اصلاح مورد آزمون طرح شده اقدام خواهیم کرد. مجری آزمون در این قسمت، Junit4 است. این برنامه‌ها، زیرمجموعه بخش Junit هستند. محتوای هر برنامه Junit، یک مورد آزمون از پنج مورد آزمونی است که شرط پوشش لبه در آن رعایت شده است. برنامه‌های مورد آزمون، در ابزار Muclipse باید با قالب خاصی طراحی شوند. برنامه مورد آزمون، شامل دستورالعمل مقادیر (value)، دستورالعمل مقدار مورد انتظار (expectedOrder)، دستورالعمل مقایسه مقدار حاصل از اجرای مورد آزمون‌ها و مقدار مورد انتظار (assertEquals) است. در این حالت، مورد آزمون مدنظر باید بتواند برنامه اصلی را به درستی آزمایش کند. اگر بتواند جهش مدنظر را پیدا کند که همان حالت کشتن جهش است، این جهش به لیست جهش‌های کشته شده اضافه می‌شود. اما اگر نتواند این جهش را کشف کند، گوییم جهش مدنظر زنده مانده است.

شکل ۳، کد برنامه FindMax را نشان می‌دهد که یکی از برنامه‌های محک انتخاب شده برای آزمون روش پیشنهادی است و به زبان جاوا نوشته شده است. این برنامه، در بستر Muclipse برای اجرای جهش‌ها نوشته شده است.

```

Analysis of test case
test 1 kill ==> AOIS_10 AOIS_11 AOIS_12 AOIS_14
AOIS_15 AOIS_16 AOIS_17 AOIS_19 AOIS_21 AOIS_22
AOIS_23 AOIS_24 AOIS_25 AOIS_26 AOIS_28 AOIS_37
AOIS_38 AOIS_5 AOIS_7 AOIS_9 AOIU_3 ASRS_1 ASRS_2
ASRS_3 ASRS_4 COI_2 LOI_11 LOI_4 LOI_6 LOI_7
-----
Live: 26
Dead: 30
Active: 0
-----

Results for class maxx.Max2
-----
Live mutants: 26
Killed mutants: 30
Mutation Score: 53.0
Writing to file...

```

شکل ۶. خروجی به‌دست‌آمده از اجرای برنامه‌های خطادار در برنامه

محک FindMax

```

AOIU_1:27:int_Numbers(int):x.length => -x.length
AOIU_2:33:int_Numbers(int):i => -i
AOIS_9:30:int_Numbers(int):max => ++max
AOIS_10:30:int_Numbers(int):max => --max
AOIS_11:30:int_Numbers(int):max => max++
AOIS_17:31:int_Numbers(int):i => ++n
AOIS_18:31:int_Numbers(int):n => --n
AOIS_19:31:int_Numbers(int):n => n++
AOIS_20:31:int_Numbers(int):n => n--
AOIS_21:32:int_Numbers(int):max => ++max
AOIS_22:32:int_Numbers(int):max => --max
AOIS_23:32:int_Numbers(int):max => max++
AOIS_24:32:int_Numbers(int):max => max--
AOIS_25:32:int_Numbers(int):i => ++i
AOIS_26:32:int_Numbers(int):i => --i
AOIS_27:32:int_Numbers(int):i => i++
AOIS_28:32:int_Numbers(int):i => i--
AOIS_29:33:int_Numbers(int):i => ++i
AOIS_30:33:int_Numbers(int):i => --i
AOIS_31:33:int_Numbers(int):i => i++
AOIS_32:33:int_Numbers(int):i => i--
AOIS_33:36:int_Numbers(int):n => ++n
AOIS_34:36:int_Numbers(int):n => --n
AOIS_35:36:int_Numbers(int):n => n++
AOIS_36:36:int_Numbers(int):n => n--
AOIS_37:36:int_Numbers(int):max => max++
AOIS_38:36:int_Numbers(int):max => max--
AOIS_39:37:int_Numbers(int):max => max++
AOIS_40:37:int_Numbers(int):max => max--
COI_1:31:int_Numbers(int): i < n => !(i < n)
COI_2:32:int_Numbers(int): max < x[i] => !(max < x[i])
LOI_1:27:int_Numbers(int):x.length => -x.length
LOI_2:28:int_Numbers(int):n => ~n
LOI_3:30:int_Numbers(int):max => ~max
LOI_4:31:int_Numbers(int):i => ~i
LOI_5:31:int_Numbers(int):n => ~n
LOI_6:32:int_Numbers(int):max => ~max
LOI_7:32:int_Numbers(int):i => ~i
LOI_8:33:int_Numbers(int):i => ~i
LOI_9:36:int_Numbers(int):n => ~n
LOI_10:36:int_Numbers(int):max => ~max
LOI_11:37:int_Numbers(int):max => ~max
ASRS_1:31:int_Numbers(int):i += 1 => i /= 1
ASRS_2:31:int_Numbers(int):i += 1 => i *= 1
ASRS_3:31:int_Numbers(int):i += 1 => i -= 1
ASRS_4:31:int_Numbers(int):i += 1 => i %= 1

```

شکل ۷. بخشی از عملگرهای جهشی ایجاد شده برای برنامه محک

FindMax

```

* Compiling traditional mutants into bytecode
+AOIS_1 +AOIS_10 +AOIS_11 +AOIS_12 +AOIS_13 +AOIS_14
+AOIS_15 +AOIS_16 +AOIS_17 +AOIS_18 +AOIS_19 +AOIS_2
+AOIS_20 +AOIS_21 +AOIS_22 +AOIS_23 +AOIS_24
+AOIS_25 +AOIS_26 +AOIS_27 +AOIS_28 +AOIS_29 +AOIS_3
+AOIS_30 +AOIS_31 +AOIS_32 +AOIS_33 +AOIS_34
+AOIS_35 +AOIS_36 +AOIS_37 +AOIS_38 +AOIS_39 +AOIS_4
+AOIS_40 +AOIS_41 +AOIS_42 +AOIS_43 +AOIS_44 +AOIS_5
+AOIS_6 +AOIS_7 +AOIS_8 +AOIS_9 +AORB_1 +AORB_2
+AORB_3 +AORB_4 +AORB_5 +AORB_6 +AORB_7
+AORB_8 +ASRS_1 +ASRS_2 +ASRS_3 +ASRS_4 +COR_1
+COR_2 +LOI_1 +LOI_10 +LOI_11 +LOI_2 +LOI_3 +LOI_4
+LOI_5 +LOI_6 +LOI_7 +LOI_8 +LOI_9 +ROR_1 +ROR_10
+ROR_11 +ROR_12 +ROR_13 +ROR_14 +ROR_15 +ROR_16
+ROR_17 +ROR_18 +ROR_19 +ROR_2 +ROR_20 +ROR_3
+ROR_4 +ROR_5 +ROR_6 +ROR_7 +ROR_8 +ROR_9
-----
Mutants have been created!

```

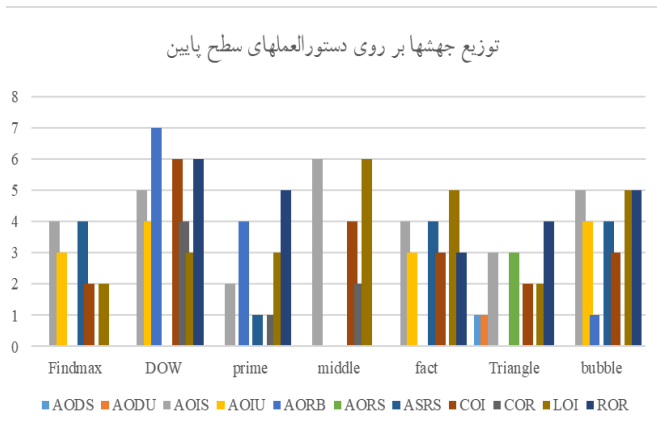
شکل ۵. نتیجه اجرای برنامه جهشی FindMax در ابزار Muclipse

در شکل ۵ نتیجه اجرای برنامه‌های آزمون بر روی برنامه‌های خطادار برای برنامه محک FindMax، نشان داده شده است. ابزار Muclipse، این قابلیت را دارد تا بتواند تعداد جهش‌های کشته شده، تعداد جهش‌های زنده، تعداد جهش‌هایی که فعال مانده‌اند و امتیاز جهش را نشان دهد. شایان ذکر است، جهش‌های زنده، ترکیبی از جهش‌های معادل و جهش‌های سرسخت است که ابزار Muclipse قابلیت تفکیک این دو جهش از هم را ندارد. گزارش نشان داده شده در شکل ۶، حاوی اطلاعاتی در زمینه اجرای موردآزمون‌ها و نمایش نتیجه مورد آزمون بر روی تک تک جهش‌های موجود در برنامه خطادار است. گزارش اجرای آزمون‌ها نشان می‌دهند که کدام جهش در برنامه، توسط مورد آزمون کشته شده است و کدامیک توسط مورد آزمون، تشخیص داده نشده است.

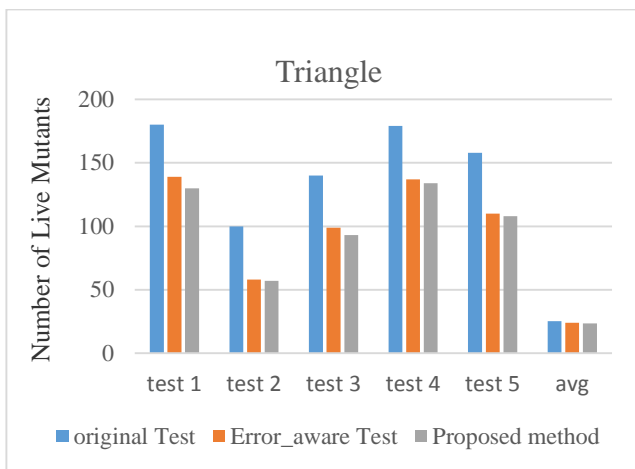
شکل ۷ نمایی کلی از محتوای فایل ایجاد شده از عملگرهای جهش تولیدشده برای برنامه FindMax را نشان می‌دهد.

۶- نتایج

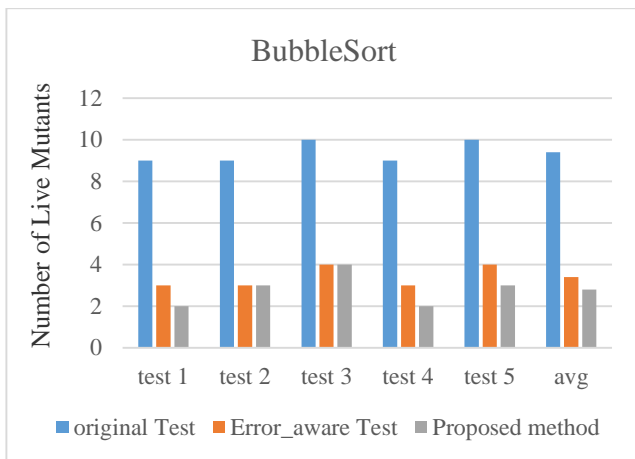
هدف از روش پیشنهادی، کاهش تعداد جهش‌های زنده در فرایند آزمون جهش است. جهش‌های زنده از دو جهش مختلف تشکیل شده است. یکی جهش‌های معادل است که این نوع جهش‌ها با هیچ مورد آزمونی قابل شناسایی نیستند. دسته دوم، جهش‌های سرسخت هستند. در روش قبلی، به فرارگیری نادرست جهش‌های سرسخت در دسته جهش‌های معادل پرداخته نشده بود. چون ممکن است به دلیل رفتار مشابه جهش‌های سرسخت با جهش‌های معادل، طبقه‌بندی نادرستی از این دو جهش رخ دهد. در این مقاله، با بررسی ویژگی‌های مهم و اثرگذار جهش‌ها، این ضعف را برطرف کرده‌ایم. نتیجه آزمایش‌ها حاکی از متمایز نمودن جهش‌های سرسخت از دسته جهش‌های معادل به دلیل کاهش تعداد جهش‌های زنده است. این امر همچنین باعث افزایش تعداد جهش‌های کشته شده نیز می‌شود.



شکل ۸. توزیع عملگرهای جهشی برای هفت برنامه محک



شکل ۹. تعداد جهش‌های زنده در برنامه Triangle برای سه روش سنتی، آگاه به انتشارخطا و روش پیشنهادی با پنج مورد آزمون



شکل ۱۰. تعداد جهش‌های زنده در برنامه BubbleSort برای سه روش سنتی، آگاه به انتشارخطا و روش پیشنهادی با پنج مورد آزمون

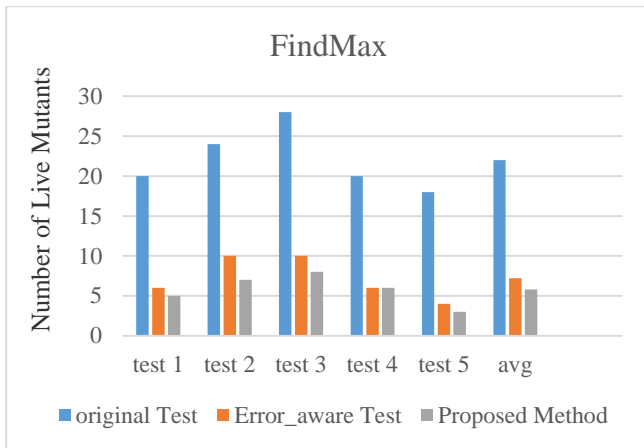
شکل ۸ تعداد کل عملگرهای تزریق شده در تمامی دستورالعمل‌های برنامه‌های محک را نشان می‌دهد. همانطور که در نمودار مشخص است، برنامه DOW، بیشترین تعداد دستورالعمل‌های سطح پایین و در نتیجه آن، بیشترین مشارکت عملگرها را دارد. جدول ۴، برای هر برنامه، تعداد کل جهش‌های تولید شده، تعداد جهش‌های تولید شده برای دستورالعمل‌های سطح بالا و تعداد جهش‌های تولید شده برای دستورالعمل‌های سطح پایین را نشان می‌دهد. تعداد کل جهش‌ها، به تفکیک نوع دستورالعمل‌ها، در سطر آخر مشخص شده است.

در روش پیشنهادی، آزمایش‌های انجام گرفته بر روی دستورالعمل‌های هفت برنامه محک با دو روش سنتی و روش آگاه به انتشار خطا مقایسه شده است. نتایج حاکی از بهبود کارایی روش پیشنهادی با کاهش تعداد جهش‌های زنده است. شکل ۹ نتایج پنج مورد آزمون مختلف بر روی برنامه Triangle را نشان می‌دهد. شکل ۱۰، تعداد جهش‌های زنده در برنامه BubbleSort برای سه روش سنتی، آگاه به انتشارخطا و روش پیشنهادی با پنج مورد آزمون را نشان می‌دهد. در شکل ۱۱، تعداد جهش‌های زنده در برنامه Factorial برای سه روش سنتی، آگاه به انتشارخطا و روش پیشنهادی با پنج مورد آزمون نمایش داده شده است. تعداد جهش‌های زنده در برنامه Prime برای سه روش سنتی، آگاه به انتشارخطا و روش پیشنهادی با پنج مورد آزمون، در شکل ۱۲ آمده است. لازم به ذکر است، برنامه‌های محک انتخاب شده برای آزمایش روش پیشنهادی، براساس آزمون‌های مختلف استاندارد جهانی انتخاب شده است.

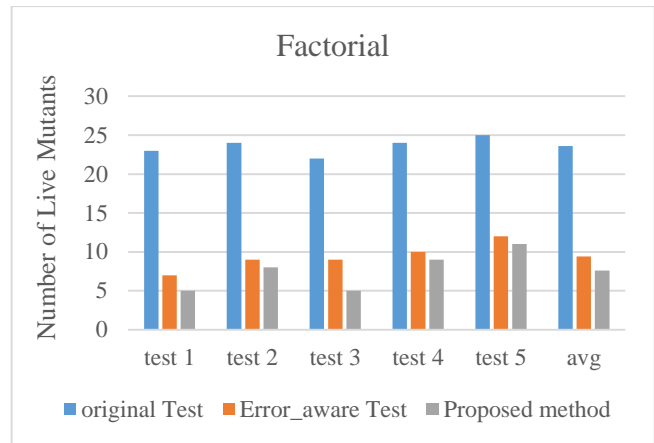
شکل ۱۳، تعداد جهش‌های زنده در برنامه Middle را نشان می‌دهد. نتایج حاکی از کاهش محسوس تعداد جهش‌های زنده است. شکل ۱۴ کاهش تعداد جهش‌های زنده برنامه FindMax را نشان می‌دهد. شکل ۱۵ نیز تعداد جهش‌های زنده در هر سه روش سنتی، آگاه به انتشار خطا و روش پیشنهادی را برای برنامه DOW نشان می‌دهد. جدول ۴. تعداد جهش‌های تولید شده براساس نوع دستورالعمل‌های

برنامه محک

نام برنامه	تعداد کل جهش‌های تولید شده برای هر برنامه	تعداد جهش‌های تولید شده برای دستورالعمل‌های سطح بالا	تعداد جهش‌های تولید شده برای دستورالعمل‌های سطح پایین
Triangle	252	208	44
Bubble Sort	97	87	10
Factorial	68	50	18
Prime	66	51	15
Middle	121	106	15
FindMax	56	38	18
Dow	294	260	34
Total Mutants	954	800	154



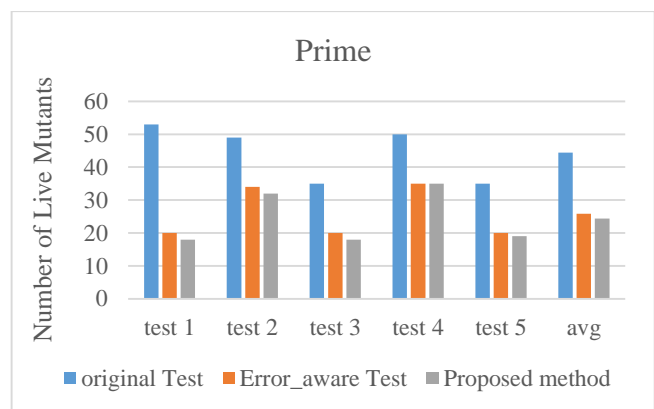
شکل ۱۴. تعداد جهش‌های زنده در برنامه FindMax برای سه روش سنتی، آگاه به انتشار خطا و روش پیشنهادی با پنج مورد آزمون



شکل ۱۱. تعداد جهش‌های زنده در برنامه Factorial برای سه روش سنتی، آگاه به انتشار خطا و روش پیشنهادی با پنج مورد آزمون



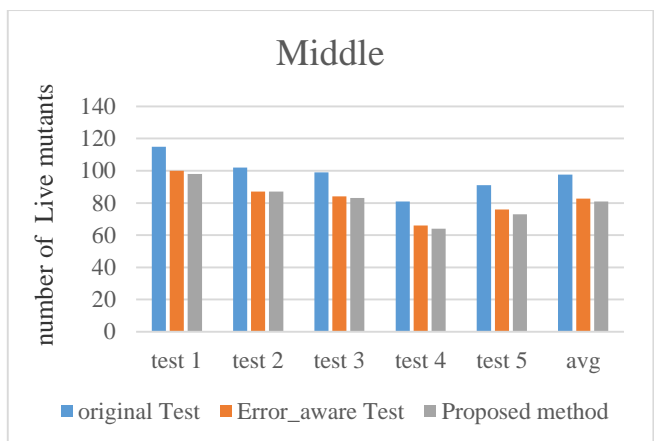
شکل ۱۵. تعداد جهش‌های زنده در برنامه DOW برای سه روش سنتی، آگاه به انتشار خطا و روش پیشنهادی با پنج مورد آزمون



شکل ۱۲. تعداد جهش‌های زنده در برنامه Prime برای سه روش سنتی، آگاه به انتشار خطا و روش پیشنهادی با پنج مورد آزمون

مراجع

- [1] Richard A. DeMillo, Richard J. Lipton, and Fred G. Sayward. "Hints on test data selection: Help for the practicing programmer." *IEEE Computer*, 11(4): 34-41, 1978.
- [2] Jeff Offutt and Ammei Lee. (1998) "An empirical evaluation of weak mutation." *IEEE Transactions on Software Engineering*, 649-660.
- [3] Mark Harman. "The current state and future of search-based software engineering." In *Proceedings of the 28th International Conference on Software Engineering (ICSE'06)*, pages 342-351. ACM, 2006.
- [4] Simon Poulding and James A. Jones. "An empirical study of the robustness of MacOS applications using random testing." In *Proceedings of the 10th International Workshop on Dynamic Analysis (WODA'12)*, pages 35-40. IEEE, 2012.
- [5] López, J., Kushik, N., & Yevtushenko, N. (2018). Source Code Optimization using Equivalent Mutants. *Inf. Softw. Technol.*, 103, 138-141. <https://doi.org/10.1016/j.infsof.2018.06.013>.
- [6] Ghiduk, A., Girgis, M., & Shehata, M. (2019). Employing Dynamic Symbolic Execution for Equivalent Mutant Detection. *IEEE Access*, 7, 163767-163777. <https://doi.org/10.1109/ACCESS.2019.2952246>.
- [7] Zeinab Asghari, Bahman Arasteh, and Abbas Koochari. 2024. Effective Software Mutation-Test Using Program Instructions



شکل ۱۳. تعداد جهش‌های زنده در برنامه Middle برای سه روش سنتی، آگاه به انتشار خطا و روش پیشنهادی با پنج مورد آزمون

- [16] Ayad, A., Marsit, I., Loh, J., Omri, M., & Mili, A. (2019). Estimating the Number of Equivalent Mutants. 2019 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW), 112-121. <https://doi.org/10.1109/ICSTW.2019.00039>.
- [17] Ayad, A., & Mili, A. (2020). Automated Estimation of the Rate of Equivalent Mutants. 2020 International Conference on Computational Science and Computational Intelligence (CSCI), 1794-1799. <https://doi.org/10.1109/CSCI51800.2020.00331>.
- [18] Kazerouni, A., Davis, J., Basak, A., Shaffer, C., Servant, F., & Edwards, S. (2021). Fast and accurate incremental feedback for students' software tests using selective mutation analysis. *J. Syst. Softw.*, 175, 110905. <https://doi.org/10.1016/j.jss.2021.110905>.
- [19] Wu, X., Zhang, J., & Zhao, Y. (2023). Model-based testing and mutation analysis: Recent advances and future directions. *Journal of Software: Evolution and Process*. <https://doi.org/10.1002/smr.2602>
- [20] Williams, J. L., Gupta, P. S., & Rodrigues, S. C. (2023). A comparative study of mutation testing techniques: Recent developments and future directions. *Journal of Software: Evolution and Process*. <https://doi.org/10.1002/smr.2667>
- [21] Khan, R., & Amjad, M. (2019). Mutation-based genetic algorithm for efficiency optimisation of unit testing. *Int. J. Adv. Intell. Paradigms*, 12, 254-265. <https://doi.org/10.1504/IJAIP.2019.10019862>.
- [22] Petrović, G., Ivankovic, M., Fraser, G., & Just, R. (2021). Practical Mutation Testing at Scale: A view from Google. *IEEE Transactions on Software Engineering*, 48, 3900-3912. <https://doi.org/10.1109/TSE.2021.3107634>.
- [23] Lee, J., Patel, A., Zhang, K., & Chen, Y. (2023). Advancements in mutation testing using deep learning techniques: A review and future perspectives. *IEEE Access*, 11, 158264158283. <https://doi.org/10.1109/ACCESS.2023.3290548>.
- Classification. *J. Electron. Test.* 39, 5–6 (Dec 2023), 631–657. <https://doi.org/10.1007/s10836-023-06089-0>
- [8] Ghiduk, A., Girgis, M., & Shehata, M. (2019). Employing Dynamic Symbolic Execution for Equivalent Mutant Detection. *IEEE Access*, 7, 163767-163777. <https://doi.org/10.1109/ACCESS.2019.2952246>.
- [9] Jammalamadaka, K., & Parveen, N. (2021). Equivalent mutant identification using hybrid wavelet convolutional rain optimization. *Software: Practice and Experience*, 52, 576 - 593. <https://doi.org/10.1002/spe.3038>.
- [10] Souza, B., & Gheyi, R. (2020). A Lightweight Technique to Identify Equivalent Mutants. https://doi.org/10.5753/CBSOFT_ESTENDIDO.2020.14630.
- [11] Tenorio, M., Lopes, R., Fachine, J., Marinho, T., & Costa, E. (2019). Subsumption in Mutation Testing: An Automated Model Based on Genetic Algorithm. 16th International Conference on Information Technology-New Generations (ITNG 2019). https://doi.org/10.1007/978-3-030-14070-0_24.
- [12] Fernandes, L., Ribeiro, M., Gheyi, R., Delamaro, M., Guimarães, M., & Santos, A. (2022). Put Your Hands In The Air! Reducing Manual Effort in Mutation Testing. *Proceedings of the XXXVI Brazilian Symposium on Software Engineering*. <https://doi.org/10.1145/3555228.3555233>.
- [13] Abuljadayel, A., & Wedyan, F. (2018). An Approach for the Generation of Higher Order Mutants Using Genetic Algorithms. *International Journal of Intelligent Systems and Applications*, 10, 34-45. <https://doi.org/10.5815/IJISA.2018.01.05>.
- [14] Basile, D., Beek, M., Cordy, M., & Legay, A. (2020). Tackling the equivalent mutant problem in real-time systems: the 12 commandments of model-based mutation testing. *Proceedings of the 24th ACM Conference on Systems and Software Product Line: Volume A - Volume A*. <https://doi.org/10.1145/3382025.3414966>.
- [15] Naem, M., Lin, T., Naem, H., & Liu, H. (2020). A machine learning approach for classification of equivalent mutants. *Journal of Software: Evolution and Process*, 32. <https://doi.org/10.1002/smr.2238>.

Generating Personalized Explanations for Twitter List Recommendations Using Semantic Similarity of Hashtags

Havva Alizadeh Noughabi¹, Behshid Behkamal^{1*}, Saleh Naseri¹, Mohsen Kahani¹

¹ Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Received: 15 January 2024, Revised: 11 April 2024, Accepted: 05 May 2024

Paper type: Research

Abstract

Twitter List recommender systems have achieved high prediction accuracy by leveraging diverse user and List information alongside complex algorithms. However, explainability remains a significant challenge in these systems. Providing meaningful explanations along with a set of recommendations can enhance user trust and satisfaction, assisting them in informed decision-making. In this paper, we present a model for the automated generation of personalized descriptions as explanations for recommended Twitter Lists. Specifically, our model selects frequently used hashtags from the content of the recommended List, establishing semantic relationships with the user's activity history. The aim is to present the List's subject in an understandable and personalized manner through a generated description. Through experiments conducted on a real Twitter dataset, our proposed model demonstrates its capability to generate explanations for a high percentage of the recommendations provided by a recommendation model.

Keywords: Explainable Recommender System, Personalized Explanation, Twitter List, Semantic Similarity of Hashtags.

* Corresponding Author's email: behkamal@um.ac.ir

تولید توضیح شخصی سازی شده برای پیشنهادهای لیست توئیتر مبتنی بر شباهت معنایی هشتگ‌ها

حوا علیزاده نوقابی^۱، بهشید بهکمال^{۱*}، صالحه ناصری^۱، محسن کاهانی^۱
^۱ گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

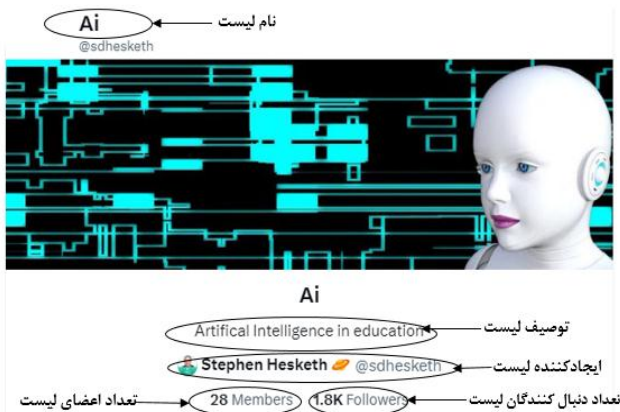
تاریخ دریافت: ۱۴۰۲/۱۰/۲۵ تاریخ بازبینی: ۱۴۰۳/۰۱/۲۳ تاریخ پذیرش: ۱۴۰۳/۰۲/۱۶
نوع مقاله: پژوهشی

چکیده

امروزه سیستم‌های پیشنهاددهنده‌ی لیست‌های توئیتر با بکارگیری اطلاعات مختلف کاربران و لیست‌ها و همچنین اعمال الگوریتم‌های پیچیده، توانسته‌اند به دقت بالایی در پیش‌بینی برسند و پیشنهادهای مرتبط برای هر کاربر تولید کنند، اما قابلیت توضیح‌پذیری در این سیستم‌ها به عنوان یک چالش مطرح می‌باشد. یک توضیح مناسب به همراه لیست توئیتر پیشنهادشده، علاوه بر افزایش اعتماد و رضایت کاربران، می‌تواند به آنها در تصمیم‌گیری آگاهانه کمک نماید. از این رو در این مقاله، یک مدل تولید توضیح، ارائه می‌شود که به صورت خودکار، یک توصیف شخصی سازی شده برای لیست توئیتر پیشنهادشده، ایجاد می‌نماید. بطور دقیق‌تر، این مدل با انتخاب مجموعه‌ای از هشتگ‌های پرتکرار لیست توئیتر که شباهت معنایی با تاریخچه فعالیت‌های قبلی کاربر دارد، سعی می‌کند موضوع لیست را به گونه‌ای قابل فهم در قالب یک توضیح، همراه با لیست پیشنهادشده، به کاربر نمایش دهد. پس از جمع‌آوری یک مجموعه داده واقعی از شبکه اجتماعی توئیتر، با انجام آزمایش‌ها نشان داده شد که مدل پیشنهادی قادر به تولید توضیح برای درصد بالایی از پیشنهادهای ایجادشده براساس یک مدل پیشنهاددهنده‌ی پایه می‌باشد.

کلیدواژگان: سیستم پیشنهاددهنده توضیح‌پذیر، توضیح شخصی سازی شده، لیست توئیتر، شباهت معنایی هشتگ‌ها.

* رایانامه نویسنده مسئول: behkamal@um.ac.ir



شکل ۱. یک نمونه لیست توئیتر

توضیحات در یک سیستم پیشنهاددهنده ممکن است برای رسیدن به اهداف مختلفی مانند شفافیت^۷، متقاعدسازی^۸، اثربخشی^۹، رضایت‌مندی^{۱۰} و غیره فراهم شده باشند [۸، ۹]. به عنوان مثال یک توضیح ممکن است دلیل پیشنهاد یک قلم خاص به کاربر را بیان کند: «این قلم [لنز] به شما پیشنهاد شده است، چون شما قبلاً قلم مرتبطی [دوربین] خریداری کرده‌اید» [۶]. بطور مشابه هدف یک توضیح می‌تواند فراهم نمودن اطلاعات در مورد قلم پیشنهادشده برای کمک به اتخاذ تصمیم آگاهانه توسط کاربر باشد. مانند بیان یک توضیح حاوی اطلاعات راجع به محتوا و موضوع یک لیست توئیتر که می‌تواند به کاربر در شناخت آن لیست و در نتیجه تصمیم‌گیری برای پذیرش یا رد آن پیشنهاد کمک بسزایی نماید. با توجه به اینکه در اکثر مواقع حدس زدن موضوع یا موضوعات اصلی یک لیست توئیتر فقط با دیدن نام آن برای کاربران سخت است، معمولاً کاربران به صورت دستی تعدادی از توئیتهای لیست را بررسی می‌کنند تا موضوع آن را درک کنند که خسته‌کننده و زمان‌بر است. بنابراین تولید خودکار یک توصیف برای یک لیست توئیتر می‌تواند یک نوع توضیح آگاهی‌دهنده باشد.

از آنجا که هشتک‌ها در شبکه‌های اجتماعی به عنوان کلمات کلیدی بسیار استفاده می‌شوند و به موضوع یا مفهوم خاصی در یک توئیت اشاره می‌کنند، در این تحقیق توصیف یک لیست توئیتر براساس هشتک‌های پرتکرار محتوای آن در نظر گرفته می‌شود. اگرچه که استخراج هشتک‌های پرتکرار لیست و ارائه آن به کاربر می‌تواند توصیفی مناسب راجع به لیست فراهم نماید، اما برآیند تا این توصیف، برای کاربر مدنظر، شخصی‌سازی شود.

۱- مقدمه

افزایش قابل توجه محتوای تولیدشده در شبکه‌های اجتماعی، باعث شده است که جهت سازماندهی اطلاعات مرتبط در یک زمینه، راه‌حلهایی توسط پلتفرم‌های مختلف به کاربران ارائه داده شود. به عنوان مثال، لیست‌ها در توئیتر^۱ برای مقابله با مشکل سربار اطلاعات در سال ۲۰۱۳ معرفی شدند [۱]. یک لیست توئیتر^۲، گروهی از حساب‌های کاربری است که به موضوع آن لیست علاقه‌مند هستند. همانطور که در شکل ۱ نشان داده می‌شود هر لیست تعدادی عضو دارد که امکان درج توئیت در مورد موضوع آن لیست را دارند و سایر کاربران می‌توانند لیست را دنبال کنند و از مطالب آن در زمینه‌ی مشخص شده استفاده کنند. در همین راستا سیستم‌های پیشنهاددهنده لیست توئیتر [۲، ۳] طراحی شدند که با استفاده از ویژگی‌های مختلف کاربر و همچنین ویژگی‌های لیست، سعی در بهبود عملکرد در جهت ارائه پیشنهادات مرتبط برای هر کاربر دارند و اما چالش اصلی آنها عدم فراهم نمودن توضیح^۳ مناسب برای هر پیشنهاد می‌باشد. عبارتی قابلیت توضیح‌پذیری^۴ در این سیستم‌ها اهمیت بسیاری دارد که به عنوان یک چالش مهم مطرح است.

در سال‌های اخیر فاکتورهایی مانند اعتماد و رضایت کاربران در شبکه‌های اجتماعی بسیار مورد توجه قرار گرفته است و تحت‌تأثیر عواملی مانند قابلیت توضیح‌پذیری در سیستم‌های پیشنهاددهنده اجتماعی می‌باشد [۴]. در واقع ارائه یک توضیح مناسب همراه هر قلم پیشنهادشده (مانند یک لیست توئیتر)، علاوه بر جلب رضایت کاربران، باعث تعامل بیشتر آنها با پیشنهاد ارائه شده می‌شود. از این رو هدف اصلی این تحقیق فراهم نمودن توضیحی قابل‌فهم و شخصی‌سازی‌شده^۵ می‌باشد. از آنجا که توسعه روش‌های تولید توضیح تعقیبی^۵ [۵، ۶، ۷]، که در آنها توضیح‌ها پس از ایجاد پیشنهادها و بصورت مستقل از مدل پیشنهاددهنده، تولید می‌شوند، مورد توجه زیادی قرار گرفته است، این تحقیق مدل تولید توضیحی را ارائه می‌دهد که دارای انعطاف‌پذیری برای بکارگیری طیف وسیعی از مدل‌های پیشنهاددهنده می‌باشد. عبارت دیگر هر نوع سیستم پیشنهاددهنده لیست توئیتر می‌تواند از این مدل برای تولید توضیح استفاده نماید.

^۱ اخیراً این شبکه اجتماعی، نام جدید X گرفته است.

^۲ Twitter List

^۳ explanation

^۴ explainability

^۵ personalized

^۶ post-hoc explanation

^۷ transparency

^۸ persuasiveness

^۹ effectiveness

^{۱۰} satisfaction

۲- کارهای گذشته

امروزه توضیح پذیر بودن مدل‌های پیشنهاددهنده یک مساله حائز اهمیت به شمار می‌رود. بعبارت دیگر یک سیستم پیشنهاددهنده علاوه بر آن که عملکرد خوبی در تولید پیشنهادهای مرتبط به کاربران دارد، بایستی توضیح‌های مفید به همراه اقلام پیشنهاددهنده به کاربران ارائه نماید. مدل‌های تولید توضیح را می‌توان به دو دسته کلی طبقه‌بندی کرد: مستقل از مدل پیشنهاددهنده و یا وابسته به مدل پیشنهاددهنده [۶, ۸, ۱۳]. رویکردهای دسته اول با استفاده از مدل‌های جداگانه‌ای که مستقل از الگوریتم پیشنهاددهنده هستند، توضیحاتی را ایجاد می‌کنند که به آنها توضیح تعقیبی می‌گویند. اینگونه رویکردها اغلب انعطاف پذیر هستند و می‌توانند برای طیف وسیعی از الگوریتم‌های پیشنهاددهنده اعمال شوند. تکنیک‌های متداول برای ایجاد توضیحات تعقیبی شامل مدل‌های تقلیدکننده [۴, ۱۴]، مدل‌های مبتنی بر روش‌های داده‌کاوی مثل قانون‌کاوی انجمنی [۵, ۱۵] و یا مبتنی بر زیرگراف [۱۶] است. از طرف دیگر توضیحات ذاتی^۶ توسط مدل پیشنهاددهنده ایجاد می‌شوند و اغلب هدف آنها ارائه توضیحی در مورد اینکه چرا الگوریتم پیشنهاددهنده یک قلم خاص را به کاربر پیشنهاد داده، می‌باشد. از جمله تحقیقات انجام شده در این زمینه می‌توان روش‌های مبتنی بر مدل‌های فاکتورسازی^۷ [۱۷, ۱۸]، مدل‌های مبتنی بر گراف دانش [۱۹] تا [۲۲]، مدل‌های یادگیری عمیق [۲۳] تا [۲۵] و مدل‌های مبتنی بر قاعده^۸ [۲۶, ۲۷] را نام برد.

همچنین توضیح پذیر بودن سیستم‌های پیشنهاددهنده اجتماعی^۹ در ایجاد اعتماد کاربران بسیار مورد توجه قرار گرفته است، زیرا برای حفظ پایداری شبکه‌های اجتماعی، ضروری دانسته شده است [۲۸]. در برخی از تحقیقات قبلی مرتبط با تولید توضیح برای پیشنهادهای اجتماعی، توضیحات را بر اساس شبکه کاربران ارائه می‌دهند. به عنوان مثال، وانگ و همکاران [۲۹] توضیحی را فراهم می‌کنند که از قالب «کاربر آ و کاربر ب نیز این قلم را می‌پسندند.» پیروی می‌کند. برای ایجاد این توضیح، آنها الگوریتمی را برای شناسایی مجموعه بهینه کاربران برای گنجاندن در توضیح طراحی کردند که کاربر مدنظر را در مورد قلم پیشنهاددهنده متقاعد کنند. در برخی دیگر از مدل‌های توضیح‌پذیر، یک شبکه اطلاعاتی نامگون^{۱۰} از اطلاعات کاربران (مانند تعاملات قبلی آنها) و اقلام (مانند ویژگی‌های

به صورت مشخص تر می‌توان گفت از آنجا که هر کاربر در توئیت دارای یک تاریخچه‌ی پست می‌باشد، اگر توضیح ارائه شده، بجای بکارگیری تمام هشتگ‌های پرتکرار لیست پیشنهاددهنده، تنها موارد مرتبط با تاریخچه‌ی پست‌های کاربر را دربرگیرد، دو مزیت به همراه خواهد داشت: (۱) توضیح برای کاربر، قابل فهم^۱ خواهد شد که فراهم نمودن توضیح قابل فهم در بافتار^۲ سیستم‌های پیشنهاددهنده‌ی توضیح‌پذیر اهمیت ویژه‌ای دارد [۱۰]، به عبارتی هر توضیح براساس میزان دانش هر کاربر می‌تواند شخصی سازی شود تا کاربر قادر باشد شناخت بهتری از قلم پیشنهاددهنده با کمک توضیح فراهم شده به دست آورد [۱۱]. (۲) چنین توضیحی می‌تواند به متقاعدسازی کاربر نسبت به چرایی پیشنهاد این لیست توسط سیستم پیشنهاددهنده، کمک کند. از آنجا که این توضیح به ارتباط لیست پیشنهاددهنده با فعالیت‌های قبلی کاربر (در اینجا تعاملات قبلی از طریق پست‌های توئیت) می‌پردازد، کاربر دلیل پیشنهاد چنین قلمی از سمت سیستم پیشنهاددهنده را درک می‌کند و به مدل پیشنهاددهنده اعتماد می‌کند که یکی از اهداف مهم مدل‌های تولید توضیح است [۶].

در راستای شخصی سازی توضیح، برای محاسبه شباهت بین هشتگ‌های لیست توئیت پیشنهاددهنده و هشتگ‌های استخراج شده از تاریخچه پست‌های کاربر از روش مبتنی بر BERT^۳ [۱۲] استفاده می‌شود که قادر است بصورت معنایی، ارتباط هشتگ‌ها را مشخص نماید. بنابراین اهداف اصلی این مقاله در موارد زیر خلاصه می‌شود:

- ارائه یک روش تولید توضیح توصیفی برای لیست‌های توئیت پیشنهاددهنده، به گونه‌ای که به کاربران برای تصمیم‌گیری آگاهانه کمک نماید.
- شخصی سازی توضیح توصیفی برای افزایش قابلیت فهم آن که براساس شباهت معنایی بین هشتگ‌های لیست توئیت پیشنهاددهنده و هشتگ‌های موجود در تاریخچه پست‌های کاربر انجام می‌شود.
- یک مجموعه داده از توئیت جمع‌آوری می‌گردد و جهت ارزیابی روش پیشنهادی، آزمایش‌های مختلف انجام می‌شود.

در ادامه مقاله، در بخش ۲ به بررسی کارهای گذشته‌ی مرتبط با موضوع تحقیق پرداخته می‌شود و سپس روش پیشنهادی به همراه جزئیات آن در بخش ۳ شرح داده می‌شود. نتایج آزمایش‌ها و تحلیل آنها در بخش ۴ آورده شده و در بخش ۵، جمع‌بندی آمده است.

⁶ intrinsic

⁷ factorization model

⁸ rule

⁹ social recommender systems

¹⁰ heterogeneous information network (HIN)

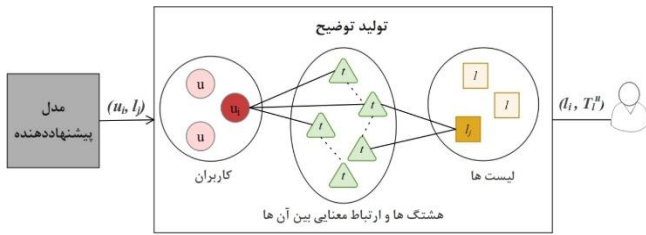
¹ understandable

² context

³ Bidirectional Encoder Representations from Transformers

⁴ surrogate models

⁵ association rule mining



شکل ۲. توضیح مبتنی بر هشتگ برای لیست توئیتر پیشنهاددهنده

از آنجایی که در این روش، توضیحات از نوع تعقیبی هستند، مدل پیشنهاددهنده به شکل جعبه سیاه رسم شده است، زیرا برای هر لیست پیشنهاددهنده بدون در نظر گرفتن الگوریتم پیشنهاددهنده، یک توضیح فراهم خواهد شد.

بنابراین ورودی مولفه^۷ تولید توضیح شامل جفت (کاربر u ، لیست پیشنهاددهنده l) می‌باشد. آنگاه براساس این ورودی و همچنین هشتگ‌های پرتکرار کاربر و لیست و ارتباط معنایی بین آنها، یک توضیح تولید می‌شود که نهایتاً همراه با لیست پیشنهاددهنده، یعنی جفت (l, T_l^u) ، به کاربر ارائه داده می‌شود و در نتیجه قابلیت توضیح‌پذیری به مدل پیشنهاددهنده افزوده می‌گردد.

همانطور که در کامپوننت تولید توضیح در شکل ۲ مشاهده می‌شود نیاز به دانستن ارتباط معنایی بین هشتگ‌ها (خط‌چین) می‌باشد. از آنجایی که مدل BERT، یک مدل مطرح در زمینه پردازش زبان طبیعی است و به عنوان یکی از پیشروهای این حوزه به حساب می‌آید، برای استخراج روابط معنایی بین هشتگ‌ها از آن استفاده می‌شود. در واقع این مدل می‌تواند مفاهیم پیچیده و ارتباطات معنایی کلمات را با دقت خوبی مشخص کند و بنابراین در این مقاله به عنوان یک ابزار قدرتمند مورد استفاده قرار می‌گیرد تا شباهت معنایی بین دو هشتگ را تعیین کند.

الگوریتم شکل ۳، نحوه انتخاب زیرمجموعه‌ی مناسبی از هشتگ‌های پرتکرار لیست با در نظر داشتن هشتگ‌های شناخته‌شده کاربر را بیان می‌کند. بطور دقیق‌تر می‌توان گفت اگر شباهت معنایی بین یک هشتگ پرتکرار از لیست و یک هشتگ پرتکرار کاربر از یک مقدار آستانه بیشتر باشد به عنوان عضوی از مجموعه هشتگ‌های خروجی الگوریتم خواهد بود. مقدار این آستانه می‌تواند به هر عددی بین صفر و ۱ تنظیم شود که هر چه این عدد به ۱ نزدیکتر باشد، یعنی شباهت هشتگ‌های خروجی الگوریتم از نظر معنایی بسیار شبیه به هشتگ‌های قبلی کاربر خواهد بود.

مختلف) ایجاد می‌شود و سپس توضیحات به شکل مسیر^۱ اتصال کاربر به قلم پیشنهاددهنده در گراف، ارائه می‌شود [۳۰، ۳۱]. اخیراً، محققان محتوای تولید شده توسط کاربران مانند نظرات^۲ را در تولید توضیح بکار می‌گیرند [۳۲، ۳۳]. به عنوان مثال، رن و همکاران [۳۳] مدلی را معرفی کردند که از دیدگاه‌ها^۳ به عنوان توضیح استفاده می‌کند. دیدگاه‌ها مستخرج از نظرات کاربران بوده و با سه‌گانه‌های مفهوم^۴، موضوع^۵ و برجسب احساس^۶ نشان داده می‌شوند.

در مقایسه با مدل‌های متنوع تولید توضیح برای اقلام پیشنهاددهنده، روش این مقاله بر ارائه توضیحی تعقیبی و آگاهی‌دهنده برای لیست توئیتر تاکید دارد. عبارت دیگر، مدل ارائه شده می‌تواند مستقل از مدل پیشنهاددهنده و پیچیدگی‌های ذاتی الگوریتم آن، به ایجاد خودکار توضیحات شخصی‌سازی‌شده مبتنی بر هشتگ بپردازد. چنین توضیحاتی به کاربر در شناخت لیست توئیتر پیشنهاددهنده، کمک خواهد کرد.

۳- تولید توضیح شخصی‌سازی‌شده برای لیست توئیتر پیشنهاددهنده

اگر U مجموعه کاربران و L مجموعه لیست‌های توئیتر باشد، آنگاه برای هر لیست $l \in L$ با توجه به محتوای آن می‌توان مجموعه هشتگ‌های پرتکرار را به شکل $T_l = \{t_l^1, t_l^2, \dots, t_l^m\}$ در نظر گرفت. از طرف دیگر برای هر کاربر $u \in U$ با توجه به تاریخچه‌ی پست‌های او می‌توان هشتگ‌های پرتکرار را استخراج نموده و به شکل $T_u = \{t_u^1, t_u^2, \dots, t_u^n\}$ نمایش داد. آنگاه می‌توان مساله این تحقیق که تولید توضیح شخصی‌سازی‌شده برای لیست‌های توئیتر پیشنهاددهنده است، را به شکل زیر تعریف نمود.

تعریف مساله. فرض کنید به کاربر $u \in U$ ، لیست توئیتر $l \in L$ ، توسط یک سیستم پیشنهاددهنده، پیشنهاد شده است، آنگاه یافتن زیرمجموعه‌ی مناسب از هشتگ‌های لیست (T_l) ، با در نظر داشتن هشتگ‌های پرتکرار کاربر $u (T_u)$ ، مدنظر است. این زیرمجموعه از هشتگ‌ها یعنی $T_l^u \subset T_l$ به عنوان یک توضیح توصیفی شخصی‌سازی‌شده برای لیست پیشنهاددهنده l تعریف می‌شود.

شکل ۲ بخش‌های مدل تولید توضیح برای لیست‌های توئیتر پیشنهاددهنده به کاربران را نشان می‌دهد.

⁵ topic

⁶ sentiment label

⁷ component

¹ path

² reviews

³ viewposts

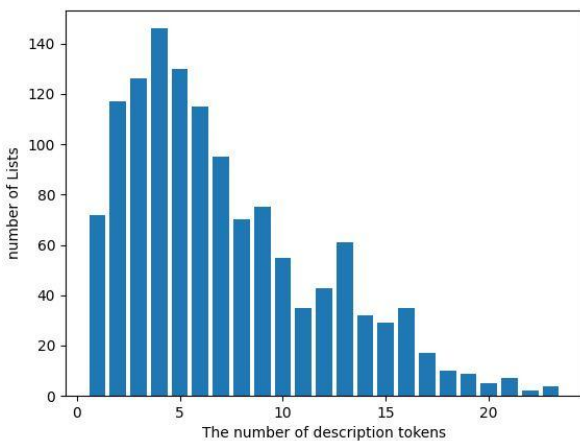
⁴ concept

جدول ۱ برخی از آمارهای اصلی در مورد مجموعه داده را نمایش می دهد.

با بررسی زیرمجموعه تصادفی شامل ۳۰۰۰ لیست، مشخص می شود که برای ۶۱٪ آنها توصیفی توسط ایجادکننده لیست، درج نشده است (مشابه توصیف ذکر شده در مثال شکل ۱) و برای ۳۹٪ درصد باقیمانده، توزیع لیستهای توئیت بر اساس تعداد توکنهای توصیف در شکل ۴ آورده شده است که بیان می کند طول توصیف لیست در بیشتر موارد کوتاه است. بنابراین، ارائه توضیحات با مقدار اطلاعات کافی در راستای کمک به کاربران برای شناخت بهتر لیست ضروری می باشد.

جدول ۱. اطلاعات آماری در مورد مجموعه داده

#کاربران	# لیست			#رابطه کاربر-لیست
	مجموع	به ازای یک کاربر		
۱۲۸۴	۱۵۳۵	حداکثر	۱۰	۶۰۵۹
		حداقل	۱	
		متوسط	۵	



شکل ۴. توزیع لیستهای توئیت بر اساس تعداد توکنهای توصیف

۴-۲- مدل پیشنهاددهنده

همانطور که در بخش ۳ گفته شد تمرکز این تحقیق بر ارائه توضیحات تعقیبی برای لیستهایی است که توسط یک مدل پیشنهاددهنده تولید شده است. از این رو برای تولید پیشنهادات برای هر کاربر، یک مدل پیشنهاددهنده لیست توئیت بر اساس [۲] بکار گرفته می شود. در این مدل، ابتدا باید نمایه ای برای کاربران و لیستها بر اساس محتوای آنها ساخته شود و سپس بر اساس شباهت بین نمایه ها، k لیست مرتبط به هر کاربر پیشنهاد داده شود. ساخت

```

Algorithm: Tag-based Explanation Generation
Input: (user  $u$ , recommended List  $l$ ), frequent tags of  $u$  ( $T_u$ ),
frequent tags of  $l$  ( $T_l$ )
Output: the subset of  $l$ 's tags ( $T_l^u \subset T_l$ )
 $T_l^u = \square$ 
for  $t_l$  in  $T_l$  do
  for  $t_u$  in  $T_u$  do
     $sim = BERT.get\_semantic\_similarity(t_l, t_u)$ 
    if  $sim > threshold$  and  $t_l$  not in  $T_l^u$  then:
       $T_l^u.append(t_l)$ 
    end if
  end for
end for
return  $T_l^u$ 

```

شکل ۳. الگوریتم تولید توضیح مبتنی بر هشتگ

در آزمایشات این مقاله با توجه به اعداد مشاهده شده مربوط به شباهت هشتگها بر اساس BERT، مقدار آستانه به ۰٫۷ تنظیم گردید. پس از تولید توضیح، می توان آن را در قالبهای مختلف به کاربر نمایش داد که در این تحقیق همانند [۳۵، ۳۴] از ابرکلمه^۱ برای نمایش هشتگها استفاده می شود.

۴- آزمایشها و ارزیابی نتایج

۴-۱- مجموعه داده

برای انجام آزمایشها، یک مجموعه داده از توئیت شامل کاربران و لیستهایی که هر کاربر دنبال می کند، با استفاده از Tweepy^۲ جمع آوری گردید. مشابه تحقیق [۳۶]، از 'Ashton Kutcher'، بازیگر و کارآفرین محبوب توئیت با بیش از ۱۷ میلیون دنبال کننده، شروع شد و سپس با توجه به لیستهای او، برای افزایش کاربران مجموعه داده، از هر لیست حداکثر ۵۰ دنبال کننده بصورت تصادفی انتخاب شدند. بعد از آن با توجه به کاربران جدید، مجموعه لیستها با جمع آوری حداکثر ۱۰ لیست به ازای هر کاربر جدید، گسترش داده شد. این چرخه ی گسترش کاربران و لیستها، ۴ بار تکرار شد که منجر به مجموعه داده ای با ۲۵،۷۸۸ کاربر و ۱۷،۶۰۴ لیست و ۶۶،۵۸۳ رابطه ی کاربر-لیست شد. سپس، توئیتهای کاربران و لیستها نیز جمع آوری گردید، به گونه ای که برای هر کاربر حداکثر ۲۰۰۰ توئیت و برای هر لیست حداکثر ۵۰۰ توئیت اخیر گردآوری شدند. مشابه [۳۷]، برای تمرکز بر کاربران فعال توئیت، فقط کاربرانی که حداقل ۲۰۰ توئیت داشته اند، به عنوان کاربران نهایی در نظر گرفته شد. پس از این پیش پردازشها، مجموعه داده شامل ۱۲۸۴ کاربر، ۱۵۳۵ لیست و ۶۰۵۹ رابطه کاربر-لیست می باشد.

^۲ <https://www.tweepy.org/>

^۱ word-cloud

۳-۴- شباهت معنایی بین هشتگ‌ها

همانطور که در بخش ۳ گفته شد برای تولید توضیح مبتنی بر هشتگ برای لیست توئیتر پیشنهاد شده، نیاز است تا ارتباط معنایی بین هشتگ‌های لیست و هشتگ‌های کاربری که این لیست به او پیشنهاد شده است، مشخص گردد. در این تحقیق برای اندازه‌گیری شباهت معنایی بین هر جفت هشتگ (هشتگ لیست، هشتگ کاربر)، BERT بکار گرفته می‌شود. در جدول ۳، چند زوج هشتگ و میزان شباهت معنایی آنها آورده شده است که بیانگر عملکرد قابل قبول BERT در این زمینه می‌باشد.

۴-۴- ارزیابی مدل تولید توضیح

در این تحقیق یک مدل جهت تولید توضیح مبتنی بر هشتگ برای لیست‌های توئیتر ارائه شد. مهمترین معیار برای ارزیابی یک مدل تولید توضیح تعقیبی، معیار وفاداری^۵ [۱۵] می‌باشد، که بیانگر درصد اقلام پیشنهاد شده‌ای است که مدل تولید توضیح می‌تواند برای آنها توضیح ایجاد کند. عبارت دیگر یک مدل تولید توضیح باید قادر باشد برای درصد قابل توجهی از پیشنهادات، توضیح فراهم نماید. معادله ۱ نحوه محاسبه معیار وفاداری را بیان می‌کند، که در آن R_u اقلام پیشنهاد شده به کاربر u با طول k است. برای هر کاربر، تعداد اقلام پیشنهاد شده‌ای که توضیح متناظر با آن، یعنی e_i ، مخالف تهی باشد، بر تعداد کل اقلام (k) تقسیم می‌شود. آنگاه میانگین مقادیر برای تمام کاربران به عنوان مقدار نهایی این معیار خواهد بود.

$$fidelity = \frac{1}{|U|} \sum_{u \in U} \frac{|\{r_i \in R_u | e_i \neq \emptyset\}|}{K} \quad (1)$$

جدول ۳. چند مثال از شباهت معنایی بین هشتگ‌ها

شباهت معنایی	جفت هشتگ
0.7101	(#ArtificialIntelligence, #Python)
0.8873	(#deepLearning, #machineLearning)
0.8190	(#AI, #ML)
0.5544	(#football, #deepLearning)
0.5946	(#politics, #algorithm)

نمایه‌ها^۱ مبتنی بر روش‌های مدل‌سازی موضوع^۲، انجام می‌شود، بدین طریق که پس از اعمال یک روش مدل‌سازی موضوع و استخراج Z موضوع نهفته در متون (در این تحقیق توئیته‌ها)، برای هر کاربر و هر لیست، موضوع‌ها و وزن هر یک مشخص می‌گردد. بر این اساس دو ماتریس ساخته می‌شود: اولی ماتریس کاربر-موضوع $A = |U| \times |Z|$ و دومی ماتریس موضوع-لیست $B = |Z| \times |L|$. برای تولید پیشنهادات، ماتریس جدیدی مبتنی بر این دو ماتریس به صورت $R = A \times B$ ایجاد می‌شود و با توجه به آن میتوان k -برترین پیشنهادات را برای هر کاربر مشخص نمود.

برای اعمال این الگوریتم پیشنهاددهنده، روش مدل‌سازی موضوع جدید به نام BERTopic بکار گرفته شد. روش BERTopic ابتدا با استفاده از مدل‌های زبانی مبتنی بر ترانسفورماتور از قبل آموزش دیده، تعبیه^۳ متون را بدست می‌آورد. در گام بعدی، پس از کاهش ابعاد تعبیه، آنها را خوشه‌بندی^۴ می‌کند و در نهایت موضوعات و کلمات متناظر آن را با روش C-TF-IDF ارائه می‌دهد [۳۸، ۳۹]. در جدول ۲ تعدادی از موضوعات و کلمه‌های متناظر آن که توسط BERTopic از مجموعه داده این تحقیق استخراج شده است، نشان داده می‌شود.

جدول ۲. برخی از موضوعات استخراج شده از مجموعه داده

موضوع	کلمات موضوع
AI	'ai', 'machinelearning', 'artificialintelligence', 'datascience', 'robot', 'bigdata', 'ml', 'learning', 'machine learning', 'artificial'
Music	'music', 'song', 'album', 'kanye', 'songs', 'wizkid', 'spotify', 'tickets', 'dj', 'dance'
Cybersecurity	'cybersecurity', 'cyber', 'security', 'infosec', 'cyber security', 'hacking', 'cyberwarrior', 'cybercrime', 'privacy', 'malware'
Cancer	'cancer', 'breast', 'breast cancer', 'heart', 'prostate', 'patients', 'prostate cancer', 'treatment', 'disease', 'surgery'
Bitcoin	'bitcoin', 'crypto', 'btc', 'blockchain', 'defi', 'ethereum', 'cryptocurrency', 'dip', 'market', 'buy'

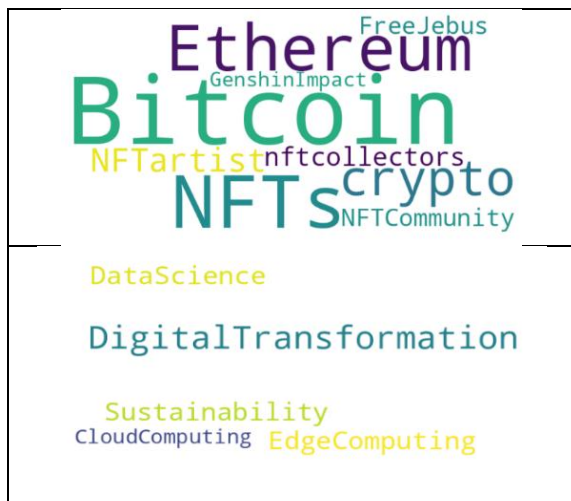
⁴ clustering

⁵ fidelity

¹ profile

² topic modeling

³ embedding



شکل ۶. ارائه توضیح به کاربر در قالب ابرکلمه

۵-۴- مطالعه موردی

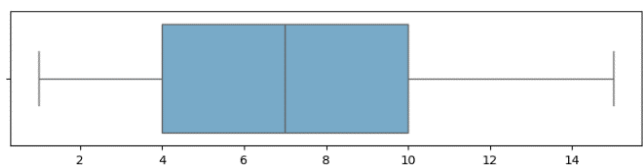
برای ارزیابی کیفی توضیحات، همانند [۱۸] با طرح پرسش‌های مناسب بررسی می‌شود که آیا مدل تولید توضیح، موفق شده است در راستای اهداف خود، توضیحات مفیدی فراهم نماید. بر همین اساس در این تحقیق دو پرسش اصلی در نظر گرفته شده است: (۱) آیا توضیح فراهم شده باعث شناخت بیشتر لیست توئیتر پیشنهادشده می‌شود؟ (۲) آیا کاربر با مشاهده توضیح، می‌تواند بینشی راجع به اینکه چرا این لیست توئیتر به او پیشنهاد شده است، بدست آورد؟ در واقع اولین پرسش به ارزیابی جنبه‌ی آگاهی‌بخش بودن توضیحات اشاره دارد و مورد دوم بررسی می‌کند که آیا توضیحات شخصی‌سازی‌شده می‌تواند باعث افزایش اعتماد کاربران به سیستم پیشنهاددهنده شود، زیرا که هر توضیح، ارتباط بین قلم پیشنهادشده و فعالیت‌های قبلی کاربر را نیز در نظر دارد.

برای انجام این مطالعه موردی، به طور تصادفی ۱۰ کاربر از مجموعه داده جمع‌آوری‌شده انتخاب گردید و با کمک مدل پیشنهاددهنده پایه برای هر کدام ۳ لیست توئیتر به عنوان اقلام پیشنهادی فراهم شد. سپس مدل تولید توضیح برای این اقلام، توضیحات مبتنی بر هشتگ را ایجاد نمود. آنگاه ۲ ارزیاب انسان مسلط به زبان انگلیسی، با در دست داشتن اطلاعات مربوط به کاربران و لیست‌های توئیتر، به ارزیابی توضیحات براساس دو پرسش فوق پرداختند. قابل بیان است که برای هر پرسش، ۵ گزینه پاسخ شامل به شدت منفی، منفی، خنثی، مثبت و به شدت مثبت وجود دارد.

براساس نتایج ارزیابی که در شکل ۷ مشاهده می‌شود، با توجه به پاسخ‌های مربوط به پرسش ۱ (بخش بالای شکل ۷) می‌توان گفت

در جدول ۴ به ازای k های مختلف که در آن k تعداد پیشنهاد ارائه شده به هر کاربر توسط مدل پیشنهاددهنده‌ی پایه (بخش ۴-۲) می‌باشد، مقدار معیار وفاداری آورده شده است. به عنوان مثال در حالت $k=1$ که به هر کاربر فقط یک قلم (در اینجا یک لیست توئیتر) پیشنهاد می‌شود، اگر مدل تولید توضیح نتواند توضیح شخصی‌سازی‌شده‌ی مبتنی بر هشتگ برای این قلم فراهم کند، معیار وفاداری برای این کاربر برابر با صفر خواهد بود. در واقع جدول ۴، میانگین مقدارهای وفاداری مربوط به تمام کاربران مجموعه داده را به عنوان مقدار نهایی وفاداری برای هر k بیان می‌کند. همانطور که نتایج نشان می‌دهد، مدل تولید توضیح ارائه‌شده می‌تواند به ازای k های مختلف، بطور میانگین برای ۹۰٫۶۲ درصد لیست‌های توئیتر پیشنهادشده، توضیح فراهم نماید. در نتیجه می‌توان گفت که مدل تولید توضیح ارائه‌شده در این مقاله، قادر است برای درصد بالایی از لیست‌های توئیتر پیشنهادشده، یک توضیح دارای هشتگ‌های پرتکرار لیست ارائه دهد به گونه‌ای که ارتباط معنایی با تاریخچه پست‌های کاربر نیز دارد. علاوه بر این، نمودار جعبه‌ای شکل ۵ براساس تعداد هشتگ‌های هر توضیح رسم شده که بیانگر آن است که میانگین تعداد هشتگ‌های هر توضیح تقریباً ۷ می‌باشد.

در شکل ۶ توضیح‌ها برای دو لیست توئیتر با نام‌های 'crypto' و 'technology' در قالب ابرکلمه آورده شده است که کاربر می‌تواند با نگاهی به این ابرکلمات در مورد محتوا و موضوعات مطرح شده در لیست توئیتر پیشنهادشده، اطلاعاتی بدست آورد که مرتبط با فعالیت‌های قبلی خود کاربر نیز می‌باشد. لازم به ذکر است که این ابرکلمه‌ها از نوع مبتنی بر فراوانی^۱ می‌باشند که در آن اندازه هر کلمه، متناسب با تعداد تکرار آن است که برای این تحقیق، برابر با فراوانی هشتگ در لیست توئیتر پیشنهادشده می‌باشد.



شکل ۵. طول هر توضیح براساس تعداد هشتگ

جدول ۴. ارزیابی براساس معیار وفاداری

	k=5	k=3	k=1	
	90.75 %	90.86 %	90.24 %	معیار وفاداری

^۱ frequency-based word-cloud

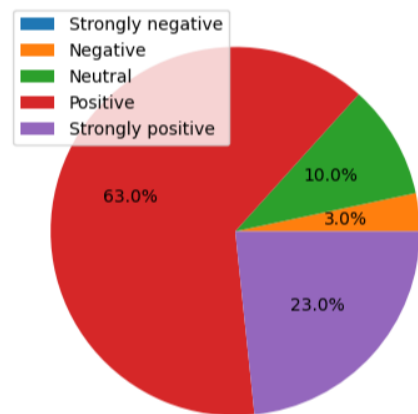
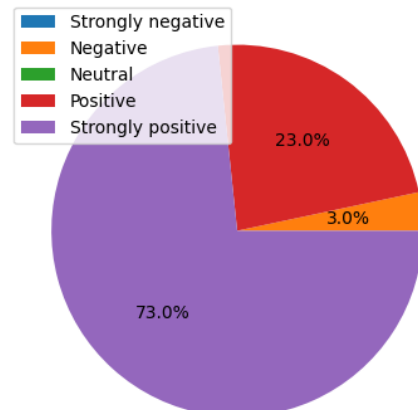
برای درصد قابل توجهی از لیست‌های توئیتر، توضیح مختص کاربر فراهم نماید. علاوه بر این طبق مطالعه موردی انجام شده، نشان داد که توضیحات فراهم‌شده علاوه بر اینکه منجر به شناخت بهتر لیست‌های توئیتر می‌گردند، باعث افزایش اعتماد کاربران از طریق ایجاد بینشی راجع به چرایی پیشنهادها می‌شوند.

در این تحقیق برای تولید پیشنهادات، از یک مدل پیشنهاددهنده پایه مبتنی بر مدل‌سازی موضوع استفاده شد. اگر چه که مدل ارائه شده در این مقاله به تولید توضیحات تعقیبی، مستقل از الگوریتم پیشنهاددهنده می‌پردازد، اما در کارهای آتی به بررسی تاثیر انواع مدل‌های پیشنهاددهنده مختلف (مانند روش‌های پالایش همکارانه، روش‌های مبتنی بر محتوا و روش‌های ترکیبی) بر روی عملکرد مدل تولید توضیح پرداخته خواهد شد.

مراجع

- [1] S. de la Rouviere and K. Ehlers, "Lists as coping strategy for information overload on Twitter," presented at the Proceedings of the 22nd International Conference on World Wide Web, 2013, pp. 199–200.
- [2] V. Rakesh, D. Singh, B. Vinzamuri, and C. K. Reddy, "Personalized recommendation of twitter lists using content and network information," presented at the Proceedings of the International AAAI Conference on Web and Social Media, 2014, pp. 416–425.
- [3] L. Chen, Y. Zhao, S. Chen, H. Fang, C. Li, and M. Wang, "iplug: Personalized list recommendation in twitter," presented at the Web Information Systems Engineering–WISE 2013: 14th International Conference, Nanjing, China, October 13-15, 2013, Proceedings, Part II 14, Springer, 2013, pp. 88–103.
- [4] C.-H. Tsai and P. Brusilovsky, "The effects of controllability and explainability in a social recommender system," *User Modeling and User-Adapted Interaction*, vol. 31, pp. 591–627, 2021.
- [5] D. Shmaryahu, G. Shani, and B. Shapira, "Post-hoc Explanations for Complex Model Recommendations using Simple Methods.," presented at the IntRS@ RecSys, 2020, pp. 26–36.
- [6] Y. Zhang and X. Chen, "Explainable recommendation: A survey and new perspectives," *Foundations and Trends® in Information Retrieval*, vol. 14, no. 1, pp. 1–101, 2020.
- [7] V. Hassija *et al.*, "Interpreting black-box models: a review on explainable artificial intelligence," *Cognitive Computation*, vol. 16, no. 1, pp. 45–74, 2024.
- [8] I. Nunes and D. Jannach, "A systematic review and taxonomy of explanations in decision support and recommender systems," *User Modeling and User-Adapted Interaction*, vol. 27, pp. 393–444, 2017.
- [9] K. Balog and F. Radlinski, "Measuring recommendation explanation quality: The conflicting goals of explanations," presented at the Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval, 2020, pp. 329–338.
- [10] R. Confalonieri, T. Weyde, T. R. Besold, and F. M. del Prado Martín, "Using ontologies to enhance human understandability of global post-hoc explanations of black-box models," *Artificial Intelligence*, vol. 296, p. 103471, 2021.
- [11] J. Schneider and J. Handali, "Personalized explanation in machine learning: A conceptualization," *arXiv preprint arXiv:1901.00770*, 2019.

توضیح تولیدشده توسط مدل پیشنهادی در اکثر مواقع توانسته است لیست توئیتر را به خوبی توصیف کند تا کاربران شناخت مناسبی از قلم پیشنهادشده پیدا کنند. همچنین با توجه به پاسخ‌های مربوط به پرسش ۲ (بخش پایین شکل ۷) در درصد قابل توجهی از مواقع، ارتباط بین توضیح فراهم‌شده با فعالیت‌های قبلی کاربران، باعث ایجاد بینشی راجع به چرایی پیشنهاد این لیست توئیتر شده است.



شکل ۷. نتایج مطالعه موردی (پاسخ‌های مربوط به پرسش اول و دوم به ترتیب در بخش بالا و پایین شکل آمده است).

۵- جمع‌بندی

با توجه به اهمیت قابلیت توضیح‌پذیری در سیستم‌های پیشنهاددهنده‌ی امروزی، در این تحقیق یک مدل تولید توضیح برای لیست‌های توئیتر پیشنهادشده به کاربران شبکه اجتماعی ارائه شد. با توجه به اینکه هدف از فراهم نمودن توضیح در این تحقیق، ارائه اطلاعات بیشتر در مورد لیست توئیتر پیشنهادشده جهت تصمیم‌گیری آگاهانه کاربران است، یک توصیف شخصی‌سازی شده مبتنی بر هشتگ در قالب ابرکلمه ایجاد می‌شود.

برای انجام آزمایش‌ها و بررسی قابلیت عملیاتی بودن مدل تولید توضیح ارائه‌شده، یک مجموعه داده از توئیتر جمع‌آوری گردید و سپس نتایج آزمایش‌ها نشان دادند که مدل تولید توضیح قادر است

- [25] Z. Chen *et al.*, “Co-attentive multi-task learning for explainable recommendation,” presented at the IJCAI, 2019, pp. 2137–2143.
- [26] X. Wang, X. He, F. Feng, L. Nie, and T.-S. Chua, “Tem: Tree-enhanced embedding model for explainable recommendation,” presented at the Proceedings of the 2018 world wide web conference, 2018, pp. 1543–1552.
- [27] W. Ma *et al.*, “Jointly learning explainable rules for recommendation with knowledge graph,” presented at the The world wide web conference, 2019, pp. 1210–1221.
- [28] W. Sherchan, S. Nepal, and C. Paris, “A survey of trust in social networks,” *ACM Computing Surveys (CSUR)*, vol. 45, no. 4, pp. 1–33, 2013.
- [29] B. Wang, M. Ester, J. Bu, and D. Cai, “Who also likes it? generating the most persuasive social explanations in recommender systems,” presented at the Proceedings of the AAAI Conference on Artificial Intelligence, 2014.
- [30] C. Shi, Z. Zhang, P. Luo, P. S. Yu, Y. Yue, and B. Wu, “Semantic path based personalized recommendation on weighted heterogeneous information networks,” presented at the Proceedings of the 24th ACM International Conference on Information and Knowledge Management, 2015, pp. 453–462.
- [31] Y. Zhang, X. Xu, H. Zhou, and Y. Zhang, “Distilling structured knowledge into embeddings for explainable and accurate recommendation,” presented at the Proceedings of the 13th international conference on web search and data mining, 2020, pp. 735–743.
- [32] J. Zheng, Z. Qin, S. Wang, and D. Li, “Attention-based explainable friend link prediction with heterogeneous context information,” *Information Sciences*, vol. 597, pp. 211–229, 2022.
- [33] Z. Ren, S. Liang, P. Li, S. Wang, and M. de Rijke, “Social collaborative viewpoint regression with explainable recommendations,” presented at the Proceedings of the tenth ACM international conference on web search and data mining, 2017, pp. 485–494.
- [34] Y. Wu and M. Ester, “Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering,” presented at the Proceedings of the eighth ACM international conference on web search and data mining, 2015, pp. 199–208.
- [35] Y. Zhang, “Incorporating phrase-level sentiment analysis on textual reviews for personalized recommendation,” presented at the Proceedings of the eighth ACM international conference on web search and data mining, 2015, pp. 435–440.
- [36] D. Kim, Y. Jo, I.-C. Moon, and A. Oh, “Analysis of twitter lists as a potential source for discovering latent characteristics of users,” presented at the ACM CHI workshop on microblogging, Citeseer, 2010.
- [37] C. Lu, W. Lam, and Y. Zhang, “Twitter user modeling and tweets recommendation based on wikipedia concept graph,” presented at the Workshops at the Twenty-Sixth AAAI conference on artificial intelligence, 2012.
- [38] R. Egger and J. Yu, “A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts,” *Frontiers in sociology*, vol. 7, p. 886498, 2022.
- [39] M. Grootendorst, “BERTopic: Neural topic modeling with a class-based TF-IDF procedure,” *arXiv preprint arXiv:2203.05794*, 2022.
- [12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
- [13] M. Caro-Martínez, G. Jiménez-Díaz, and J. A. Recio-García, “Conceptual modeling of explainable recommender systems: an ontological formalization to guide their design and development,” *Journal of Artificial Intelligence Research*, vol. 71, pp. 557–589, 2021.
- [14] C. Nóbrega and L. Marinho, “Towards explaining recommendations through local surrogate models,” presented at the Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, 2019, pp. 1671–1678.
- [15] G. Peake and J. Wang, “Explanation mining: Post hoc interpretability of latent factor models for recommendation systems,” presented at the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2060–2069.
- [16] C. Lonjarret, C. Robardet, M. Planchevit, R. Auburtin, and M. Atzmueller, “Why should i trust this item? explaining the recommendations of any model,” presented at the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), IEEE, 2020, pp. 526–535.
- [17] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, “Explicit factor models for explainable recommendation based on phrase-level sentiment analysis,” presented at the Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, 2014, pp. 83–92.
- [18] N. Wang, H. Wang, Y. Jia, and Y. Yin, “Explainable recommendation via multi-task learning in opinionated text data,” presented at the The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 165–174.
- [19] H. Wang *et al.*, “Ripplenet: Propagating user preferences on the knowledge graph for recommender systems,” presented at the Proceedings of the 27th ACM international conference on information and knowledge management, 2018, pp. 417–426.
- [20] Q. Ai, V. Azizi, X. Chen, and Y. Zhang, “Learning heterogeneous knowledge base embeddings for explainable recommendation,” *Algorithms*, vol. 11, no. 9, p. 137, 2018.
- [21] Y. Xian, Z. Fu, S. Muthukrishnan, G. De Melo, and Y. Zhang, “Reinforcement knowledge graph reasoning for explainable recommendation,” presented at the Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, 2019, pp. 285–294.
- [22] X. Wang, Q. Li, D. Yu, Q. Li, and G. Xu, “Reinforced path reasoning for counterfactual explainable recommendation,” *IEEE Transactions on Knowledge and Data Engineering*, 2024.
- [23] F. Fusco, M. Vlachos, V. Vasileiadis, K. Wardatzky, and J. Schneider, “RecoNet: An Interpretable Neural Architecture for Recommender Systems,” presented at the IJCAI, 2019, pp. 2343–2349.
- [24] Y. Lu, R. Dong, and B. Smyth, “Why I like it: multi-task learning for recommendation and explanation,” presented at the Proceedings of the 12th ACM Conference on Recommender Systems, 2018, pp. 4–12.

Providing a Blockchain-Based Method to Protect Users' Privacy in Social Networks

Ibrahim Zamani Babgohari¹, Monireh Hosseini^{2*}

¹ Department of Information Technology Engineering, KNTU, Tehran, Iran

² Department of Information Technology Engineering, KNTU, Tehran, Iran

Received: 27 August 2023, Revised: 11 May 2024, Accepted: 13 May 2024
Paper type: Research

Abstract

Today, social networks have brought a lot of convenience to their users, but there are still problems of not respecting privacy. This has caused the security and privacy protection of social network users to be important. Users share a large amount of personal data on social networks, attackers can obtain sensitive personal information simply by using social networks. and carry out various types of attacks and identity theft. Many social networks do not have their own data centers and usually store user data in third-party data centers. These centers can share this data with others and provide user data to other organizations without their knowledge. Owners of social networks can provide users' data to information agencies or use them for advertising purposes. In this research, in the first part, we have described the introduction of the subject, and then in the second part, we have discussed the literature on the subject, which is mostly focused on the centralized methods of privacy protection, and then in the third part, we have described the research method. And we have presented a method that increases the privacy of users in social networks to a great extent by using blockchain. In the fourth part, we have described the implementation of the research and finally we have concluded and summarized this research.

Keywords: Privacy, Information Security, Blockchain.

* Corresponding Author's email: hosseini@kntu.ac.ir

ارائه روشی مبتنی بر بلاکچین برای حفظ حریم خصوصی کاربران در شبکه‌های اجتماعی

ابراهیم زمانی بابگهری^۱، منیره حسینی^{۲*}

^۱ گروه مهندسی فناوری اطلاعات، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

^۲ گروه مهندسی فناوری اطلاعات، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۶/۰۵ تاریخ بازبینی: ۱۴۰۳/۰۲/۲۲ تاریخ پذیرش: ۱۴۰۳/۰۲/۲۴

نوع مقاله: پژوهشی

چکیده

امروزه شبکه‌های اجتماعی راحتی زیادی را برای کاربران خود به ارمغان آورده‌اند، اما هنوز مشکلات عدم رعایت حریم خصوصی وجود دارد. همین امر باعث شده امنیت و حفاظت از حریم خصوصی کاربران شبکه‌های اجتماعی مورد اهمیت قرار بگیرد. کاربران مقدار زیادی از داده‌های شخصی در شبکه‌های اجتماعی به اشتراک می‌گذارند، مهاجمان می‌توانند اطلاعات شخصی حساس را به سادگی با استفاده از شبکه‌های اجتماعی به دست آورند و انواع مختلفی از حملات و سرقت هویت را انجام دهند. بسیاری از شبکه‌های اجتماعی مراکز داده خود را ندارند و معمولاً داده‌های کاربر را در مراکز داده شخص ثالث ذخیره می‌کنند. این مراکز می‌توانند این داده‌ها را با سایرین اشتراک بگذارند و داده‌های کاربر را بدون اطلاع آنها در اختیار سازمان‌های دیگر قرار دهند. مالکان شبکه‌های اجتماعی می‌توانند داده‌های کاربران را در اختیار آژانس‌های اطلاعاتی قرار دهند و یا از آنها در زمینه‌های تبلیغاتی استفاده کنند. در این تحقیق در بخش اول مقدمه موضوع را شرح داده‌ایم و پس از آن در بخش دوم به ادبیات موضوع پرداخته‌ایم که بیشتر تمرکز کارهای قبلی بروی روش‌های متمرکز حفظ حریم خصوصی می‌باشد و سپس در بخش سوم روش تحقیق را شرح داده‌ایم و روشی را ارائه داده‌ایم که با استفاده از بلاکچین حریم خصوصی کاربران در شبکه‌های اجتماعی تا حد زیادی افزایش پیدا می‌کند. در بخش چهارم اجرای تحقیق را شرح داده‌ایم و در نهایت به نتیجه‌گیری و جمع‌بندی از این تحقیق پرداخته‌ایم.

واژه‌های کلیدی: حریم خصوصی، امنیت اطلاعات، بلاکچین.

* رایانامه نویسنده مسؤول: hosseini@kntu.ac.ir

۱- مقدمه

در جدول ۱ به تعاریفی پرداخته شده است که رعایت آنها باعث افزایش حفظ حریم خصوصی کاربران در شبکه‌های اجتماعی می‌شود. جدول ۲ مطالعات گذشته بر روی مسائل حریم خصوصی را نشان می‌دهد.

در این تحقیق با استفاده از بلاکچین و الگوریتم رمزنگاری بلوفیش روشی را برای افزایش حفظ حریم خصوصی کاربران در شبکه‌های اجتماعی ارائه می‌دهیم. در بخش دوم تحقیق به ادبیات موضوع می‌پردازیم و پس از آن در بخش سوم مدل را شرح می‌دهیم و سپس در بخش چهارم مدل را اجرا کرده و در نهایت در بخش پنجم به جمع‌بندی می‌پردازیم.

۲- ادبیات موضوع

در این بخش از تحقیق در جدول ۳ به مروری بر ادبیات موضوع می‌پردازیم.

شبکه‌های اجتماعی در دنیای امروز بسیار محبوب هستند. شبکه‌های اجتماعی به افراد اجازه می‌دهند تا با دوستان و خانواده خود ارتباط برقرار کنند و اطلاعات خصوصی خود را به اشتراک بگذارند. بدین خاطر مسائل مربوط به حفظ حریم خصوصی کاربر نمایان می‌شود. تعداد کاربران شبکه‌های اجتماعی در سراسر جهان هر ساله به طور مداوم در حال افزایش است [17]. براساس گزارش امنیتی سوفوس^۱ در سال ۲۰۱۱، فیس بوک ۰٫۵ میلیارد کاربر دارد اما در حال حاضر (2023) بیش از ۲٫۸ میلیارد کاربر دارد. این گزارش نشان می‌دهد که فیس بوک بزرگ‌ترین خطرات امنیتی را دارد که به طور قابل توجهی از مای اسپیس، تویی تر، و لینکدین جلوتر است. این وب سایت محبوب‌ترین وب سایت برای کاربران فعال در وب است. با توجه به این محبوبیت، تعداد زیادی از کاربران توسط دشمنان از طریق انواع مختلفی از حملات مانند بدافزار، فیشینگ، اسپم‌نویسی و غیره مورد هدف قرار می‌گیرند این حملات به طور مداوم در حال افزایش هستند.

جدول ۱. تعاریف موارد حفظ حریم خصوصی

مقاوم در برابر حملات اختلال سرویس	به معنای توانایی یک سیستم یا شبکه در مقابل حملات و اختلالات مختلف بوده و قادر به حفظ کارایی و عملکرد صحیح خود در مقابل این حملات باشد. این شامل حفظ دسترسی، حفظ امنیت و حفظ کارایی سرویس‌ها می‌شود.
جلوگیری از ذخیره داده در سرورها	به معنای اعمال سیاست‌ها یا محدودیت‌ها برای جلوگیری از ذخیره و نگهداری داده‌ها در سرورها است.
رمزگذاری داده‌ها	داده‌ها قبل از ذخیره‌سازی در پایگاه‌های داده به صورت رمزگذاری شده ذخیره می‌شوند.
حفاظت داده در برابر مالک شبکه اجتماعی	به معنای عدم توانایی دخالت مالکان شبکه‌های اجتماعی در داده‌های کاربران است.
محرمانه بودن داده‌ها	به معنای حفظ حریم خصوصی است. وقتی داده‌ها به عنوان محرمانه تشخیص داده می‌شوند، این به معنای این است که دسترسی به آنها تنها برای افراد یا سازمان‌های مجاز ممکن است.
اطلاعات درهم‌سازی شده	منظور هش است که یک تابع ریاضی است که یک ورودی را به یک مقدار خروجی ثابت و یکتا تبدیل می‌کند.
غیرمتمرکز بودن	شبکه‌هایی هستند که اطلاعات و ارتباطات بین کاربران بدون وابستگی به یک مرکز مشخصی انجام می‌شود.
تبادل کلید امن	به فرآیند تبادل کلیدهای رمزنگاری برای ایجاد ارتباط امن بین دو یا چند نهاد یا دستگاه می‌پردازد.
کنترل دسترسی دقیق	به مفهوم مدیریت و کنترل دسترسی به منابع و اطلاعات در یک سیستم یا شبکه است.
رمزگذاری سرتاسر	یک روش رمزنگاری است که در آن اطلاعات از زمان ارسال تا زمان دریافت، به صورت رمزنگاری شده باقی می‌ماند و تنها افرادی که دسترسی به کلید رمزگشایی دارند می‌توانند این اطلاعات را مشاهده کنند.

^۱ Sophos^۲ <https://tavaana.org/sites/default/files/>

جدول ۲. مطالعات گذشته بر روی مسائل حفظ حریم خصوصی

منبع	شرح
[25]	یک سیستمی را با استفاده از فناوری تشخیص چهره ارائه داده‌اند در صورتی که کاربری رضایت نداشته باشد از اشتراک عکس، عکس آن در صورت به اشتراک گذاشته شدن به صورت بلوری و یا غیرقابل تشخیص نشان داده می‌شود. روش ارائه شده از نقص حریم خصوصی یک کاربر توسط کاربر دیگری جلوگیری می‌کند اما روش ارائه شده فاقد کارایی در مقابل نقص حریم خصوصی توسط مالک شبکه است. همین طور به دیگر ابعاد حفظ حریم خصوصی مانند مقاوم در برابر حملات اختلال سرویس، غیرمتمرکز بودن، تبادل کلید امن، کنترل دسترسی دقیق و... پرداخته نشده است.
[26]	یک طرحی ارائه شده است که با استفاده از الگوریتم‌های رمزنگاری به حفظ داده‌ها و افزایش حریم خصوصی کاربران در شبکه‌های اجتماعی می‌پردازد. اما در این طرح بعضی از ابعاد حفظ حریم خصوصی مانند مقاوم در برابر حملات اختلال سرویس، جلوگیری از ذخیره داده در سرورها، حفاظت داده در برابر مالک شبکه اجتماعی و ... پرداخته نشده است.
[1]	مسائل حفظ حریم خصوصی عمده در شبکه‌های اجتماعی را به چهار دسته تقسیم کرده‌اند: (الف) مسائل حریم خصوصی، (ب) بازاریابی ویروس، (ج) حملات ساختاری شبکه و (د) حملات مخرب تحقیق آن‌ها شامل یک بحث عمیق در مورد هر موضوع و مکانیزم‌های دفاعی مربوطه بود.
[17]	مسائل امنیتی و حریم خصوصی کلی در شبکه‌های اجتماعی را مورد بررسی قرار داده‌اند و همین‌طور تکنیک‌های مختلفی را مورد بحث قرار دادند.
[2]	یک بررسی از تهدیدهای امنیتی و حریم خصوصی مختلف در شبکه‌های اجتماعی ارائه کرده است.
[58]	ارائه یک راهکار برای نبرد با حملات neighbors با استفاده از یک تکنیک دو مرحله‌ای جهت ناشناس سازی در شبکه‌های اجتماعی
[59]	سناریوهای مختلف مربوط به تهدیدات شبکه اجتماعی آنلاین و راه‌حل‌های آن‌ها را با استفاده از مدل‌ها، چارچوب‌ها و تکنیک‌های رمزگذاری مختلف ارائه کرده و با این کار از کاربران شبکه اجتماعی در برابر حملات مختلف محافظت می‌کند. علاوه بر این راه‌حل‌های مختلف را مورد تجزیه و تحلیل قرار داده است.
[60]	به بررسی امنیت و حریم خصوصی در شبکه‌های اجتماعی می‌پردازد.
[62]	ارائه یک طبقه بندی در مورد چالش‌های حریم خصوصی و یک بررسی عمیق در مورد برخی از راه‌حل‌های ارائه شده اخیر در شبکه‌های اجتماعی.
[61]	به مروری برای پوشش تهدیدات ناشی از حریم خصوصی و امنیت اطلاعات در شبکه اجتماعی می‌پردازد و به صورت کلی حریم خصوصی را مورد بحث قرار می‌دهد، امنیت اطلاعات را با تمرکز بر تهدیدات مهندسی اجتماعی مورد بحث قرار می‌دهد و تهدیدات حساب‌های جعلی، سرقت هویت و نیز فیشینگ به طور خاص در شبکه‌های اجتماعی مورد بحث قرار می‌گیرد. در نهایت تلاش می‌کند تا دستورالعملی برای حفظ حریم خصوصی و تهدیدات فعلی آن در شبکه‌های اجتماعی ارائه دهد.

جدول ۳. مروری بر ادبیات موضوع

منبع	یافته‌ها	روش تحقیق	هدف تحقیق	سال	نویسنده
[39]	توییت و فیس‌بوک حریم خصوصی را حفظ نمی‌کنند.	ارائه مدلی منطقی برای مشخص کردن ویژگی‌ها و استدلال درباره دانش کاربران و یک زبان رسمی برای نوشتن سیاست‌های حفظ حریم خصوصی.	ارائه چارچوبی را برای نوشتن خط‌مشی‌های حفظ حریم خصوصی و استدلال درباره چنین سیاست‌هایی که شبکه را تکامل می‌دهند.	۲۰۱۷	رائول پاردو و همکاران
[26]	طراحی ایمن و کلیدها به صورت امن مبادله می‌شوند.	مدلی با رمزگذاری مبتنی بر ویژگی‌های خط‌مشی رمزنگاری (CP-ABE) و استاندارد رمزگذاری AES	رفع چالش افشا و دسترسی غیرمجاز به اطلاعات، داده‌ها و ارتباط بین کاربران در شبکه‌های اجتماعی.	۲۰۲۲	سید محمد صفی و همکاران
[40]	کاربران می‌توانند افشا و محافظت از محتوای خود را کنترل کنند	ارائه معماری مبتنی بر یک پلت فرم مدیریت حقوق باز که این معماری مکانیسم‌های امنیتی و حریم خصوصی لازم را اعمال می‌کند.	کنترل بیشتر کاربر بعد از قرارگیری داده‌ها بر روی شبکه‌ی اجتماعی	۲۰۱۳	خواکیم مارکزا و همکاران
[41]	عدم اطلاع بسیاری از کاربران در مورد تنظیمات حریم خصوصی	طرح سؤالات پرسش‌نامه‌ای، سؤالات در زمینه حفظ حریم خصوصی کاربران در شبکه‌های اجتماعی از بین دانشجویان کارشناسی.	ارائه توصیه‌هایی به کاربران شبکه‌های اجتماعی جهت افزایش حفظ حریم خصوصی آنها.	۲۰۱۸	شوکت علی و همکاران
[42]	کاربران از مسائل حریم خصوصی آشنا نیستند، افزایش اعتماد به مالکان شبکه‌های اجتماعی	ارائه مدلی از حفظ حریم خصوصی که توسط کاربر تنظیم می‌شود و مالک شبکه اجتماعی به آن نظارت می‌کند.	جلوگیری از افشای اطلاعات حساس کاربران در شبکه‌های اجتماعی	۲۰۱۴	الکینی و همکاران
[25]	عملکرد آن در ۸۷.۳۵٪ رضایت بخش بوده است.	با استفاده از فناوری تشخیص چهره جهت بلوری کردن تصاویر غیرمجاز.	حفظ حریم خصوصی در اشتراک‌گذاری تصاویر در شبکه‌های اجتماعی	۲۰۱۵	پاناجیوتیس و همکارانش
[43]	عملی بودن بر اساس آزمایش‌های انجام شده.	روشی بر اساس رمزگذاری همومورفیک ساخته شده است و به کاربر این امکان را می‌دهد تا با کمک چندین سرور نمایه‌های مشابه کاربر را پیدا کند.	حفظ حریم خصوصی در تطبیق پروفایل در شبکه‌های اجتماعی	۲۰۲۰	ژون یی و همکاران
[44]	تخمین بهتر از اندازه‌گیری افشای حریم خصوصی کاربران در چندین شبکه	با استفاده از سیستم‌های آماری و فازی، عوامل اصلی تأثیرگذار بر حریم خصوصی کاربران، یعنی حساسیت و دید را شناسایی می‌کند تا امتیاز افشای نهایی را برای هر کاربر به دست آورد.	اندازه‌گیری ریسک و افشای حریم خصوصی کاربران در چندین شبکه اجتماعی	۲۰۱۷	عرفان آقاسیان و همکاران

منبع	یافته‌ها	روش تحقیق	هدف تحقیق	سال	نویسنده
	اجتماعی				
[45]	فراهم شدن ناشناس‌سازی، اثربخشی و سودمندی عملی	یک رویکرد ناشناس‌سازی آگاه از ساختار را پیشنهاد می‌کند که ساختار شبکه اصلی و همچنین ویژگی‌های ساختاری آن را درحالی‌که ناشناس می‌کند، حداکثر حفظ هم می‌کند.	حفظ حریم خصوصی با استفاده از ناشناس‌سازی گراف	۲۰۰۹	زایون هی و همکاران
[46]	حفظ حریم خصوصی در زمینه تعقیب سایبری.	دو مدل کلیدی حریم خصوصی تعریف شده است، یعنی عدم پیوند وزن لبه و عدم پیوندناپذیری گره برای جلوگیری از پیوند اطلاعات کمکی به یک فرد هدف با احتمال بالا.	حفظ حریم خصوصی در تعقیب سایبری در شبکه‌های اجتماعی	۲۰۲۱	کاه منگ و همکاران
[47]	یکنواختی چارچوب زمینه مشترکی را برای مقایسه مدل‌های حمله موجود در اختیار قرار می‌دهد.	یک چارچوب منطقی که می‌تواند مدل‌های حمله مختلف را به طور یکسان نشان دهد.	حفظ حریم خصوصی در زمینه انتشار داده‌های شبکه اجتماعی.	۲۰۱۴	شنگ هسو و همکاران
[48]	نشان‌دادن عوامل اساسی پشت تهدیدات، توصیه‌های فنی برای بهبود حریم خصوصی.	مزایای تجاری و اجتماعی استفاده ایمن و آگاهانه از شبکه‌های اجتماعی را برجسته می‌کند و بر مهم‌ترین تهدیدات برای کاربران تأکید می‌کند.	تهدیدات شبکه‌های اجتماعی	۲۰۰۹	عبدالله الحصب
[49]	غیرمتمرکز و مبتنی بر اتریوم برای تأیید اثربخشی.	تغییرات در تنظیمات حریم خصوصی کاربر شبکه اجتماعی توسط یک قرارداد هوشمند تأیید می‌شود تا اطمینان حاصل شود که با انتظارات کاربران مطابقت دارد.	ارائه رویکرد غیرمتمرکز مبتنی بر بلاکچین برای مدیریت تنظیمات حریم خصوصی یک کاربر.	۲۰۲۱	جیانلوکا لاکس و همکاران
[50]	بلاکچین به‌عنوان ابزاری برای مدیریت کنترل دسترسی	مدلی ارائه می‌دهد که مبتنی بر بلاکچین به‌جای نقش محتوا نقش کاربر را به‌عنوان مرکز سیستم در نظر می‌گیرد.	ارائه روشی برای مدیریت مشکل کنترل دسترسی در شبکه‌های اجتماعی مبتنی بر بلاکچین.	۲۰۲۰	لارگو برونو

خود جلب کرده است. راه‌حل‌های مختلفی برای مقابله با این تهدیدات پیشنهاد شده است. در این قسمت، در جدول ۴ روش‌ها و رویکردهای مختلفی که در مقالات مربوط به حریم خصوصی در شبکه‌های اجتماعی ارائه شده‌اند مورد بحث و بررسی قرار گرفته است.

۲-۱- راه‌حل‌های حفظ حریم خصوصی در شبکه‌های اجتماعی

در چند سال گذشته، نقض حریم خصوصی در شبکه‌های اجتماعی توجه بسیاری از محققان در هر دو حوزه صنعتی و دانشگاهی را به

جدول ۴: خلاصه‌ای از روش‌های جلوگیری از نقض حریم خصوصی در شبکه‌های اجتماعی [17]

منابع	شرح	راه‌حل
[5,6,7,8,9]	روشی برای جاسازی داده‌ها در محتوای رسانه باهدف اثبات مالکیت محتوای رسانه است.	واترمارکینگ
[10,11,12]	به کاربران متعددی اجازه می‌دهد تا سیاست‌های حفظ حریم خصوصی خود را بر روی فیلم‌ها، تصاویر مشترک با مالکیت مشترک اعمال کنند.	مالکیت مشترک
[13,14,15]	مکانیزمی برای یافتن اطلاعات مخرب در داده‌های چندرسانه‌ای است.	استگانالیز
[16,18,19]	اطلاعاتی را که در اختیار ارائه‌دهنده خدمات شخص ثالث مانند ارائه‌دهندگان خدمات ابری قرار می‌دهند، رمزگذاری شوند.	رمزگذاری ذخیره‌سازی
[20,21,22]	مکانیسم‌های متفاوتی برای تشخیص انتشار بدافزار در شبکه‌های اجتماعی وجود دارد.	تشخیص بدافزار
[23,24,28,29]	ابزارها و تکنیک‌هایی برای شناسایی پروفایل‌های جعلی و دفاع در برابر حملات سیبیل ایجاد شده است. اکثر تکنیک‌ها به نمودارهای اجتماعی یا مفهوم مسیرهای تصادفی متکی هستند.	شناسایی پروفایل‌های جعلی و دفاعی سیبیل
[30,31,32,17]	رویکردهای موجود برای شناسایی اسپم در شبکه‌های اجتماعی استخراج یک مجموعه ویژگی است که کاربران اسپم را از افراد مجاز جدا کرده و آن ویژگی را به مدل‌های مختلف طبقه‌بندی کننده یادگیری ماشین برای شناسایی فعالیت‌های نامناسب ارائه می‌دهد.	تشخیص اسپم
[17,33]	توسط چندین شرکت امنیتی برای محافظت از کاربران شبکه‌های اجتماعی در برابر تهدیدات امنیتی تولید شده‌اند.	راه‌حل‌های تجاری
[34,35,17,36]	شبکه‌های اجتماعی راه‌حل‌های امنیتی داخلی مختلفی مانند تنظیمات حریم خصوصی کاربر، مکانیزم‌های مجوز را ارائه می‌دهند.	راه‌حل‌های داخلی در شبکه‌های اجتماعی
[37,17,38]	بسیاری از شبکه‌های اجتماعی ویژگی‌ای دارند که به‌صورت خودکار مشخصات شبیه‌سازی شده را شناسایی کرده و به کاربران در مورد چنین نمایه‌ای اطلاع می‌دهد.	تشخیص شبیه‌سازی مشخصات

۲-۲- شکاف ادبیات موضوع

امروزه شبکه‌های اجتماعی به کاربران خود اجازه می‌دهند تا حریم خصوصی خود را خودشان تنظیم کنند البته در سال ۲۰۱۹ یک مطالعه انجام شده است که نشان می‌دهد بسیاری از کاربران نمی‌دانند و یا نمی‌توانند چگونه حریم خصوصی خود را تنظیم کنند [3]. روش‌های مختلفی برای شناسایی افراد یا احراز هویت در شبکه‌های اجتماعی وجود دارد اما رایج‌ترین آنها استفاده از شماره همراه می‌باشد به طوری که ابتدا کاربر شماره همراه خود را وارد می‌کند و یک کد تایید برای او ارسال می‌شود و پس از ثبت نام می‌تواند با نام کاربری و رمز عبور وارد حساب کاربری خود شود. در روش پیشنهادی همین روش انجام شده است. اما تهدید برون سپاری و شفافیت مراکز داده یکی از تهدیدهای نقض حریم خصوصی در شبکه‌های اجتماعی محسوب می‌شود [17]. چرا که ممکن است داده‌های ارسالی ما در شبکه‌های اجتماعی مورد دستکاری قرار بگیرد یا داده‌های ما در شبکه‌های اجتماعی از طرف شرکت ارائه دهنده خدمات یا مالک شبکه اجتماعی بفروش برسد. برای جلوگیری از این کار ما در این تحقیق بلاکچین و برای رمزگذاری داده‌ها الگوریتم بلوفیش را پیشنهاد می‌کنیم.

۳- مدل پیشنهادی

در این بخش مدل پیشنهادی خود را تشریح می‌کنیم شکل ۱ نمای کلی مدل پیشنهادی را نشان می‌دهد.

استفاده از بلاکچین باعث می‌شود که مالک داده‌های تولید شده در شبکه‌های اجتماعی خود کاربر باشد و دیگر مالکین شبکه‌های اجتماعی قادر به دستکاری آنها و فروش آنها نیستند. یعنی شبکه اجتماعی کاربر محور می‌شود و هر کس مالک داده‌های خود است [51]. با توجه به اینکه در این مدل حفظ حریم خصوصی و اجرای دقیق و امن قراردادهای هوشمند از اهمیت بالایی برخوردار است روش مورد نظر ما برای ذخیره داده‌ها در بلاکچین درون زنجیر^۱ است.

الگوریتم گواه اثبات کار^۲، در حقیقت یک نوع الگوریتم اجماع است که از حملات DDoS و سایر سوءاستفاده‌ها در شبکه و همچنین از ایجاد اسپم در شبکه جلوگیری می‌کند. الگوریتم اثبات کار در بلاکچین و رمزرها کاربردهای بسیار زیادی دارد.



شکل ۱. نمای کلی مدل پیشنهادی

به صورت کلی بلاکچین‌های مختلفی وجود دارد که هر کدام کاربردهایی دارند. بلاکچین عمومی معروف‌ترین است، به عنوان مثال می‌توان بیت‌کوین را به عنوان بلاکچین عمومی نام برد. بلاکچین‌های عمومی به شفافیت و امنیت شهرت دارند و هیچ نهادی وجود ندارد که شبکه را کنترل کند. اما بلاکچین خصوصی سریع‌تر است ولی امکان مشاهده و کنترل داده‌ها به علت ماهیت خصوصی بودن آن وجود دارد [63]. Steemit یک شبکه اجتماعی مبتنی بر بلاکچین عمومی می‌باشد. در مدل پیشنهادی بلاکچین عمومی مدنظر است.

مزایا:

- بلاکچین می‌تواند مکانیسم‌های کنترل کاربر بهتری ایجاد کند
- حفظ تقریباً همه ابعاد حریم خصوصی
- غیرمتمرکز بودن، این امر باعث می‌شود که کاربران آزادی بیشتری داشته باشند و از سانسورها در امان باشند.
- مالکیت بر داده‌های شخصی و کنترل بهبودیافته بر محتوای تولید شده توسط کاربر
- بلاکچین می‌تواند امکانات جدیدی را برای شبکه‌های اجتماعی ارائه دهد. به عنوان مثال، این شبکه‌ها می‌توانند از قراردادهای هوشمند برای ایجاد اقتصادهای درون پلتفرمی و افزایش حریم خصوصی استفاده کنند.

معایب:

- ممکن است از تکنولوژی بلاکچین برای عملیات مجرمانه استفاده شود.
- سرعت پایین

۳-۱- عملکرد مدل پیشنهادی

در این قسمت عملکرد روش پیشنهادی که در شکل ۱ نمایش داده

² Proof Of Work

¹ On-Chain

حفظ می‌شود. با توجه به تحقیقات قبلی در عمل الگوریتم بلوفیش نسبت به دیگر الگوریتم‌های متقارن برای رمزگذاری فایل‌های تصویری سرعت بهتری نشان داده است [4,53,54,55,56,57]. البته الگوریتم‌های سریعتری هم نسبت به بلوفیش وجود دارد مثل ChaCha20 که ما از این الگوریتم به علت کوتاه‌تر بودن تاریخچه و محبوب نبودن آن نسبت به بلوفیش استفاده نکرده‌ایم. با توجه به این که در شبکه‌های اجتماعی از فایل‌های تصویری زیادی استفاده می‌شود ترجیح داده شد از این الگوریتم استفاده شود.

۳-۳- مدیریت کلید در مدل پیشنهادی

سروری که کلیدها بر روی آن قرار گرفته‌اند سروری است که تمام پروتکل‌های امنیتی بر روی آن اعمال شده است. کلیدها به صورت روزانه توسط سیستم تولید می‌شود و بر روی جدولی در پایگاه‌داده قرار می‌گیرند. یعنی پیام‌ها در هر روز با یک کلید متفاوت رمزگذاری می‌شوند و هنگام رمزگشایی با توجه به تاریخ پیام با کلید آن روز رمزگشایی می‌شوند. شکل ۲ شمای کلی مدیریت کلید در مدل پیشنهادی را نشان می‌دهد. بدست آوردن کلید برای رمزگذاری و یا رمزگشایی در سه گام انجام می‌شود:

۱. درخواست کاربر از طریق برنامه به سرور کلیدها ارسال می‌شود.
۲. سرور درخواست را به سیستم احراز مجوز دسترسی ارسال می‌کند.
۳. سیستم احراز مجوز دسترسی درخواست را بررسی می‌کند در صورتی که درخواست احراز شد و مجوز دریافت کرد با توجه به تاریخ پیام کلید را انتخاب و در نهایت کلید را ارسال می‌کند.

۳-۴- امضای دیجیتال

امضای دیجیتال در سه گام انجام می‌شود:

۱. هش کردن
اولین گام هش کردن داده‌ها است. داده‌ها از طریق یک الگوریتم هش مانند SHA256 هش می‌شوند. اندازه یک پیام می‌تواند به طور قابل توجهی متفاوت باشد، اما پس از هش کردن، تمام مقادیر هش آن طول یکسانی خواهند داشت.



شکل ۲. مدیریت کلید در مدل پیشنهادی

شده است را به صورت گام به گام تشریح می‌کنیم:

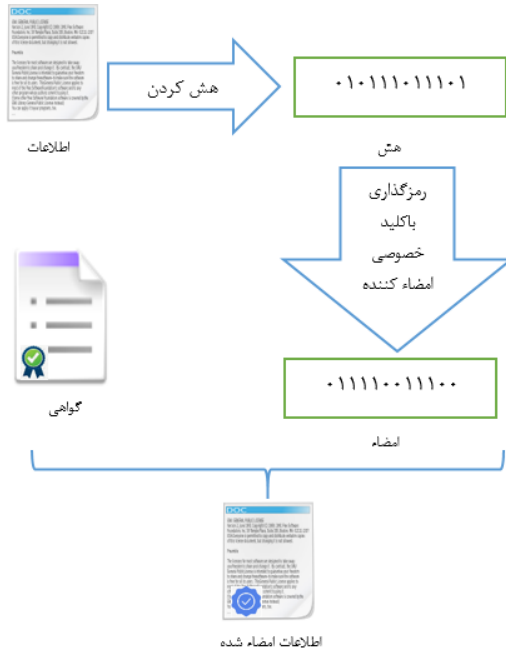
۱. کاربران در شبکه اجتماعی ثبت‌نام می‌کنند.
برای ثبت‌نام کاربران ابتدا باید شماره تلفن همراه خود را احراز هویت کنند. پس از ثبت شماره تلفن همراه کاربران در شبکه اجتماعی، کد احراز هویت برای کاربر ارسال می‌شود که از طریق آن کاربر می‌تواند اصالت شماره تلفن همراه خود را اثبات کند. پس از تأیید شماره تلفن همراه کاربران، این شماره در پایگاه‌داده ذخیره می‌شود. شماره تلفن همراه کاربران به دلیل منحصر به فرد بودن، کاربران را از یکدیگر متمایز می‌کند. پایگاه‌داده شبکه اجتماعی یک پایگاه‌داده توزیعی بلاکچینی است که اطلاعات بر روی آن قرار می‌گیرد.
۲. تنظیم پروفایل
بعد از ثبت‌نام کاربر پروفایل خود را تنظیم می‌کند یعنی حریم خصوصی خود را تنظیم می‌کند. این بخش با استفاده از قراردادهای هوشمند انجام می‌شود. تنظیمات حریم خصوصی می‌تواند شامل: پنهان کردن استوری، مسدودسازی عده‌ای از کاربران و چه کسانی او را ببینند و دیگر اطلاعات مربوط به حریم خصوصی باشد.
۳. ذخیره حریم خصوصی تنظیم شده
حریم خصوصی تنظیم شده با الگوریتم بلوفیش رمز و جهت ذخیره به بلاکچین ارسال و آنجا ذخیره خواهد شد.
۴. ارسال و دریافت داده‌ها
حال کاربر به استفاده امن از شبکه اجتماعی مبتنی بر بلاکچین می‌کند. شایان ذکر است کلیه داده‌ها قبل از قرارگیری بر پایگاه‌داده با الگوریتم بلوفیش رمز می‌شوند.

۳-۲- روش رمزگذاری داده‌ها در مدل پیشنهادی

در این مدل قبل از اینکه داده‌ها بر روی دیتابیس بلاکچین قرار بگیرند با الگوریتم متقارن رمز می‌شوند. علت این امر این است که به هر دلیلی اگر داده‌های ما مورد حمله قرار گرفته‌اند (مانند حمله ۵۱ درصد) به داده‌ها آسیبی وارد نشود. روش رمزگذاری متقارن به این علت انتخاب شده است که سرعت بالاتری نسبت به نامتقارن دارد. البته روش‌های نامتقارن امنیتی بیشتری نسبت به متقارن دارد اما چون داده‌های ما توسط خود بلاکچین هش می‌شوند و خود بلاکچین امنیت را به شدت بالا می‌برد و در نتیجه سرعت برای ما حائز اهمیت است، به همین علت از متقارن استفاده کرده‌ایم.

انتخاب الگوریتم رمزگذاری به عوامل مختلفی از قبیل سرعت، امنیت، سادگی و ... بستگی دارد. در این مدل سرعت برای ما حائز اهمیت است، به همین علت الگوریتم رمزگذاری متقارنی که استفاده کرده‌ایم، بلوفیش است چراکه امنیت ما توسط بلاکچین به شدت

۲. امضاء



شکل ۳. فرایند امضاء

پس از هش کردن داده‌ها، فرستنده پیام باید آن را امضاء کند. پیام با کلید خصوصی فرستنده امضاء می‌شود و گیرنده پیام می‌تواند صحت پیام و امضاء فرستاده شده را با استفاده از کلید عمومی مربوطه که توسط امضاکننده ارائه شده، کنترل کند. هر دو کلید عمومی و خصوصی توسط فرستنده پیام تولید می‌شوند، اما فقط کلید عمومی با گیرنده به اشتراک گذاشته می‌شود. امضا دیجیتال در بلاکچین به طور مستقیم با محتوای هر پیام مرتبط است. بنابراین، هر پیام امضاء شده دیجیتال دارای امضای دیجیتال متفاوتی خواهد بود. امضایی که هیچ‌کس نمی‌تواند آن را جعل کند.

۳. تأیید

فرایند تأیید زمانی اتفاق می‌افتد که گیرنده با استفاده از کلید عمومی ارائه شده توسط فرستنده دریافت پیام و اعتبارسنجی امضاء را تأیید کند. گام‌های فوق در شکل ۳ و ۴ نمایش داده شده است.

۴- ارزیابی

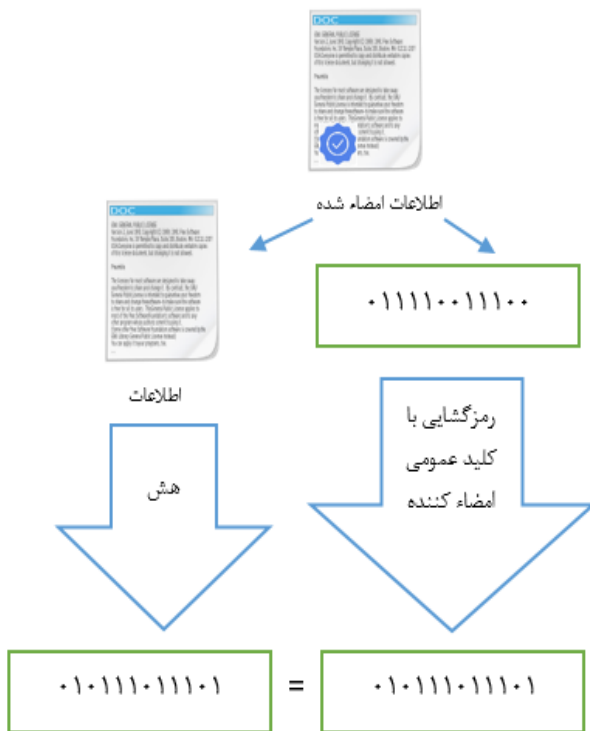
در این قسمت به اجرای تحقیق و ارزیابی روش پیشنهادی خود می‌پردازیم.

۴-۱- ارزیابی حفظ حریم خصوصی

عناصر مهم حریم خصوصی شامل محرمانه بودن داده‌ها، کنترل دسترسی دقیق است [26]. روش پیشنهادی به طور کامل هر دو عنصر را فراهم می‌کند:

• محرمانه بودن داده‌ها:

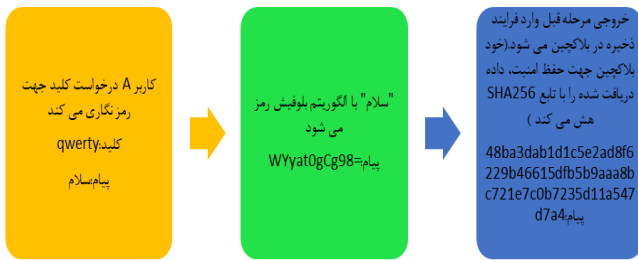
با توجه به این که مدل پیشنهادی از فناوری بلاکچین استفاده می‌کند و این فناوری برای ایجاد امنیت بیشتر از اساس به گونه‌ای طراحی شده است که برای تأیید یا ذخیره هر داده، به اعتبارسنجی نودهای مختلف شبکه احتیاج می‌باشد که از ویژگی‌های مهم برای امن نگه‌داشتن بلاکچین است. به این صورت یک عامل مخرب نمی‌تواند با دسترسی پیدا کردن به پایگاه داده اطلاعات ذخیره شده آن را مختل کند یا تغییر دهد، در نتیجه اطلاعات ثبت شده در شبکه بلاکچین به علت نیاز به امضای نودها برای تغییر، به نوعی تغییرناپذیر محسوب می‌شوند. اساس بلاکچین بر پایه رمزنگاری و حفظ امنیت اطلاعات بنا شده است و برای انتقال داده‌ها در آن از ویژگی هشینگ استفاده می‌شود و ما در این مدل قبل از هش داده‌ها ابتدا آنها را با الگوریتم بلوفیش رمز می‌کنیم و بعد رمز توسط بلاکچین با تابع هش SHA256 هش می‌شود. در هش هرگز امکان ندارد دو هش یکسان برای اطلاعات متفاوت ایجاد شود.



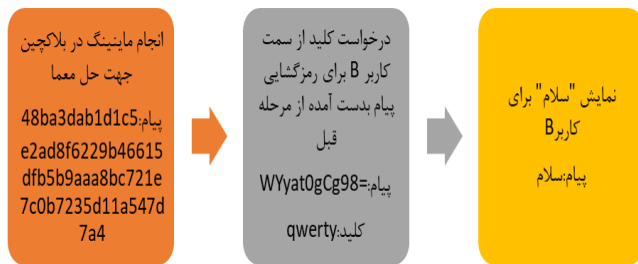
شکل ۴. فرایند تأیید

برای ساخت هر بلاک در بلاکچین نیز از هش بلاک قبل استفاده می‌شود و اگر شخصی (عامل مخرب) قصد ایجاد تغییر در اطلاعات ذخیره شده قبلی را داشته باشد، مجبور می‌شود تمام بلاک‌های بعد از آن را نیز از نو بسازد و هش کند. به علت این ویژگی فناوری بلاکچین، خرابکاری وی بلافاصله لو می‌رود.

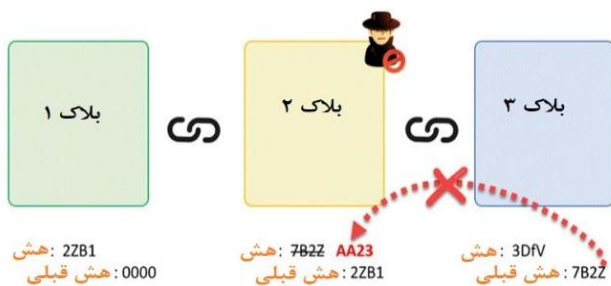
به عنوان مثال فرض کنید کاربر A می‌خواهد پیام "سلام" را به کاربر B ارسال کند مراحل آن در شکل ۵ شرح داده شده است. برای نمایش سلام برای کاربر B مراحل شکل ۶ طی می‌شود.



شکل ۵. ارسال اطلاعات از کاربر A به کاربر B



شکل ۶. نمایش پیام برای کاربر B



شکل ۷. تغییر اطلاعات توسط کاربر مخرب. [52]

گام‌های الگوریتم به شرح ذیل می‌باشند:

- ۱- دریافت داده برای رمزگذاری/رمزگشایی.
 - ۲- دریافت کلید از سرور کلیدها برای رمزگذاری/رمزگشایی.
 - ۳- رمزگذاری/رمزگشایی داده‌ها.
- شبه‌کد مربوطه در شکل ۸ نمایش داده شده است.

```
static void Main()
{
    string dataToEncryptDecrypt = "Your data to encrypt";
    byte[] keyFromServer = GetKeyFromServer();
    byte[] encryptedOrDecryptedData =
    EncryptOrDecryptedDataWithBlowFish(dataToEncryptDecr
    ypt, keyFromServer);
}

```

یک تابع برای دریافت کلید از سرور کلیدها است: GetKeyFromServer().
 :متدی که وظیفه‌ی EncryptOrDecryptedDataWithBlowFish رمزگذاری و رمزگشایی داده‌ها با الگوریتم بلوفیش را دارد.

شکل ۸. شبه‌کد الگوریتم رمزگذاری و رمزگشایی

با اجرای مراحل فوق و با استفاده از بلاکچین توانسته‌ایم داده‌های خود را نامتمرکز کنیم که این باعث می‌شود مالکان شبکه‌های اجتماعی نتوانند از داده‌های کاربران سوءاستفاده کنند.

با توجه به ذخیره‌سازی داده‌ها در بلاکچین و استفاده از مکانیزم‌های هش و رمزگذاری اگر کاربری قصد دستکاری پیام داشته باشد و به هر دلیلی موفق هم شود چون هش بلاک در بلاک قبلی است و با تغییر پیام توسط عامل مخرب، هش تغییر می‌کند و برای همه قابل مشاهده است که این پیام تغییر کرده است و به همین علت کاربر غیرمجاز امکان دستکاری در پیام را ندارد.

به عنوان مثال فرض کنید بلاک‌های شکل ۷ را داریم و کاربری قصد تغییر اطلاعات بلاک شماره ۲ را دارد و فرض می‌کنم می‌تواند این کار را انجام دهد زمانی که این کار را انجام می‌دهد هش بلاک تغییر می‌کند و با هش بلاک ۳ که هش بلاک قبلی را نگهداری می‌کند همخوانی ندارد و همه متوجه می‌شوند این بلاک تغییر پیدا کرده است.

• کنترل دقیق دسترسی:

مدل پیشنهادی با استفاده از قراردادهای هوشمند در بلاکچین کنترل دقیق دسترسی‌ها را انجام می‌دهد به عنوان مثال اگر یک کاربر یک کاربر دیگر را مسدود کرده باشد این کار با استفاده از قراردادهای هوشمند به درستی انجام می‌شود.

۴-۲- مقایسه طرح پیشنهادی با طرح‌ها دیگر

در این بخش، طرح پیشنهادی ما از عناصر حفظ حریم خصوصی و حملات متداول مانند حملات اختلال سرویس، حفاظت از داده‌ها در برابر مالکان شبکه‌های اجتماعی، تبادل کلید امن، رمزگذاری داده‌ها، کنترل دسترسی دقیق، برنامه متن‌باز، اطلاعات درهم‌سازی شده، غیرمتمرکز بودن و رمزگذاری سرتاسر با سایر طرح‌ها مقایسه شده است. نتیجه مقایسه در جدول ۴ نشان داده شده است.

۴-۳- ارزیابی عملکرد روش پیشنهادی

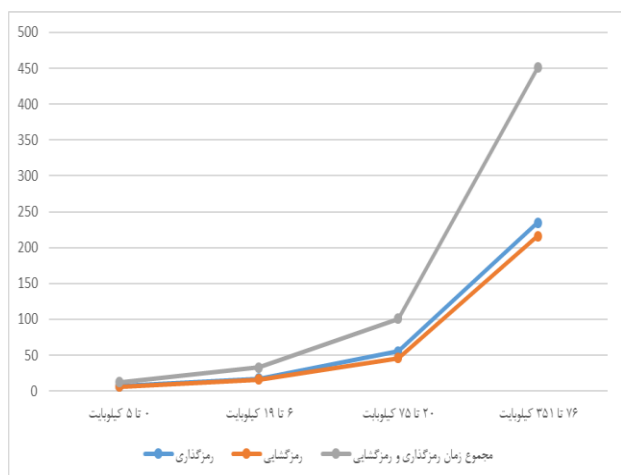
برای شبیه‌سازی یک برنامه‌ای را با استفاده از زبان سی‌شارپ پیاده‌سازی کرده‌ایم تا بتوانیم زمان رمزگذاری و رمزگشایی را محاسبه کنیم با استفاده از این می‌توانیم متن و تصاویر را رمزگذاری کنیم ما چند سناریوی زیر را برای شبیه‌سازی انجام می‌دهیم لازم به ذکر است شبیه‌سازی بر روی یک دستگاه با مشخصات زیر اجرا شده است:

CPU:AMD FX-7600P Radeon R7, 12 Compute Cores
 4C+8G 2.70 GHz, RAM:8GB

جدول ۴. مقایسه مدل پیشنهادی با دیگر طرح‌ها

مدل پیشنهادی	[26]	[46]	[43]	[25]	
مقاوم در برابر حملات اختلال سرویس	خیر	خیر	خیر	خیر	
جلوگیری از ذخیره داده در سرورها	خیر	خیر	خیر	خیر	
رمزگذاری داده‌ها	بله	-	بله	-	
حفاظت داده در برابر مالک شبکه اجتماعی	خیر	خیر	خیر	بله	
محرمانه بودن داده‌ها	بله	بله	بله	خیر	
اطلاعات درهم‌سازی شده	خیر	خیر	خیر	خیر	
غیرمتمرکز بودن	خیر	خیر	خیر	خیر	
تبادل کلید امن	بله	-	بله	-	
کنترل دسترسی دقیق	بله	بله	بله	خیر	
رمزگذاری سرتاسر	بله	خیر	خیر	خیر	

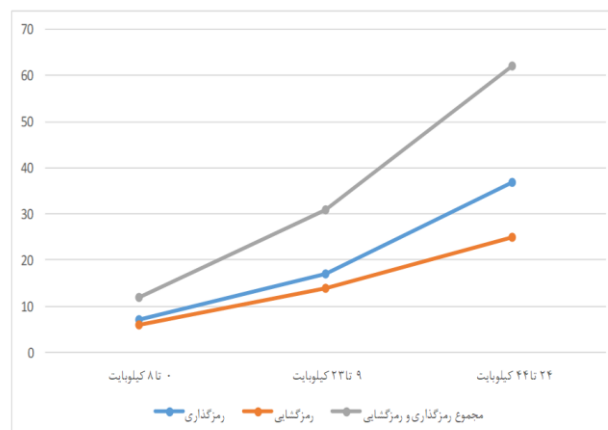
- ۱- زمان رمزگذاری تصاویر با حجم ۰ کیلوبایت تا ۵ کیلوبایت برابر است با ۴ تا ۷ میلی ثانیه
 - ۲- زمان رمزگشایی تصاویر با حجم ۰ کیلوبایت تا ۵ کیلوبایت برابر است با ۴ تا ۶ میلی ثانیه
 - ۳- زمان رمزگذاری تصاویر با حجم ۶ کیلوبایت تا ۱۹ کیلوبایت برابر است با ۸ تا ۱۷ میلی ثانیه
 - ۴- زمان رمزگشایی تصاویر با حجم ۶ کیلوبایت تا ۱۹ کیلوبایت برابر است با ۷ تا ۱۶ میلی ثانیه
 - ۵- زمان رمزگذاری تصاویر با حجم ۲۰ کیلوبایت تا ۷۵ کیلوبایت برابر است با ۱۸ تا ۵۵ میلی ثانیه
 - ۶- زمان رمزگشایی تصاویر با حجم ۲۰ کیلوبایت تا ۷۵ کیلوبایت برابر است با ۱۷ تا ۴۶ میلی ثانیه
 - ۷- زمان رمزگذاری تصاویر با حجم ۷۶ کیلوبایت تا ۳۵۱ کیلوبایت برابر است با ۵۶ تا ۲۳۵ میلی ثانیه
 - ۸- زمان رمزگشایی تصاویر با حجم ۷۶ کیلوبایت تا ۳۵۱ کیلوبایت برابر است با ۴۷ تا ۲۱۶ میلی ثانیه
- خروجی آن در شکل ۱۰ نمایش داده شده است.



شکل ۱۰. زمان رمزگذاری و رمزگشایی تصاویر بر حسب میلی ثانیه

سناریوی زیر برای متن در نظر گرفته شده است:

- ۱- زمان رمزگذاری متن با حجم صفر کیلوبایت تا ۸ کیلوبایت برابر است با ۴ تا ۷ میلی ثانیه
 - ۲- زمان رمزگشایی متن با حجم صفر کیلوبایت تا ۸ کیلوبایت برابر است با ۴ تا ۶ میلی ثانیه
 - ۳- زمان رمزگذاری متن با حجم ۹ کیلوبایت تا ۲۳ کیلوبایت برابر است با ۸ تا ۱۷ میلی ثانیه
 - ۴- زمان رمزگشایی متن با حجم ۹ کیلوبایت تا ۲۳ کیلوبایت برابر است با ۷ تا ۱۴ میلی ثانیه
 - ۵- زمان رمزگذاری متن با حجم ۲۴ کیلوبایت تا ۴۴ کیلوبایت برابر است با ۱۸ تا ۳۷ میلی ثانیه
 - ۶- زمان رمزگشایی متن با حجم ۲۴ کیلوبایت تا ۴۴ کیلوبایت برابر است با ۱۵ تا ۲۵ میلی ثانیه
- خروجی آن در شکل ۹ نمایش داده شده است.
- سناریوی زیر برای تصاویر در نظر گرفته شده است:



شکل ۹. زمان رمزگذاری و رمزگشایی متن بر حسب میلی ثانیه

۵- نتیجه گیری

شبکه‌های اجتماعی به یک رسانه ارتباطی مطلوبی برای میلیاردها کاربر وب تبدیل شده‌اند، به طوری که چنین سرویس‌هایی به مردم اجازه می‌دهند تا علایق، عکس‌ها، ویدئوها را به اشتراک بگذارند امروزه خیلی از کاربران شبکه‌های اجتماعی از این طریق کسب درآمد می‌کنند، شبکه‌های اجتماعی امروزه با کسب و کارها پیوند زده شده‌اند. اما این خدمات می‌توانند کاربران را در معرض خطرات جدی نقض حریم خصوصی قرار دهند. یکی از مهم‌ترین و بزرگترین مشکلات موارد نقض حریم خصوصی فروش و یا به اشتراک گذاری داده‌ها توسط مالکان شبکه‌های اجتماعی است که در این تحقیق روشی ارائه دادیم مبتنی بر بلاکچین و رمزنگاری تا از این نقض حریم خصوصی کاربران جلوگیری کند مدل پیشنهادی با چند طرح دیگر مقایسه و ارزیابی شد و نشان داده شد که مدل پیشنهادی ما سرعت رمزگذاری و رمزگشایی کمتری (کندتر) نسبت به طرح‌های مقایسه شده دارد اما رعایت و حفظ حریم خصوصی در طرح پیشنهادی ما نسبت به طرح‌های مقایسه شده بیشتر است. از محدودیت‌های طرح ما می‌توان به سرعت اینترنت مناسب برای اجرای صحیح طرح پیشنهادی که باید حداقل 4G یا بالاتر یا معادل آن باشد و همین‌طور دستگاه تلفن همراهی که این روش را به خوبی پشتیبانی می‌کند باید حداقل یک دستگاه متوسط از نظر سخت‌افزاری باشد.

در تحقیقات آینده ما سعی خواهیم کرد از طرح پیشنهادی و یا ترکیب بلاکچین و سیستم‌های متمرکز به عنوان درون‌زنجیر و برون‌زنجیر، برای طراحی و پیاده‌سازی یک پلتفرم شبکه اجتماعی غیرمتمرکز و مبتنی بر بلاکچین استفاده کنیم.

مراجع

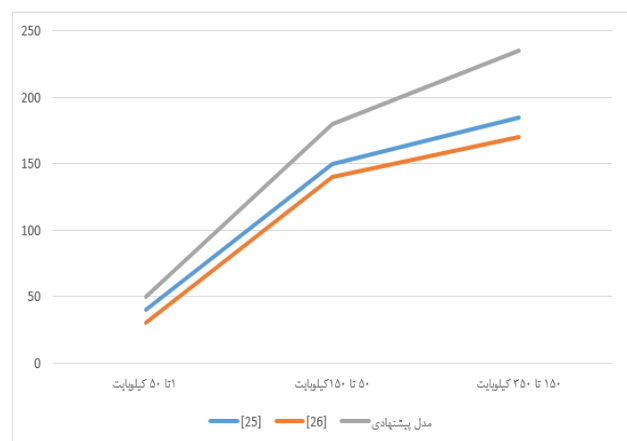
- [1] H. Gao , J. Hu , T. Huang , J. Wang , Y. Chen , "Security issues in online social networks", *IEEE Internet Computer*, Vol. 15(4), pp. 56-63, July-Aug. 2011, DOI: [10.1109/MIC.2011.50](https://doi.org/10.1109/MIC.2011.50).
- [2] M. Fire , R. Goldschmidt , Y. Elovici , "Online social networks: threats and solutions", *IEEE Communications Surveys & Tutorials*, vol. 16 (4), pp. 5-15, 02 May 2014, DOI: [10.1109/COMST.2014.2321628](https://doi.org/10.1109/COMST.2014.2321628).
- [3] Gianluca Lax , Antonia Russo, Lara Saidia Fasci, "A Blockchain-based approach for matching desired and real privacy settings of social network users", *Information Sciences*, vol. 557, pp. 220-235, May 2021, DOI: <https://doi.org/10.1016/j.ins.2021.01.004>.
- [4] Daa Salama Abd Elminaam, Hatem Mohamed Abdual Kader, and Mohiy Mohamed Hadhoud , "Evaluating The Performance of Symmetric Encryption Algorithms", *IJCSNS International Journal of Computer Science and Network Security*, VOL.8 No.12, December 2008, May 2010
- [5] [5] A.M. Alattar, N.D. Memon, C.D. Heitzenrater, "Media Watermarking", Available: <https://www.spiedigitallibrary.org/proceedings/Download?urlId=10.1117%2F12.2022495>
- [6] A. Zigomitos, A. Papageorgiou, C. Patsakis, "Social network content management through watermarking", in: *Proceedings of the 11th International Conference on Trust, Security and Privacy in Computing and Communications*, IEEE, 25-27 June 2012, DOI: [10.1109/TrustCom.2012.264](https://doi.org/10.1109/TrustCom.2012.264)

در این مدل باتوجه به اینکه ما از بلاکچین استفاده می‌کنیم حفظ حریم خصوصی کاربران تا حد بسیار زیادی افزایش پیدا می‌کند ضمن اینکه قبل از ذخیره داده در بلاکچین توسط الگوریتم بلوفیش رمز می‌شود و بعد از هش در بلاکچین ذخیره می‌شود.

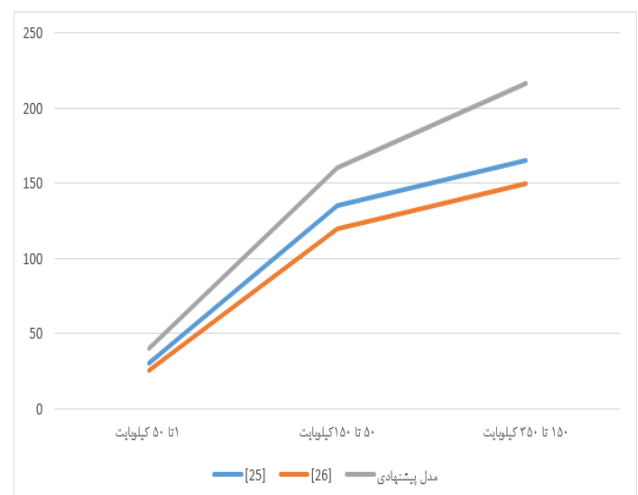
۴-۴- مقایسه طرح پیشنهادی با طرح‌های دیگر از نظر عملکرد و کارایی

در این بخش مدل پیشنهادی خود را با دو طرح دیگر از نظر زمان رمزگذاری و رمزگشایی مقایسه کرده‌ایم که نتایج آن در شکل ۱۱ و ۱۲ نشان داده شده است.

زمان رمزگذاری و رمزگشایی مدل پیشنهادی با دو طرح مقایسه شده بیشتر است. اما حفاظت بیشتری از حریم خصوصی انجام می‌دهد و با توجه به اینکه قبل از ذخیره‌سازی داده‌ها در بلاکچین، داده‌ها رمزگذاری می‌شوند و سپس بلاکچین هش می‌شوند و بعد در پایگاه داده بلاکچینی ذخیره می‌شوند حفاظت از حریم خصوصی بسیار بالا می‌رود و در نتیجه مدل پیشنهادی از حریم خصوصی نسبت به دو مدل مقایسه شده حفاظت بهتری می‌کند.



شکل ۱۱. زمان رمزگذاری برحسب میلی ثانیه



شکل ۱۲. زمان رمزگشایی برحسب میلی ثانیه

- [24] G. Wang, F. Musau, S. Guo, M.B. Abdullahi, "Neighbor similarity trust against sybil attack in P2P e-commerce", *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, pp. 824-833, March 2015, doi: [10.1109/TPDS.2014.2312932](https://doi.org/10.1109/TPDS.2014.2312932)
- [25] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, Sotiris Ioannidis, "Preventing Privacy Leakage From Photos in Social Networks", *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 781-792, October 2015, doi: <https://doi.org/10.1145/2810103.2813603>
- [26] Seyyed Mohammad Safi, Ali Movaghar, Mohammad Ghorbani, "Privacy protection scheme for mobile social network", *Journal of King Saud University - Computer and Information Sciences*, vol.34, pp. 4062-4074, July 2022, doi:<https://doi.org/10.1016/j.jksuci.2022.05.011>
- [27] Sun, X., Yao, Y., Xia, Y., Liu, X., Chen, J., Wang, Z., 2016. "Towards Efficient Sharing of Encrypted Data in Cloud-Based Mobile Social Network". *KSI Transactions on Internet and Information Systems*, VOL. 10, NO. 4, pp. 1892-1903, Apr. 2016
- [28] N. Tran, J. Li, L. Subramanian, S.S. Chow, "Optimal sybil-resilient node admission control", *Proceedings of the INFOCOM IEEE*, 10-15 April 2011, doi: [10.1109/INFCOM.2011.5935171](https://doi.org/10.1109/INFCOM.2011.5935171)
- [29] W. Wei, F. Xu, C.C. Tan, Q. Li, "SybilDefender: a defense mechanism for Sybil attacks in large social networks", *IEEE Trans*, vol. 24, pp. 2492-2502, December 2013, doi: [10.1109/TPDS.2013.9](https://doi.org/10.1109/TPDS.2013.9)
- [30] A. Hai Wang, "Don't follow me: spam detection in twitter", *Proceedings of the International Conference on Security and Cryptography (SECRYPT) IEEE*, 26-28 July 2010.
- [31] F. Ahmed, M. Abulaish, "A generic statistical approach for spam detection in Online Social Networks", *Computer Communications*, vol. 36, pp. 1120-1129, June 2013, doi: <https://doi.org/10.1016/j.comcom.2013.04.004>
- [32] H. Gao, Y. Chen, K. Lee, D. Palsetia, A.N. Choudhary, "Towards online spam filtering in social networks", 2012
- [33] M. Fire, R. Goldschmidt, Y. Elovici, "Online social networks: threats and solutions", *IEEE Communications Surveys & Tutorials*, vol. 16, pp. 2019-2036, 02 May 2014, doi: [10.1109/COMST.2014.2321628](https://doi.org/10.1109/COMST.2014.2321628)
- [34] D.H. Lee, "Personalizing information using users' online social networks: a case study of CiteULike", *Journal of Information Processing Systems*, vol. 11, pp. 1, 2015, doi: [10.3745/JIPS.04.0014](https://doi.org/10.3745/JIPS.04.0014)
- [35] D. Wang, N. Wang, P. Wang, S. Qing, "Preserving privacy for free: efficient and provably secure two-factor authentication scheme with user anonymity", *Information Sciences*, vol. 321, pp. 162-178, 10 November 2015, doi: <https://doi.org/10.1016/j.ins.2015.03.070>
- [36] T. Stein, E. Chen, K. Mangla, "Facebook immune system", *Proceedings of the 4th Workshop on Social Network Systems*, No. 8, pp. 1-8, April 2011, doi: <https://doi.org/10.1145/1989656.1989664>
- [37] G. Kontaxis, I. Polakis, S. Ioannidis, E.P. Markatos, "Detecting social network profile cloning", *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops) IEEE*, 21-25 March 2011, DOI: [10.1109/PERCOMW.2011.5766886](https://doi.org/10.1109/PERCOMW.2011.5766886)
- [38] Z. Shan, H. Cao, J. Lv, C. Yan, A. Liu, "Enhancing and identifying cloning attacks in online social networks", *Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication*, NO. 59, pp. 1-6, January 2013, doi: <https://doi.org/10.1145/2448556.2448615>
- [39] Raúl Pardo a, Musard Balliu a, Gerardo Schneider, "Formalising privacy policies in social networks, Journal of Logical and Algebraic Methods in Programming", *Journal of Logical and Algebraic Methods in Programming*, vol. 90, pp. 125-157, August 2017, doi: <https://doi.org/10.1016/j.jlamp.2017.02.008>
- [40] Joaquim Marquesa, Carlos Serrão, "Improving content privacy on social networks using open digital rights management solutions", *Procedia Technology*, vol. 9, pp. 405-410, 2013, doi: <https://doi.org/10.1016/j.protcy.2013.12.045>
- [7] C. Ho Sin, N.A. Kim, B.W. Go, K.S. Min, J.D. Lee, J.H. Park, "Realizing the Right to Be Forgotten in an SNS Environment", *Advances in Computer Science and its Applications*, vol. 279, pp. 1443-1449, (2014)
- [8] C. Patsakis, A. Zigomitos, A. Papageorgiou, E. Galván-López, "Distributing privacy policies over multimedia content across multiple online social networks" *Computer Networks*, vol. 75, pp. 531-543, 24 December 2014, DOI: <https://doi.org/10.1016/j.comnet.2014.08.023>.
- [9] K. Thongkor, N. Mettripun, T. Pramoun, T. Amornraksa, "Image watermarking based on DWT coefficients modification for social networking services", in: *Proceedings of the 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) IEEE*, 15-17 May 2013, DOI: [10.1109/ECTICon.2013.6559626](https://doi.org/10.1109/ECTICon.2013.6559626)
- [10] A.C. Squicciarini, H. Xu, X.L. Zhang, "CoPE: enabling collaborative privacy management in online social networks", *J. Am. Soc. Inf. Sci. Technol.*, vol. 62, 25 January 2011, DOI: <https://doi.org/10.1002/asi.21473>
- [11] A.C. Squicciarini, M. Shehab, J. Wede, "Privacy policies for shared content in social network sites", *VLDB J*, vol. 19, pp. 779-796, 30 June 2010.
- [12] H. Hu, G.J. Ahn, J. Jorgensen, "Multiparty access control for online social networks: model and mechanisms", *IEEE Trans*, vol. 25, pp. 1614-1627, 7 July 2013, DOI: [10.1109/TKDE.2012.97](https://doi.org/10.1109/TKDE.2012.97)
- [13] F. Li, K. Wu, J. Lei, M. Wen, Z. Bi, C. Gu, "Steganalysis over large-scale social networks with high-order joint features and clustering ensembles", *IEEE Trans Inf Forensic Secur*, vol. 11, pp. 344-357, 2 February 2016, DOI: [10.1109/TIFS.2015.2496910](https://doi.org/10.1109/TIFS.2015.2496910)
- [14] N. Venkatachalam, R. Anitha, "A multi-feature approach to detect Stegobot: a covert multimedia social network botnet", *Multimed Tools Appl*, vol. 76, pp. 6079-6096, 2017, DOI: [10.1007/s11042-016-3555-3](https://doi.org/10.1007/s11042-016-3555-3)
- [15] V. Natarajan, S. Sheen, R. Anitha, "Multilevel analysis to detect covert social botnet in multimedia social networks", *IEEE Comput*, vol. 58, pp. 679-687, April 2015, DOI: [10.1093/comjnl/bxu063](https://doi.org/10.1093/comjnl/bxu063)
- [16] M. Tierney, I. Spiro, C. Bregler, L. "Subramanian, Cryptagram: photo privacy for online social media", *Proceedings of the first ACM conference on Online social networks*, pp. 75-88, October 2013, doi: <https://doi.org/10.1145/2512938.2512939>
- [17] S. Rathore, P. Kumar Sharma, V. Loia, Y. Jeong, J. Park, "Social network security: Issues, challenges, threats, and solutions", *Information Sciences*, vol. 421, pp. 43-69, December 2017, doi: <https://doi.org/10.1016/j.ins.2017.08.063>.
- [18] P. Savla, L.D. Martino, "Content analysis of privacy policies for health social networks", *Proceedings of the International Symposium on Policies for Distributed Systems and Networks* 16-18 July 2012, DOI: [10.1109/POLICY.2012.20](https://doi.org/10.1109/POLICY.2012.20)
- [19] X. Liu, Q. Liu, T. Peng, J. Wu, "Dynamic access policy in cloud-based personal health record (PHR) systems", *Inf. Sci.*, vol. 379, pp. 62-81, 10 February 2017, doi: <https://doi.org/10.1016/j.ins.2016.06.035>
- [20] G. Yan, G. Chen, S. Eidenbenz, N. Li, "Malware propagation in online social networks: nature, dynamics, and defense implications" *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, vol. 7, pp. 196-206, March 2011, doi: <https://doi.org/10.1145/1966913.1966939>
- [21] H. Zhu, C. Huang, H. Li, "MPPM: Malware propagation and prevention model in online social network", *IEEE International Conference on Communications Workshops (ICC)*, 10-14 June 2014, doi: [10.1109/ICCW.2014.6881278](https://doi.org/10.1109/ICCW.2014.6881278)
- [22] W. Xu, F. Zhang, S. Zhu, "Toward worm detection in online social networks", *Proceedings of the 26th Annual Computer Security Applications Conference*, pp. 11-20, December 2010, doi: <https://doi.org/10.1145/1920261.1920264>
- [23] G. Danezis, P. Mittal, SybilInfer: "detecting sybil nodes using social networks", *NDSS*, 2009

- [53] Aamer Nadeem, Dr M. Younus Javed, "A Performance Comparison of Data Encryption Algorithms", *International Conference on Information and Communication Technologies*, 27-28 August 2005, doi: [10.1109/ICICT.2005.1598556](https://doi.org/10.1109/ICICT.2005.1598556)
- [54] D. Commey, S. Griffith Klogo, J. Dzisi Gadze, "Performance comparison of 3DES, AES, Blowfish and RSA for Dataset Classification and Encryption in Cloud Data Storage", *International Journal of Computer Applications*, vol. 177, February 2020
- [55] Chandrashekhar B, Dr. Mohamed Abdul Waheed, "Analysis of Possible Attacks on Data and Possible Solutions with Comparative Analysis of Various Encryption Algorithms and Evaluation", *International Journal of Innovative Research in Engineering & Management (IJIREM)*, vol. 9, April 2022, doi: <https://doi.org/10.55524/ijirem.2022.9.2.7>
- [56] C. Rathod, A. Gonsai, "Performance Analysis of AES, Blowfish and Rijndael: Cryptographic Algorithms for Audio", *Rising Threats in Expert Applications and Solutions*, pp. 203–209, 02 October 2020, doi: [10.1007/978-981-15-6014-9_24](https://doi.org/10.1007/978-981-15-6014-9_24)
- [57] H. Alabdulrazzaq, M. Alenezi, "Performance Analysis and Evaluation of Cryptographic Algorithms: DES, 3DES, Blowfish, Twofish, and Threefish", *International Journal of Communication Networks and Information Security (IJCNIS)*, vol. 14, No. 1, April 2022
- [58] Bin Zhou, Jian Pei, "Preserving Privacy in Social Networks Against Neighborhood Attacks", *IEEE 24th International Conference on Data Engineering*, 07-12 April 2008, doi: [10.1109/ICDE.2008.4497459](https://doi.org/10.1109/ICDE.2008.4497459)
- [59] A. Jain, S. Ranjan Sahoo, J. Kaubiyal, "Online social networks security and privacy: comprehensive review and analysis", *Complex & Intelligent Systems*, Vol. 7, pp. 2157–2177, 01 June 2021, doi: [10.1007/S40747-021-00409-7](https://doi.org/10.1007/S40747-021-00409-7)
- [60] Seyed Hossein Mousavi, Hamid Barati, "SECURITY AND PRIVACY IN SOCIAL NETWORKS", *Journal of Positive School Psychology*, Vol. 6, No. 5, 2022
- [61] Ahmed Al-Charchafchi, S. Manickam, Zakaria N. M. Alqattan, "Threats Against Information Privacy and Security in Social Networks: A Review", *Springer Nature Singapore Pte Ltd*, pp. 358–372, 2020, doi: https://doi.org/10.1007/978-981-15-2693-0_26
- [62] R. Malekhosseini, M. Hosseinzadeh, K. Navi, "An investigation into the requirements of privacy in social networks and factors contributing to users' concerns about violation of their privacy", *Social Network Analysis and Mining*, 11 June 2018, doi: <https://doi.org/10.1007/s13278-018-0518-x>
- [63] P.K. Paul, P.S. Aithal, R. Saavedra, Su. Ghosh, "Blockchain Technology and its Types—A Short Review", *International Journal of Applied Science and Engineering*, December 2021
- [41] Shaukat Ali, Naveed Islam, Azhar Rauf, Ikram Ud Din and Mohsen Guizani and Joel J. P. C. Rodrigues, "Privacy and Security Issues in Online Social Networks", *future internet*, 2018
- [42] Nader Yahya Alkeinaya, Norita Md. Norwawi, "User Oriented Privacy Model for Social Networks", *Procedia - Social and Behavioral Sciences*, vol. 129, pp. 191-197, 15 May 2014, doi: <https://doi.org/10.1016/j.sbspro.2014.03.666>
- [43] Xun Yi, Elisa Bertino, Fang-Yu Rao, Kwok-Yan Lam, Surya Nepal and Athman Bouguettaya, "Privacy-Preserving User Profile Matching in Social Networks", *IEEE*, vol. 32, pp. 1572-1585, 01 August 2020, doi: [10.1109/TKDE.2019.2912748](https://doi.org/10.1109/TKDE.2019.2912748)
- [44] Erfan Aghasian, Saurabh Garg, (Member, IEEE), Longxiang Gao, (Member, IEEE), Shui Yu, Senior Member, IEEE) and James Montgomery, (Member, IEEE), "Scoring Users' Privacy Disclosure Across Multiple Online Social Networks", *IEEE Access*, vol. 5, 27 June 2017, doi: [10.1109/ACCESS.2017.2720187](https://doi.org/10.1109/ACCESS.2017.2720187)
- [45] Xiaoyun He, Jaideep Vaidya, Basit Shafiq, Nabil Adam, Vijay Atluri, "Preserving Privacy in Social Networks: A Structure-Aware Approach", *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, Milan, Italy, 15-18 September 2009, doi: [10.1109/WI-IAT.2009.108](https://doi.org/10.1109/WI-IAT.2009.108)
- [46] KAH MENG CHONG AND AMIZAH MALIP, "Trace Me If You Can: An Unlinkability Approach for Privacy-Preserving in Social Networks", *IEEE Access*, vol. 9, pp. 143950 - 143968, 17 March 2021, doi: [10.1109/ACCESS.2021.3066176](https://doi.org/10.1109/ACCESS.2021.3066176)
- [47] Tsan-sheng Hsu, Churn-Jung Liao, Da-Wei Wang, "A logical framework for privacy-preserving social network publication", *Journal of Applied Logic*, vol. 12, pp. 151-174, June 2014, doi: <https://doi.org/10.1016/j.jal.2013.12.001>
- [48] Abdullah Al Hasib, "Threats of Online Social Networks", *International Journal of Computer Science and Network Security*, vol. 9, November 2009.
- [49] Gianluca Lax, Antonia Russo, Lara Saidia Fasci, "A Blockchain-based approach for matching desired and real privacy settings of social network users", *Information Sciences*, vol. 557, pp. 220-235, May 2021, doi: <https://doi.org/10.1016/j.ins.2021.01.004>
- [50] Largo Bruno Pontecorvo, "When Blockchain meets Online Social Networks", *Pervasive and Mobile Computing*, vol. 62, pp. 101-131, February 2020, doi: <https://doi.org/10.1016/j.pmcj.2020.101131>
- [51] Le Jiang, Xinglin Zhang, "A Blockchain-Based Decentralized Online Social Network", *IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS*, VOL. 6, pp. 1454-1466, DECEMBER 2019, doi: [10.1109/TCSS.2019.2941650](https://doi.org/10.1109/TCSS.2019.2941650)
- [52] <https://www.guru99.com>

Providing an Intrusion Detection System in the Industrial Internet of Things Using the Gray Wolf Algorithm

Sajjad Alimohammadi^{1*}, Mohammad Fathi²

¹ Faculty Member of Electrical Engineering, University of Kurdistan, Sanandaj, Iran

² Faculty Member of Electrical Engineering, University of Kurdistan, Sanandaj, Iran

Received: 19 April 2024, Revised: 26 June 2024, Accepted: 07 October 2024

Paper type: Research

Abstract

Security is a main goal in the design of the Industrial Internet of Things network. Due to the ever-increasing developments in the Internet of Things, it is necessary to use new methods in detecting active network attacks. In this article, an intrusion detection system for industrial Internet of Things is proposed. This system uses a combination of gray wolf meta-heuristic (GWO) and decision tree (DT), nearest neighbor (KNN) and artificial neural network (ANN) classification algorithms. First, the data is pre-processed and then normalized, in the next step, data feature extraction is performed using the gray wolf algorithm to extract its independent and effective features. Then it is trained and finally evaluated using classification algorithms. The obtained results show that the use of the combined GWO-ANN algorithm with 93.22% accuracy has a better performance in detecting attacks. Also, the ANN algorithm is more accurate than the DT and KNN algorithms in combination with the GWO algorithm.

Keywords: Intrusion Detection System, Gray Wolf Algorithm, Industrial Internet of Things.

* Corresponding Author's email: sajad.alimohamadi@uok.ac.ir

ارائه سیستم تشخیص نفوذ در اینترنت اشیاء صنعتی با استفاده از الگوریتم گرگ خاکستری

سجاد علی محمدی^{۱*}، محمد فتحی^۲

^۱ دانشکده مهندسی برق، دانشگاه کردستان، سنندج، ایران

^۲ دانشکده مهندسی برق، دانشگاه کردستان، سنندج، ایران

تاریخ دریافت: ۱۴۰۳/۰۱/۳۱ تاریخ بازبینی: ۱۴۰۳/۰۴/۰۶ تاریخ پذیرش: ۱۴۰۳/۰۷/۱۶

نوع مقاله: پژوهشی

چکیده

امنیت یک هدف اصلی در طراحی شبکه اینترنت اشیاء صنعتی است. با توجه به پیشرفت‌های روزافزون در اینترنت اشیاء لازم است از روش‌های جدید در تشخیص حملات فعال شبکه استفاده شود. در این مقاله یک سیستم تشخیص نفوذ برای اینترنت اشیاء صنعتی پیشنهاد شده است. این سیستم از ترکیب الگوریتم‌های فراابتکاری گرگ خاکستری (GWO) و الگوریتم‌های طبقه‌بندی درخت تصمیم (DT)، نزدیک‌ترین همسایه (KNN) و شبکه عصبی مصنوعی (ANN) استفاده می‌کند. ابتدا داده‌ها پیش‌پردازش و سپس نرمال‌سازی شده، در مرحله بعد استخراج ویژگی داده‌ها با استفاده از الگوریتم گرگ خاکستری برای استخراج ویژگی‌های مستقل و مؤثر آن انجام می‌شود. سپس با استفاده از الگوریتم‌های طبقه‌بندی، آموزش و در نهایت ارزیابی می‌شود. نتایج به‌دست‌آمده نشان می‌دهد که استفاده از الگوریتم ترکیبی GWO-ANN با دقت ۹۳/۲۲ درصد در میزان تشخیص حملات عملکرد بهتری دارد. همچنین الگوریتم ANN نسبت به الگوریتم‌های DT و KNN در تلفیق با الگوریتم GWO دارای دقت بالاتری است.

کلیدواژگان: سیستم تشخیص نفوذ، الگوریتم گرگ خاکستری، اینترنت اشیاء صنعتی.

* رایانامه نویسنده مسؤول: sajad.alimohamadi@uok.ac.ir

۱- مقدمه

حال افزایش است. کنترل‌کننده‌های منطقی قابل‌برنامه‌ریزی با دستگاه‌های فیزیکی سایبری پیشرفته‌تر جایگزین شده‌اند که سبب شده تا دستگاه‌های نهفته^۳ به صورت آزادانه فرآیندهای فیزیکی را برنامه‌ریزی و کنترل کنند. در دستگاه‌های کنترل صنعتی، مفهوم امنیت به طور عمومی تقریباً همان معنای ایمنی را دارد، یعنی حفاظت از انسان، محیط‌زیست و ماشین‌ها در برابر پیامدهای ناشی از خرابی دستگاه‌ها است [۳].

امروزه برقراری امنیت سایبری در زیرساخت‌های اینترنت اشیا که در صنعت مورد استفاده قرار می‌گیرند دارای اهمیت ویژه‌ای است. برنامه‌هایی که در صنعت بر پایه اینترنت اشیا استفاده می‌شوند در برابر هرگونه حمله، اختلال و سرقت اطلاعات، آسیب‌پذیر می‌باشند. این برنامه‌ها ممکن است بسیار حساس و حفظ حریم خصوصی و امنیت آن‌ها دارای اهمیت بالای باشد [۴].

از این رو در این مقاله تلاش می‌گردد که به مسئله امنیت در IIoT که از اهمیت بالایی برخوردار است، پرداخته شود و با استفاده از روش‌های مبتنی بر یادگیری ماشین به بررسی این مسئله پرداخته شود. در این مقاله اثر ترکیب الگوریتم فرا ابتکاری گرگ خاکستری با مدل‌های یادگیری ماشین مرسوم از جمله درخت تصمیم، شبکه عصبی و k- نزدیک‌ترین همسایه برای سیستم تشخیص نفوذ در اینترنت اشیا مورد ارزیابی قرار می‌گیرد.

دستگاه‌های کنترل صنعتی بخشی جدایی‌ناپذیر از زیرساخت‌های حیاتی هستند و مدت‌زمان طولانی در نظارت بر ماشین‌آلات صنعتی و فرآیندها مورد استفاده قرار گرفته‌اند. این دستگاه‌ها بر دستگاه‌های صنعتی موجود در شرکت نظارت کرده، با آن‌ها تعامل داشته و به صورت بلادرنگ نسبت به جمع‌آوری و تجزیه و تحلیل داده‌ها اقدام کرده و همچنین ثبت تمام رویدادهایی که در دستگاه‌های صنعتی اتفاق می‌افتد را انجام می‌دهند. سیستم کنترل نظارت و اکتساب داده^۴ (SCADA) بزرگ‌ترین زیرمجموعه از دستگاه‌های کنترل صنعتی است. این سیستم یک رابط کاربری گرافیکی را از طریق رابط انسانی-دستگاه^۵ HMI فراهم می‌کند. HMI مشاهده وضعیت سیستم، برقرار ارتباط با دستگاه‌های IIoT و دریافت هشدار که نشان‌دهنده رفتارهای غیرطبیعی است را برای کاربران امکان‌پذیر می‌سازد. طرح کلی دستگاه‌های SCADA به همراه یک سیستم تشخیص نفوذ در شکل ۱ نشان داده شده است [۵].

با گسترش روزافزون شبکه‌های مبتنی بر اینترنت اشیا چالش‌های این حوزه روزبه‌روز در حال افزایش است. اینترنت اشیا در زمینه‌های مختلف مانند صنعت، پزشکی و غیره کاربرد دارد. حفظ امنیت یکی از چالش‌های مهم شبکه‌های اینترنت اشیا است. شبکه‌های مبتنی بر اینترنت اشیا همواره با محدودیت‌هایی از قبیل تأمین انرژی، قدرت پردازش کم و غیره روبرو می‌باشند. از این رو، تأمین امنیت شبکه‌های اینترنت اشیا از اهمیت بالایی برخوردار است زیرا مهاجم بانفوذ به شبکه قادر به انجام اعمال مخربی مانند سرقت اطلاعات یا اختلال در شبکه است. مهاجم با کشف یک آسیب‌پذیری و بهره‌برداری از آن به اهداف مخرب خود دست می‌یابد. با توجه به حساسیت شبکه اینترنت اشیا صنعتی (IIoT) در محیط‌های مختلف مانند صنایع پتروشیمی، خودروسازی و غیره نیاز به ایجاد یک شبکه امن لازم و ضروری است.

اینترنت اشیا امکان جمع‌آوری، انتقال و پردازش داده‌ها را در یک شبکه صنعتی فراهم می‌کند. از طرفی، پیچیدگی تجهیزات یک شبکه صنعتی و نیاز به هماهنگی تمام اجزاء شبکه مبتنی بر اینترنت اشیا دشوار است. از این رو، وجود آسیب‌پذیری در یک شبکه IIoT اجتناب‌ناپذیر است. شناسایی سریع حملات فعال خواهد توانست از دسترسی مهاجم به منابع و اختلال در عملکرد شبکه جلوگیری کند [۱].

با گسترش فناوری‌های مختلف و ظهور مفاهیم جدید مانند کلان داده، محاسبات ابری، سیستم فیزیکی سایبری^۲ و غیره، صنایع به سطح بالایی از توسعه ارتقا یافته‌اند. در نتیجه، با استفاده از IIoT دستگاه‌های صنعتی هوشمند به وجود آمدند که به دنبال تحقق تولید هوشمند هستند. IIoT ارتباط انواع مختلف تجهیزات صنعتی موجود در یک محیط صنعتی هوشمند را به صورت تعاملی با سایر دستگاه‌ها برقرار می‌کنند. در یک محیط تعاملی داده‌ها دیگر مستقل از یکدیگر نیستند. با ارتباط تجهیزات صنعتی به یکدیگر می‌توان دگرگونی در فرآیند تولید هوشمند ایجاد کرد. در حقیقت، روند فعلی صنعت به دنبال استفاده از اینترنت برای اتصال تجهیزات صنعتی و تحقق چهارمین انقلاب صنعتی است [۲]. در دهه اخیر، مهندسی تولید، هوشمندسازی و روش‌های محاسباتی به دنبال تکامل IIoT هستند. اجزای محاسباتی و ارتباطی ادغام شده در دستگاه‌های کنترل صنعتی و دستگاه‌های تولید در کارخانه‌ها، به طور پیوسته در

⁴ Supervisory Control and Data Acquisition

⁵ Human Machine Interface

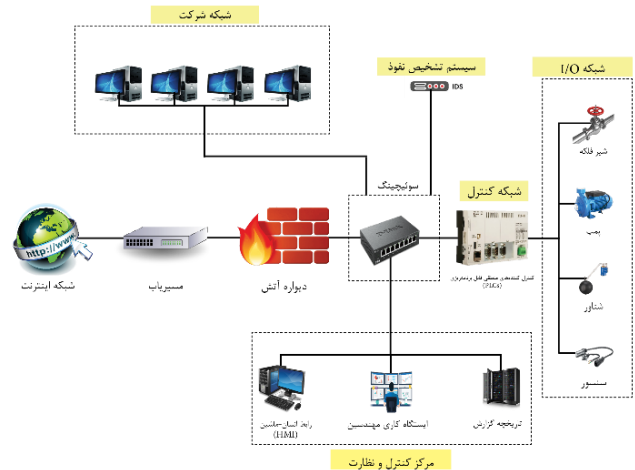
¹ Industrial Internet of Things

² Cyber Physical System

³ Embedded Devices

۲- مرور ادبیات

معماری امنیتی IIoT با ساختار امنیتی در شبکه‌های سازمانی متفاوت است. معماری IIoT دارای یک ساختار سلسله مراتبی از فناوری‌های مختلف شبکه و پروتکل‌های ارتباطی که دستگاه‌های سخت‌افزاری را با نرم‌افزار کنترلی به هم متصل می‌کند تشکیل شده است. مسائل امنیتی مطرح شده در IIoT بیشتر مربوط به طراحی پروتکل‌های ارتباطی ناامن هستند که با وجود چنین معماری، پیشگیری از حملات سایبری در شبکه IIoT را پیچیده می‌کند [۶]. در ادامه با بررسی تحقیقات پیشین روش‌های مختلف امن‌سازی شبکه‌های IIoT را بررسی خواهیم کرد.



شکل ۱. طرح کلی سیستم SCADA

آرزم و براتی [۷] در مقاله خود به ارائه روشی برای احراز هویت در اینترنت اشیاء مبتنی بر موقعیت گیرنده‌های Wi-Fi و فناوری زنجیره بلوکی پرداختند. روش پیشنهادی، شامل دو مرحله است. در مرحله اول، هر دستگاه اثبات‌کننده، مقداردهی اولیه می‌شود که برای اجرای پروتکل تصدیق و احراز هویت به آن نیاز است. در مرحله دوم، برای احراز هویت و حفظ حریم خصوصی، پروتکل زنجیره بلوکی استفاده می‌شود. با توجه به نتایج حاصل شده با تزریق خطاهای ناشی از تهدیدات امنیتی، نشان داد که تهدیدات امنیتی گذرا باعث بروز پنج خطا و تزریق تهدیدات امنیتی پایدار باعث بروز ۱۰۰ خطا در سیستم می‌شوند. بدون اعمال تمهیدات امنیتی، ۱۰۵ نمونه از داده‌های اینترنت اشیاء مورد تهدید قرار گرفت. با توجه به تعداد کل داده‌های قرائت شده در پنج ثانیه شبیه‌سازی که برابر با ۵۰۰ نمونه است، میزان داده‌های مورد تهدید قرار گرفته برابر با ۲۱ درصد است.

سید ترابی و پهلوان [۸]، در مقاله خود به رمزگذاری در احراز هویت دستگاه‌های اینترنت اشیاء پرداختند. در این مقاله روشی جهت حفظ امنیت اطلاعات مبتنی بر احراز هویت به کمک رمزنگاری و امضای دیجیتال و همچنین شکست فایل به بخش‌های کوچک‌تر ارائه شده است. آزمایش‌ها بر اساس تعداد مختلف بسته‌های انتقالی برای روش پیشنهادی نشان می‌دهد که این روش به دلیل استفاده از رمزنگاری تصادفی باعث شده است که معیار عملکرد خوبی به ازای بسته‌های مختلف از نظر معیارهای نظیر مصرف انرژی، دقت و غیره داشته باشد. همچنین بر اساس روش پیشنهادی ضمن احراز هویت کاربران، سرویس محرمانگی و دقت اطلاعات نیز به نحوی مناسب در روش پیشنهادی مورد توجه قرار گرفته شده است.

عیسی‌لو و سلیمانی [۹] در مقاله خود به ارائه روشی کم‌بار برای

همان‌گونه که در شکل ۱ نشان داده شده است ساختار SCADA شامل چهار زیرسیستم مختلف است. شبکه I/O، مرکز کنترل و نظارت، شبکه کنترل و شبکه شرکت. شبکه I/O شامل دستگاه‌های IIoT مستقر شده متشکل از حس‌گرها و محرک‌ها در فرآیند صنعتی است. کنترل نظارتی زیرسیستم مسئول اصلی امنیت، کنترل و نظارت بر دستگاه‌های IIoT است. شبکه کنترل شامل کنترل‌کننده‌های منطقی قابل برنامه‌ریزی^۱ (PLCs) است که به طور مستقیم دریافت داده و مدیریت فرآیندهای فیزیکی را انجام می‌دهد. از آنجاکه حس‌گرها و محرک‌ها نمی‌توانند به طور مستقیم ارتباط برقرار کنند، PLC ها برای جمع‌آوری داده‌ها و ارسال دستورات به محرک‌ها استفاده می‌شوند. در نهایت، شبکه شرکت شامل سرورها، کامپیوترها و دیگر کاربران متصل به شبکه برای سایر خدمات عمومی مانند انتقال فایل، میزبانی وبسایت، ایمیل، برنامه‌ریزی منابع و غیره است [۵]. در این مقاله، یک سیستم تشخیص نفوذ^۲ (IDS) مبتنی بر یادگیری ماشین که با بررسی بسته‌ها و ارتباطات ایجاد شده به دنبال کشف حملات فعال در شبکه‌های IIoT است ارائه می‌شود. توضیح اینکه سیستم تشخیص نفوذ معمولاً در لبه شبکه داخلی بعد از دیواره آتش و در اتصال با سویچ هسته شبکه است.

این مقاله در ادامه به صورت زیر سازمان‌دهی می‌شود. در بخش دوم تحقیقات مرتبط با امنیت در شبکه IIoT مورد بررسی قرار می‌گیرد. در بخش سوم طرح پیشنهادی یک IDS در شبکه IIoT ارائه می‌شود. در بخش چهار ارزیابی از روش پیشنهادی ارائه و در بخش پنجم جمع‌بندی و پیشنهادهایی برای تحقیقات آینده آورده می‌شود.

² Intrusion Detection System

¹ Programmable Logic Controllers

گردیده است. نتایج نشان می‌دهد که رویکرد اتخاذ شده قادر به شناسایی حملات با دقت رضایت بخشی است.

زولانواری و همکارانش [۱۴] در مقاله خود بر روی استفاده از یادگیری ماشین (ML) برای افزایش امنیت دستگاه‌های صنعتی اینترنت اشیا (IIoT) تمرکز دارد. به دلیل پیامدهای مخرب احتمالی حملات سایبری، بر نیاز حیاتی به ایمن‌سازی دستگاه‌های IIoT تأکید می‌کند. نویسندگان بر اهمیت ML و تجزیه و تحلیل داده‌های بزرگ در تجزیه و تحلیل و ایمن‌سازی فناوری IIoT تأکید می‌کنند و اینکه چگونه این تکنیک‌ها می‌توانند امنیت سیستم‌های IIoT را بهبود بخشند. این مطالعه شامل مروری بر پروتکل‌های رایج IIoT و آسیب‌پذیری‌های مرتبط با آن‌ها، ارزیابی آسیب‌پذیری سایبری و بررسی ادبیات راه‌حل‌های تشخیص نفوذ موجود با استفاده از مدل‌های ML است. علاوه بر این، نویسندگان یک مطالعه موردی شامل اجرای یک بستر آزمایشی در دنیای واقعی برای انجام حملات سایبری و طراحی یک سیستم تشخیص نفوذ^۳ (IDS) با استفاده از تشخیص ناهنجاری مبتنی بر ML ارائه می‌کنند. این مطالعه همچنین عملکرد IDS را از طریق معیارهای نماینده برای ارزیابی اثربخشی روش‌ها ارزیابی می‌کند. پس‌زمینه مقاله مروری جامع از آسیب‌پذیری‌ها در سیستم‌های IIoT، استفاده از ML برای تشخیص نفوذ و توسعه یک بستر آزمایشی در دنیای واقعی برای انجام حملات سایبری و طراحی IDS ارائه می‌کند. هدف این مطالعه پرداختن به چالش‌های امنیتی در سیستم‌های IIoT و نشان دادن پتانسیل راه‌حل‌های مبتنی بر ML در افزایش امنیت این سیستم‌ها است.

یعقوب کایوده سعید و سانجی میسرادر [۱۵] در مقاله خود تحت عنوان «مدل‌های یادگیری گروهی مبتنی بر بهینه‌ساز گرگ خاکستری برای تشخیص نفوذ در اینترنت اشیا» از تکنیک‌های بهره اطلاعاتی^۴ IG و PCA^۵ برای انتخاب و استخراج مجموعه‌ای از ویژگی‌ها بهره می‌برند و سپس از روش گرگ خاکستری برای بهینه‌سازی پارامترهای یک مدل ماشین ترکیبی از چهار ماشین دیگر شامل درخت تصمیم، جنگل تصادفی، شبکه عصبی و k نزدیک‌ترین همسایه استفاده می‌کنند.

مائولی و همکاران [۱۶] در این مقاله خود، مسائل امنیتی اینترنت اشیا صنعتی را از سه جنبه بررسی کرده‌اند شامل: (۱) تهدیدات امنیتی و مکانیسم‌های حمله آن‌ها برای نشان دادن آسیب‌پذیری اینترنت صنعتی اشیا. (۲) روش‌های تشخیص نفوذ از دیدگاه

احراز هویت اشیا در اینترنت اشیا پرداختند. در این مقاله به بررسی مدل احراز هویت EAP پرداخته و با سبک‌بار نمودن زیرشاخه‌ای از آن یعنی EAP-PSK روشی کم‌بار برای احراز هویت اشیا در اینترنت اشیا را ارائه کرده‌اند. روش ارائه‌شده با ابزار شبیه‌سازی Cooja مورد آزمایش قرار گرفته و در مقایسه با روش پایه در شاخص‌های مانند مجموع تعداد پیام‌های احراز هویت، زمان احراز هویت، نرخ موفقیت احراز هویت و نرخ مصرف انرژی بهبود یافته است.

IDS یک چهارچوب دفاعی امیدوارکننده در امنیت سایبری است که برای تشخیص حملات در شبکه استفاده می‌شود. در سال‌های اخیر این مکانیسم در شبکه‌های IIoT نیز مورد توجه قرار گرفته است. در ادامه سیستم‌های تشخیص نفوذ ارائه‌شده برای شبکه IIoT که مبتنی بر الگوریتم‌های یادگیری ماشین هستند مورد بررسی قرار می‌گیرند. هاب و همکاران، یادگیری زیر فضای تصادفی و الگوریتم نزدیک‌ترین همسایه^۱ (KNN) را برای بهبود دقت تشخیص KNN ترکیب کرده‌اند. نتایج به دست آمده بهبود عملکرد KNN را نشان می‌دهد [۱۰].

ردی و همکاران، یک مطالعه تجربی با استفاده از چندین روش یادگیری ماشین برای تشخیص ناهنجاری و طبقه‌بندی حمله ارائه کردند. آن‌ها یک چهارچوب یادگیری با استفاده از XGBoost ارائه نمودند. نتایج نشان می‌دهد که XGBoost یک رویکرد امیدوارکننده برای تشخیص نفوذ در طبقه‌بندی حملات است [۱۱].

زانگ و همکاران [۱۲] با استفاده از روش‌های دسته‌بندی KNN و RF^۲ سیستم تشخیص نفوذ و قابل اعتماد برای تشخیص حملات مردمیانی و منع سرویس توزیع شده را در مقاله خود ارائه نموده‌اند.

ایگنر و همکاران [۱۳] در مقاله خود به شناسایی حملات مردمیانی در شبکه‌های کنترل صنعتی پرداختند. روش پیشنهادی از تشخیص ناهنجاری با توسعه مدلی از رفتار عادی شبکه سیستم کنترل صنعتی استفاده می‌کند. یک سیستم صنعتی ساده، متشکل از حس‌گرها و محرک‌ها، با کنترل‌کننده‌هایی که به‌طور گسترده در صنعت استفاده می‌شوند، راه‌اندازی شده است. سپس از یک رویکرد یادگیری ماشین بر اساس الگوریتم k- نزدیک‌ترین همسایه با واگرایی برگمن برای تعریف مدلی از رفتار عادی (معتبر) استفاده شده است. پس از آن حملات مردمیانی علیه سیستم انجام شده و رفتار آن در حین حمله با مدل رفتاری معتبر مقایسه

⁴ Information Gain (IG)

⁵ Principal Component Analysis (PCA)

¹ K Nearest Neighbor

² Random Forest

³ Intrusion Detection Systems

۳-۱- مجموعه داده

مجموعه داده‌های متفاوت مانند UNSW-NSL-KDD، IoTID20، NB15، CICIDS2017 و Kddcup99 برای بررسی امنیت IIoT ارائه شده است. در این مقاله با توجه به تعداد ویژگی‌ها، تعداد نمونه‌های زیاد و گستردگی استفاده در تحقیقات پیشین از مجموعه داده KDDcup99 استفاده می‌شود [۲۱]. این مجموعه داده دارای ۴۱ ویژگی برابر جدول ۱ است که این ویژگی‌ها به ویژگی‌های اساسی، ویژگی محتوایی و ویژگی ترافیک تقسیم‌بندی می‌شوند.

جدول ۱. دسته‌بندی ویژگی‌ها

شماره	نام ویژگی	تعریف ویژگی	طبقه اختصاص یافته
ویژگی‌های پدای	۱	duration	طول اتصال
	۲	protocol_type	نوع پروتکل (TCP, UDP, ...)
	۳	service	سرویس مقصد (ftp, telnet, ...)
	۴	Flag	وضعیت اتصال
	۵	src_bytes	شماره B از مبدأ تا مقصد
	۶	dst_bytes	شماره B از مقصد تا مبدأ
	۷	land	اگر آدرس مبدأ و مقصد همان باشد /land=1، اگر نه، ۰ است
	۸	wrong_fragment	تعداد قطعات اشتباه
	۹	urgent	تعداد بسته‌های فوری
ویژگی‌های محتوایی	۱۰	hot	تعداد نشانگرهای داغ
	۱۱	num_failed_logins	تعداد تلاش‌های ناموفق برای ورود
	۱۲	logged_in	اگر وارد شوید=1/اگر ورود ناموفق بود ۰
	۱۳	num_compromised	تعداد کشورهای در معرض خطر
	۱۴	root_shell	اگر یک مفسر دستور با حساب ریشه در حال اجرا=1/shell، اگر نه، ۰ باشد
	۱۵	su_attempted	اگر دستور su تلاش شد=1/su tried، اگر نه، سپس ۰ (ورود موقت به سیستم با سایر اطلاعات کاربری کاربر)
	۱۶	num_rot	تعداد دسترسی‌های ریشه
	۱۷	num_file_creations	تعداد عملیاتی که فایل‌های جدید ایجاد می‌کند
	۱۸	num_shells	تعداد مفسران فرمان فعال
	۱۹	num_access_files	تعداد عملیات ایجاد فایل
	۲۰	num_outbound_cmds	تعداد دستورات خروجی در یک جلسه ftp
	۲۱	is_host_login	اگر لاگین در لیست ورود میزبان باشد، login=1/اگر نه، ۰ است
	۲۲	is_guest_login	اگر مهمان وارد سیستم شده باشد،

شناسایی حمله و (۳) برخی از استراتژی‌های دفاعی به طور جامع خلاصه شده است. در پایان این مقاله چندین نکته پایانی و دستورالعمل برای تحقیقات در آینده ارائه شده است.

فتانی و همکاران [۱۷] در این مقاله خود یک مدل IDS جدید مبتنی بر ترکیب روش‌های یادگیری عمیق و بهینه‌سازی را پیشنهاد کرده‌اند. در این مقاله ابتدا یک روش استخراج ویژگی مبتنی بر CNN توسعه داده شده است. سپس، یک روش انتخاب ویژگی جدید بر اساس نسخه اصلاح شده بهینه‌سازی رشد^۱ (GO) به نام MGO استفاده می‌شود. آن‌ها از الگوریتم بهینه‌سازی نهنگ^۲ (WOA) برای تقویت روند جستجوی GO استفاده کرده‌اند. ارزیابی و مقایسه‌های گسترده‌ای برای ارزیابی کیفیت روش پیشنهادی با استفاده از مجموعه داده‌های عمومی محیط‌های ابری و اینترنت اشیا (IoT) انجام شده است. آن‌ها به این نتیجه رسیده‌اند که MGO در تمام مقایسه‌های تجربی بهتر از چندین روش قبلی ذکر شده در مقاله عمل کرده است.

سلطانی و همکاران [۱۸] در نیز مقاله خود تحت عنوان «تشخیص نفوذ قوی برای ارتباطات شبکه در اینترنت اشیا: یک رویکرد یادگیری ماشین ترکیبی» یک رویکرد یادگیری ماشین ترکیبی (با استفاده از k-نزدیک‌ترین همسایگان و جنگل‌های تصادفی به عنوان طبقه‌بندی‌کننده نظارت‌شده) برای افزایش دقت سیستم‌های تشخیص نفوذ و به حداقل رساندن خطر حملات احتمالی ارائه کرده‌اند. همچنین در این مقاله، از الگوریتم‌های حذف به عقب و تحلیل تفکیک خطی برای کاهش ویژگی و کاهش هزینه‌های محاسباتی استفاده شده است. عملکرد مدل پیشنهادی در چارچوب برنامه‌نویسی پایتون، با استفاده از مجموعه داده‌های CICIDS 2017، NSL-KDD و TON-IoT ارزیابی شده.

۳- روش پیشنهادی

در این بخش سیستم تشخیص نفوذ پیشنهادی برای تشخیص حملات سایبری IIoT ارائه شده است که در این روش از ترکیب الگوریتم فرا ابتکاری گرگ خاکستری (GWO) [۱۹] و الگوریتم‌های یادگیری ماشین KNN، DT^۳ و ANN^۴ استفاده شده است [۲۰]. در ادامه ابتدا مجموعه داده استاندارد که برای تشخیص حملات در IIoT استفاده می‌شود معرفی خواهد شد سپس به معرفی روش پیشنهادی خواهیم پرداخت.

³ Decision Tree

⁴ Artificial Neural Network

¹ Growth Optimizer (GO)

² Whale Optimization Algorithm

جدول ۲. جزئیات مجموعه داده ۹۹ KDD Cup

دسته‌بندی حملات	نام حمله	تعداد رکورد
DoS (Denial of Service)	back, land, neptune, pod, smurf, teardrop	۵۴۲۹۴
R2L (Remote to Local)	ftp_write, guesspasswd, imap, multihop, phf, spy, warezlient, warezmaster	۵۳۹۴
U2R (User to Root)	buffer_overflow, loadmodule, perl, rootkit	۱۲۶
Probe	ipsweep, nmap, portsweep, satan	۱۳۸۵۹

۳-۱-۱- پیش‌پردازش داده‌ها

ویژگی‌های مجموعه داده در جدول ۱ گروه‌بندی شده‌اند. با توجه به اینکه ویژگی‌ها دارای مقادیر غیر عددی می‌باشند بایستی این ویژگی‌ها را به روش نگاشت همانی با اعداد حقیقی جایگزین کرد. همچنین گروه ویژگی‌های دارای مقدار صفر در تصمیم‌گیری بی‌تأثیر بوده‌اند و از مجموعه داده حذف می‌شوند. مجموعه داده، ممکن است برای یک نمونه خاص، یک ویژگی که لازم باشد را نداشته باشد و باعث ایجاد خطا در روند برنامه شود؛ بنابراین حداکثر مقدار ممکن آن ویژگی در جایگاه خالی قرار خواهد گرفت. مقادیر ویژگی‌های مجموعه داده در دامنه متفاوتی قرار دارند و احتمال بروز خطا در نتایج افزایش می‌یابد، بنابراین ۲۲ ویژگی باقیمانده با استفاده از رابطه (۱) در مقیاس [۰، ۱] نرمال‌سازی می‌شود.

$$\bar{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

در این رابطه X معرف داده‌های اولیه، \bar{X} مقدار ویژگی نرمالیزه شده، X_{max} و X_{min} به ترتیب کمترین و بیشترین مقدار ویژگی در مجموعه داده هستند. برچسب‌های داده‌ها به دو گروه حمله و نرمال تقسیم می‌شود. IDS حملات مختلف را در یک گروه کلی در نظر خواهد گرفت. پس از آماده‌سازی مجموعه داده‌ها به دو بخش آموزش شامل ۸۰ درصد نمونه‌ها و ارزیابی شامل ۲۰ درصد نمونه‌ها تقسیم می‌شوند. داده‌های بخش آموزش برای تولید مدل ماشین مورد استفاده قرار می‌گیرند. داده‌های بخش ارزیابی برای بررسی دقت مدل تولیدشده با داده‌های آموزش مورد استفاده قرار می‌گیرد.

۳-۲- استخراج ویژگی

بعد از آماده‌سازی مجموعه داده، استخراج ویژگی با استفاده از الگوریتم گرگ خاکستری^۱ (GWO) بر روی داده‌ها انجام می‌شود [۲۲]. هرکدام از دسته گرگ‌ها به‌عنوان اجزای یک راه‌حل در نظر گرفته می‌شوند. همه گرگ‌های مهاجم که در یک راه‌حل می‌باشند به‌عنوان یک واحد کلی در نظر گرفته می‌شوند که به سمت یک

ردیف	ویژگی	توضیح	تعداد رکورد
۲۳	Count	تعداد اتصالات به همان میزبان به‌عنوان اتصال فعلی در یک بازه زمانی معین	۱۳۸۵۹
۲۴	srv_count	تعداد اتصالات به همان سرورس با اتصال فعلی در یک بازه زمانی معین	۱۳۸۵۹
۲۵	serror_rate	% از اتصالات با خطاهای SYN	۱۳۸۵۹
۲۶	srv_error_rate	% از اتصالات با خطاهای SYN	۱۳۸۵۹
۲۷	rerror_rate	% از اتصالات با خطاهای REJ	۱۳۸۵۹
۲۸	srv_error_rate	% از اتصالات با خطاهای REJ	۱۳۸۵۹
۲۹	same_srv_rate	% از اتصالات به همان سرورس	۱۳۸۵۹
۳۰	diff_srv_rate	% از اتصالات به سرورس‌های مختلف	۱۳۸۵۹
۳۱	srv_diff_host_rate	% از اتصالات به هاست‌های مختلف	۱۳۸۵۹
۳۲	dst_host_count	تعداد اتصالات به همان مقصد	۱۳۸۵۹
۳۳	dst_host_srv_count	تعداد اتصالات به یک مقصد که از همان سرورس استفاده می‌کنند	۱۳۸۵۹
۳۴	dst_host_same_srv_rate	% از اتصالات به یک مقصد که از یک سرورس استفاده می‌کنند	۱۳۸۵۹
۳۵	dst_host_diff_srv_rate	% از اتصالات به هاست‌های مختلف در یک سیستم	۱۳۸۵۹
۳۶	dst_host_same_src_port_rate	% از اتصالات به یک سیستم با پورت منبع یکسان	۱۳۸۵۹
۳۷	dst_host_srv_diff_host_rate	% از اتصالات به یک سرورس که از میزبان‌های مختلف می‌آید	۱۳۸۵۹
۳۸	dst_host_serror_rate	% از اتصالات به یک میزبان با خطای S0	۱۳۸۵۹
۳۹	dst_host_srv_serror_rate	% از اتصالات به هاست و سرورس مشخص شده با خطای S0	۱۳۸۵۹
۴۰	dst_host_rerror_rate	% از اتصالات به هاست و سرورس مشخص شده با خطای S0	۱۳۸۵۹
۴۱	dst_host_srv_rerror_rate	% از اتصالات به هاست و سرورس مشخص شده با خطای S0	۱۳۸۵۹

ویژگی‌های اساسی مانند Protocol_type، Duration و Service و... شامل خصوصیات می‌باشند که از سرآیند یک اتصال TCP/IP استخراج می‌گردند. ویژگی‌های محتوایی مانند Num_compromised، Logged_in، Num_failed_logins و ... امکان تشخیص رفتارهای مشکوک در شبکه را فراهم می‌نمایند. ویژگی‌های ترافیکی که خود به دو ویژگی، میزبان یکسان و سرورس یکسان تقسیم می‌شوند نیز بر مبنای تجزیه و تحلیل تعداد ارتباط برقرار شده و بر اساس دو نمونه قبلی محاسبه می‌شوند. این مجموعه داده شامل ۲۲ حمله در چهار دسته DOS، U2R، R2L و Probe است. داده‌های ترافیک نرمال شبکه دارای برچسب Normal هستند. در جدول ۲ جزئیات مجموعه داده KDDcup99 بیان شده است.

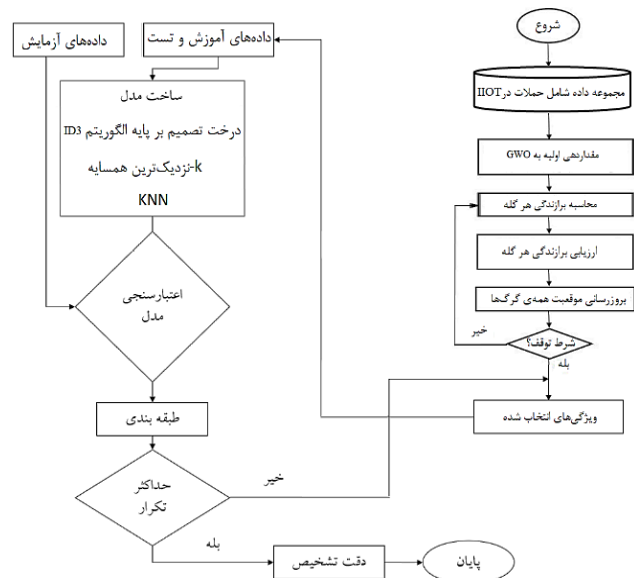
^۱ Gery Wolf Algorithm

با استفاده از الگوریتم گرگ خاکستری زیرمجموعه‌ای از ویژگی‌ها که به بهینه‌ترین مقدار منجر می‌شوند، انتخاب می‌گردد. تابع برازندگی برای انتخاب ویژگی از هر دسته گرگ طبق رابطه (۹) تعریف می‌شود. در رابطه (۹)، $|n|$ تعداد کل ویژگی‌ها و $|S|$ تعداد ویژگی‌های انتخاب شده است. پارامتر Accuracy درصد دقت و مقدار پارامترهای δ و ρ ثابت هستند و مقدار آن‌ها به ترتیب برابر با ۰,۷ و $1 - \delta$ است.

$$Fitness = \delta \cdot Accuracy + \rho \cdot \frac{|n| - |S|}{|n|} \quad (9)$$

استخراج ویژگی با ابعاد بالا امکان ایجاد راه‌حل‌های جدید را فراهم می‌کند اما در هنگام محاسبات چالش‌های ایجاد می‌شود. چالش‌های مانند اینکه در اکثر مواقع همه ویژگی‌ها برای یافتن جواب مناسب که در آن‌ها نهفته است مهم نیستند. راهکار مناسب برای انتخاب ویژگی، حذف آن قسمت از زیرمجموعه ویژگی‌های ورودی است که اطلاعات کمی در خود دارند. این امر باعث ایجاد ویژگی‌های باقیمانده که در افزایش کارایی طبقه‌بندی مؤثر هستند و باعث بالا رفتن دقت می‌شود می‌گردند.

هنگامی که با استفاده از الگوریتم گرگ خاکستری ویژگی‌های جدید از مجموعه داده استخراج شده، با استفاده از الگوریتم‌های یادگیری ماشین مانند DT، KNN و ANN مدل تصمیم‌گیری را آموزش می‌دهیم. در این مدل نتایج بررسی هر کدام از بسته‌ها با استفاده از مدل به دو گروه کلی ترافیک حمله و نرمال تقسیم‌بندی می‌شوند. نتایج به‌دست‌آمده در بخش بعدی مورد ارزیابی و مقایسه قرار می‌گیرد. فلوجارت اجرایی این روش در شکل ۲ ارائه شده است.



شکل ۲. فلوجارت پیشنهادی

هدف حرکت می‌نماید. نحوه شکار گرگ خاکستری به این شکل است که دسته گرگ‌ها به چهار نوع تقسیم می‌شوند:

الف- گرگ آلفا که رهبر گروه است و همیشه بهترین موقعیت را نسبت به هدف دارد.

ب- گرگ بتا که پایین‌تر از گرگ آلفا و بالاتر از سایرین است و تابع دستورات آلفا است و در واقع نقش مشاور را دارد.

ج- گرگ امگا که از دودسته بالا پایین‌تر و از سایرین بالاتر است و نقش محافظ در گله را بر عهده دارد.

د- گرگ‌های پیر و سالخورده دسته که گرگ دلتا نامیده می‌شوند. در این الگوریتم بهترین موقعیت برای راه‌حل مسئله به ترتیب مربوط به گرگ آلفا، گرگ بتا، گرگ دلتا و سایر راه‌حل‌ها نیز مربوط به گرگ امگا است. الگوریتم گرگ خاکستری بیشتر بر اساس موقعیت گرگ آلفا، بتا و دلتا جستجو انجام می‌دهد. در مدل‌سازی ریاضی برای محاسبه هدف از معادله (۲)، (۳) و (۴) استفاده می‌شود.

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (2)$$

$$\vec{D}(t) = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)|, t = 1, 2, \dots, t_{max} \quad (3)$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}, \vec{C} = 2 \cdot \vec{r}_2 \quad (4)$$

در این معادلات t تکرار فعلی را بیان می‌کند، $\vec{X}(t+1)$ موقعیت گرگ خاکستری در تکرار $t+1$ ، $\vec{X}_p(t)$ موقعیت هدف در تکرار t بردار \vec{D} فاصله هر کدام از گرگ‌های آلفا، بتا یا دلتا از گرگ امگا یا سایر شکارچیان گله است، بردارهای ضرایب \vec{A} و \vec{C} با توجه به تعریف، از \vec{r}_1 و \vec{r}_2 که بردارهای تصادفی بوده و به صورت تصادفی بین $[0, 1]$ هستند، به دست می‌آیند. همچنین a یک مؤلفه کاهشی است که به صورت خطی در بازه $[0, 2]$ کاهش پیدا می‌کند. با توجه به اینکه هیچ اطلاعاتی در مورد موقعیت هدف خود نداریم موقعیت هدف همان موقعیت گرگ آلفا در نظر گرفته می‌شود. به‌روزرسانی موقعیت گرگ خاکستری طبق روابط (۵)، (۶)، (۷) و (۸) انجام می‌شود [۱۵].

$$\vec{X}_1(t) = \vec{X}_\alpha(t) - \vec{A}_1 \cdot \vec{D}_\alpha(t), \vec{D}_\alpha(t) = |\vec{C}_1 \cdot \vec{X}_\alpha(t) - \vec{X}(t)| \quad (5)$$

$$\vec{X}_2(t) = \vec{X}_\beta(t) - \vec{A}_2 \cdot \vec{D}_\beta(t), \vec{D}_\beta(t) = |\vec{C}_2 \cdot \vec{X}_\beta(t) - \vec{X}(t)| \quad (6)$$

$$\vec{X}_3(t) = \vec{X}_\delta(t) - \vec{A}_3 \cdot \vec{D}_\delta(t), \vec{D}_\delta(t) = |\vec{C}_3 \cdot \vec{X}_\delta(t) - \vec{X}(t)| \quad (7)$$

$$\vec{X}(t+1) = \frac{(\vec{X}_1(t) + \vec{X}_2(t) + \vec{X}_3(t))}{3} \quad (8)$$

$\vec{X}_\alpha(t)$ ، $\vec{X}_\beta(t)$ و $\vec{X}_\delta(t)$ موقعیت گرگ‌های آلفا، بتا و دلتا در زمان t و X_1 ، X_2 ، X_3 موقعیت هر کدام از گرگ‌ها بر اساس این سه گرگ است [۱۸].

۴- ارزیابی نتایج

۹/۲۷ درصد نیازمند بهبود مدل پیش‌بینی است.

جدول ۴. نتایج GWO-DT

پارامتر	مقدار
TP	۲۸۳۲
TN	۱۱۶۰
FP	۲۰۳
FN	۲۰۵
Precision	۹۳/۳۱۱۳ درصد
Recall	۹۳/۲۴۹۹ درصد
Accuracy	۹۰/۷۲۷۳ درصد
F-Measure	۹۳/۲۸۰۶ درصد

مرحله بعد، به جای الگوریتم DT، از الگوریتم KNN جهت ارائه مدل استفاده می‌شود. نتایج این الگوریتم، در جدول ۵ ارائه شده است.

جدول ۵. نتایج GWO-KNN

پارامتر	مقدار
TP	۲۶۰۴
TN	۱۰۸۶
FP	۵۲۸
FN	۱۸۲
Precision	۸۳/۱۴۱۸ درصد
Recall	۹۳/۴۶۷۳ درصد
Accuracy	۸۳/۸۶۳۶ درصد
F-Measure	۸۸/۰۰۲۷ درصد

همان‌طور که در جدول ۵ مشاهده می‌شود، اعمال الگوریتم KNN به الگوریتم GWO در راستای کاهش ویژگی و شناسایی حمله به میزان دقت ۸۳/۸۶۳۶ درصد رسیده که در مقایسه با الگوریتم GWO-DT که ۹۰/۷۲۷۳ درصد است، در حدود ۷ درصد کمتر است. از طرفی میزان خطای به‌دست‌آمده توسط الگوریتم تلفیقی GWO-KNN برابر با ۰/۱۶۱۳ است.

در ادامه، از الگوریتم شبکه عصبی MLP استفاده می‌شود. در این مدل شبکه عصبی، الگوریتمی برای یادگیری نظارتی شبکه عصبی با استفاده از گرادینت کاهشی که در آن خطا نسبت به وزن‌های شبکه عصبی محاسبه می‌شود، تعیین می‌گردد. در جدول ۶ معماری شبکه عصبی با توجه به بیشترین ضریب همبستگی و بالاترین کارایی از لحاظ تعداد لایه و تعداد نرون، برای دستیابی به دقت بالاتر در آموزش بررسی شده است.

بر مبنای معماری‌های اشاره‌شده در جدول ۶، معماری مناسب برای شبکه عصبی مصنوعی به‌منظور تعیین عملکرد طبقه‌بندی حملات برابر با ۱۰ لایه و ۲۰ نرون برای هر لایه است؛ زیرا دارای بیشترین

در این بخش نتایج الگوریتم‌ها به‌صورت جداگانه ارائه می‌شود. برای آموزش مدل پیشنهادی از سخت‌افزار با مشخصات پردازنده Intel(R) Core(TM) U4500 - 5i و رم ۸,۰۰ GB استفاده شده است. با استفاده از ماتریس درهم‌ریختگی^۱ نتایج پیش‌بینی شده توسط مدل‌های DT، KNN و ANN توسط داده‌های استخراج‌شده از الگوریتم گرگ خاکستری بررسی می‌شود. ماتریس درهم‌ریختگی یک روش ارزیابی بازدهی مدل‌های طبقه‌بندی یادگیری ماشین است که به‌صورت جدول ۳ است و میزان انطباق مقادیر پیش‌بینی شده با مقادیر واقعی را نشان می‌دهد.

جدول ۳. ماتریس درهم‌ریختگی

	مقادیر واقعی	
	۰	۱
مقادیر پیش‌بینی شده	۰	FP
	۱	TP

برای به دست آوردن معیارهای Accuracy، Precision، Recall و F-measure با استفاده از ماتریس درهم‌ریختگی مطابق روابط (۱۰)، (۱۱)، (۱۲) و (۱۳) عمل می‌شوند.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (13)$$

Accuracy یا دقت یکی از معیارهای ارزیابی یک مدل است که نشان‌دهنده تعداد پیش‌بینی‌های صحیح است. Precision یا صحت نشان‌دهنده درصد حملاتی است که مدل به‌صورت صحیح نسبت به تمام حملات تشخیص داده است. Recall نسبت هشدارهای صحیح به تمام هشدارهای سیستم را نشان می‌دهد؛ و F-measure یک معیار هماهنگ است که به‌عنوان میانگین هندسی Precision و Recall مورد استفاده قرار می‌گیرد.

حال با توجه به معیارهای ارزیابی نتایج به‌دست‌آمده در هر کدام از الگوریتم‌ها را نشان خواهیم داد. برای الگوریتم GWO-DT نتایج مطابق جدول ۴ به‌دست‌آمده است. جدول ۴ نشان می‌دهد مدل GWO-DT میزان دقت ۹۰/۷۲۷۳ درصد و میزان خطای ۰/۰۹۲۷ با کاهش تعداد ویژگی‌های موجود در مجموعه داده توانسته وقوع یا عدم وقوع حمله را به شکل مطلوبی پیش‌بینی نماید؛ اما وجود خطا

^۱ Confusion matrix

به منظور مقایسه بهتر الگوریتم‌های بکار رفته تمامی نتایج از نظر دقت و زمان حل در جدول ۹ ارائه می‌شود.

همچنین در جدول ۱۰ مقایسه‌ای بین میزان خطای به دست آمده از تشخیص حملات سایبری توسط الگوریتم گرگ خاکستری در ترکیب با طبقه‌بندی‌های DT، KNN و ANN ارائه شده است.

بر اساس نتایج به دست آمده از نظر دقت در انتخاب ویژگی‌ها و میزان تشخیص حملات الگوریتم GWO-ANN دارای بهترین عملکرد است. پس از الگوریتم ANN، الگوریتم DT در مرتبه دوم قرار می‌گیرد و الگوریتم KNN از دقت کمتری نسبت به دو الگوریتم ANN و DT برخوردار است.

در میان سه الگوریتم بررسی شده GWO-ANN دارای بهترین دقت و کمترین خطا است از این رو نتایج به دست آمده الگوریتم را با کارهای مشابه که توسط عیسی، اورمان و بریفکانی [۲۳] که از ترکیب الگوریتم گرگ ماهی^۱ و DT در سیستم تشخیص نفوذ استفاده کرده‌اند مقایسه می‌کنیم. آن‌ها از روش‌های مختلف در راستای انتخاب تعداد ویژگی‌های مجموعه داده استفاده نموده‌اند که بهترین دقت برای ۱۰ ویژگی است. در جدول ۱۱ می‌توان نتایج مقایسه دقت و همچنین نرخ تشخیص الگوریتم GWO-ANN را با نتایج ارائه شده در مقاله اشاره شده مشاهده نمود.

جدول ۹. مقایسه دقت الگوریتم پیشنهادی با الگوریتم‌های استفاده شده

الگوریتم	دقت	تعداد ویژگی انتخاب شده
GWO - DT	۹۰/۷۲۷۳	۷
GWO - KNN	۸۳/۸۶۳۶	۹
GWO - ANN	۹۳/۲۲۷۳	۸

جدول ۱۰. میزان خطای به دست آمده از الگوریتم‌های مبتنی بر گرگ خاکستری

الگوریتم	خطا
GWO - DT	۰/۰۹۲۷
GWO - KNN	۰/۱۶۱۳
GWO - ANN	۰/۰۶۷۷

جدول ۱۱. مقایسه دقت و نرخ تشخیص

الگوریتم	دقت (%)	نرخ تشخیص (%)	تعداد ویژگی انتخاب شده
GWO-ANN	۹۳/۲۲۷۳	۹۵/۰۷۰۹	۸
CFA-DT	۹۲/۸۳۷	۹۲/۰۵۱	۱۰

کارایی و بالاترین ضریب همبستگی در بین سایر معماری‌های بررسی شده است. در جدول ۷ متغیرهای اولیه شبکه عصبی ارائه شده است. با ارائه مدل شبکه عصبی مصنوعی مناسب، مدل GWO-ANN ارائه می‌شود؛ که نتایج آن در جدول ۸ ارائه شده است.

همان‌طور که در جدول ۸ مشاهده می‌شود، اعمال الگوریتم ANN به الگوریتم گرگ خاکستری در راستای کاهش ویژگی و شناسایی حمله به میزان دقت ۹۳/۲۲۷۳ درصد رسیده که در مقایسه با الگوریتم GWO-KNN که ۹۰/۷۲۷۳ درصد است، حدود ۳ درصد افزایش داشته است.

جدول ۶. مقایسه معماری‌های مختلف شبکه عصبی MLP

ضریب همبستگی	کارایی	معماری	
		تعداد لایه	تعداد نرون
۰/۹۶۷۴	۰/۰۰۱۸	۱۰	۱۰
۰/۹۸۶۰	۰/۰۰۵۲	۲۰	۱۰
۰/۹۷۴۶	۰/۰۰۲۱	۱۰	۲۰
۰/۴۸۹۰	۰/۰۰۳۷	۱۵	۱۵
۰/۹۶۱۲	۰/۰۰۳۳	۵	۵
۰/۸۸۲۰	۰/۰۰۰۷	۸	۸
۰/۹۷۸۶	۰/۰۰۱۳	۵	۱۰
۰/۹۰۸۰	۰/۰۰۰۷	۱۰	۵
۰/۸۳۸۸	۰/۰۰۴۳	۸	۱۲
۰/۹۵۸۰	۰/۰۰۳۰	۱۵	۱۰
۰/۹۶۸۶	۰/۰۰۱۹	۲۵	۹

جدول ۷. پارامترهای شبکه عصبی

پارامتر	مقدار
Traning	Levenberg-Marquardt
Performance	Mean Squared Error
Epoch	۱۰۰۰

جدول ۸. نتایج GWO-ANN

پارامتر	مقدار
TP	۲۹۵۱
TN	۱۱۵۱
FP	۱۵۳
FN	۱۴۵
Precision	۹۵/۰۷۰۹ درصد
Recall	۹۵/۳۱۶۵ درصد
Accuracy	۹۳/۲۲۷۳ درصد
F-Measure	۹۵/۱۹۳۵ درصد

^۱ Cuttle Fish

[۹] ح. عیسی لو و ع. سلیمانی، "ارائه روشی کم‌بار برای احراز هویت اشیا در اینترنت اشیا"، کنگره ملی تحقیقات بنیادین در مهندسی کامپیوتر و فناوری اطلاعات، ص ۱۵، ۱۳۹۸.

- [10] Derhab, M. Guerroumi, A. Gumaeci, L. Maglaras, M. A. Ferrag, M. Mukherjee, et al., "Blockchain and random subspace learning-based ids for sdn-enabled industrial iot security", *Sensors*, vol. 19, no. 14, pp. 3119, 2019.
- [11] D. K. K. Reddy, H. Behera, J. Nayak, B. Naik, U. Ghosh and P. K. Sharma, "Exact greedy algorithm based split finding approach for intrusion detection in fog-enabled iot environment", *Journal of Information Security and Applications*, vol. 60, pp. 102866, 2021.
- [12] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines and J. Coble, "Multilayer data-driven cyber-attack detection system for industrial control systems based on network system and process data", *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4362-4369, 2019.
- [13] O. Eigner, P. Kreimel, P. Tavolato and P. Kreimel, "Detection of man-in-the-middle attacks on industrial control networks," in *In 2016 International Conference on Software Security and Assurance (ICSSA)*, 2016.
- [14] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan and R. Jain, "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things," *Internet of Things Journal*, vol. 6, no. 4, pp. 6822-6833, 2019.
- [15] Saheed, Yakub Kayode, and Sanjay Misra. "A voting gray wolf optimizer-based ensemble learning models for intrusion detection in the Internet of Things." *International Journal of Information Security* (2024): 1-25.
- [16] Wang, M.; Sun, Y.; Sun, H.; Zhang, B. Security Issues on Industrial Internet of Things: Overview and Challenges. *Computers* 2023, 12, 256. <https://doi.org/10.3390/computers12120256>.
- [17] Fatani, A.; Dahou, A.; Abd Elaziz, M.; Al-qaness, M.A.A.; Lu, S.; Alfidhli, S.A.; Alresheedi, S.S. Enhancing Intrusion Detection Systems for IoT and Cloud Environments Using Growth Optimizer Algorithm and Conventional Neural Networks. *Sensors* 2023, 23, 4430. <https://doi.org/10.3390/s23094430>.
- [18] Soltani, Nasim & Rahmani, Amir & Bohlouli, Mahdi & Hosseinzadeh, Mehdi. Robust intrusion detection for network communication on the Internet of Things: a hybrid machine learning approach. *Cluster Computing*, (2024). 1-17. <https://doi.org/10.1007/s10586-024-04483-7>.
- [19] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, Mar. 2014.
- [20] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2016.
- [21] M. Tavallae, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set." 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009.

[۲۲] ل. عجمی بختیاروند و ز. بهشتی، "روشی نوین برای خوشه‌بندی داده‌ها با استفاده از الگوریتم بهینه‌سازی چهارگرگ خاکستری"، *نشریه مهندسی برق و مهندسی کامپیوتر ایران*، ب- مهندسی کامپیوتر، ص ۲۷۴-۲۶۱، شماره ۴، سال ۱۹، ۱۴۰۰.

[23] S. Eesa, Z. Orman and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert systems with applications*," Elsevier, vol. 42, no. 5, pp. 2670-2679, 2015.

همان‌طور که در جدول فوق نیز مشاهده می‌شود، روش پیشنهادی GWO-ANN از نظر دقت و نرخ تشخیص همچنین انتخاب ویژگی کمتر عملکرد بهتری داشته است.

۵- جمع‌بندی

در شبکه‌های اینترنت اشیا صنعتی حملات فعال باعث ایجاد ضرر و زیان فراوانی می‌شوند از این‌رو نیاز به یک سیستم تشخیص نفوذ برای تشخیص حملات ضروری است. در این مقاله با استفاده از مجموعه داده KDD Cup ۹۹ اقدام به ارائه یک سیستم تشخیص نفوذ مبتنی بر یادگیری ماشین شد. این مدل با استفاده از الگوریتم گرگ خاکستری به‌عنوان یک واحد استخراج‌کننده ویژگی در مجموعه داده عمل می‌کند و ویژگی‌های مستقلی را از مجموعه داده استخراج و به ورودی الگوریتم‌های یادگیری ماشین می‌دهد. الگوریتم‌های که مورد بررسی قرار گرفته‌اند شامل KNN، DT و ANN است که از این میان الگوریتم ANN دارای عملکرد بهتری است. همچنین روش پیشنهادی GWO-ANN در مقایسه با روش CFA-ANN حدود ۱/۱۷۶۳ درصد بهبود را نشان می‌دهد.

مراجع

- [1] X. Fei and G. Tian, "Fault Identification and Analysis of Communication Network Based on Deep Learning," 2022.
- [2] J. Wan, J. Li, M. Imran and D. Li, "A blockchain-based solution for enhancing security and privacy in smart factory," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3652-3660, 2019.
- [3] R. Sadeghi, C. Wachsmann and M. Waidner, "Security and privacy challenges in industrial internet of thing," *In 2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, pp. 1-6, June 2015.
- [4] J. Sengupta, S. Ruj and S. D. Bit, "A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT," *Journal of Network and Computer Applications*, p. 149, 2020.
- [5] M. Zolanvari, M. A. Teixeira, L. Gupta and K. Khan, "Machine learning-based network vulnerability analysis of industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6822-6834, 2019.
- [6] Deshpande, P. Pitale and S. Sanap, "Industrial automation using Internet of Things (IOT)," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 5, no. 2, pp. 266-269, 2016.

[۷] ر. آرزوم و ع. براتی، "ارائه روشی برای احراز هویت در اینترنت اشیا مبتنی بر موقعیت گیرنده‌های Wi-Fi و فناوری زنجیره بلوکی"، پنجمین اجلاس ملی محاسبات توزیعی و پردازش داده‌های بزرگ، ص ۱۸، ۱۳۹۸.

[۸] ع. سید ترابی و ر. پهلوان، "رمزگذاری در احراز هویت دستگاه‌های اینترنت اشیا"، چهارمین اجلاس ملی ایده‌های نوین در فنی و مهندسی، ص ۲۴، ۱۳۹۸.

Trust Based Link Prediction Using Fuzzy Computational Model in Social Networks

Fatemeh Hoseinkhani^{1*}, Ali Harounabadi², Saeed Setayeshi³

¹Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

²Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

³Department of Medical Radiation Engineering, Amirkabir University of Technology, Tehran, Iran.

Received: 05 July 2023, Revised: 16 May 2024, Accepted: 12 August 2024

Paper type: Research

Abstract

Link prediction is important to check links between nodes in social networks. The modeling of social networks leads to the emergence of signed, directed, and weighted social networks. The relationships of users in social networks are characterized by subjective, asymmetric, and ambiguous aspects related to this domain, both terms of trust and distrust are challenging. To solve the problem of sparsity in networks and overcome ambiguity in relationships, a trust-distrust method based on fuzzy computational is proposed to calculate the strength of links. The purpose of the proposed link prediction is to solve problems of sparsity in signed social networks by combining descriptive features of users with the direct influence of top nodes and the indirect influence of common nodes on rating prediction. Trust is determined by a Mamdani fuzzy system based on mirroring of similarity fuzzy features, overall trust, and overall distrust. The evaluation of the proposed method was done with the accuracy measure on datasets of Epinions and Slashdot. The accuracy of the proposed method in Epinions and Slashdot datasets is 0.991 and 0.998, respectively. The obtained results show that the proposed method works well for problems of data sparsity in signed social networks and shows the effectiveness of the proposed model.

Keywords: Link Prediction, Signed Social Networks, Trust, Distrust, Fuzzy Computational.

* Corresponding Author's email: a.harounabadi@iauctb.ac.ir

پیش‌بینی پیوند مبتنی بر اعتماد با استفاده از مدل محاسبات فازی در شبکه‌های اجتماعی علامت‌دار

فاطمه حسین خانی^{۱*}، علی هارون آبادی^۲، سعید ستایشی^۳

^۱ گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران.

^۲ گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران.

^۳ گروه مهندسی پر تویزشکی، دانشکده مهندسی انرژی و فیزیک، دانشگاه صنعتی امیرکبیر، تهران، ایران.

تاریخ دریافت: ۱۴۰۲/۰۴/۱۴ تاریخ بازبینی: ۱۴۰۳/۰۲/۲۷ تاریخ پذیرش: ۱۴۰۳/۰۵/۲۲

نوع مقاله: پژوهشی

چکیده

پیش‌بینی پیوند امری ضروری برای بررسی پیوند بین گره‌ها در شبکه‌های اجتماعی است. گسترش و مدل‌سازی شبکه‌های اجتماعی منجر به پیدایش شبکه‌های اجتماعی به صورت علامت‌دار، جهت‌دار و وزنی می‌شود. روابط کاربران در شبکه‌های اجتماعی علامت‌دار جنبه‌های ذهنی و نامتقارن وابسته به این حوزه را تعریف می‌کنند، پس هر دو اصطلاح اعتماد و عدم اعتماد چالش‌برانگیز هستند. برای حل مسئله پراکندگی در شبکه‌ها و غلبه بر ابهام در روابط، یک روش اعتماد-عدم اعتماد مبتنی بر محاسبات فازی برای محاسبه قدرت پیوندها پیشنهاد می‌شود. هدف روش پیشنهادی پیش‌بینی پیوند برای حل مسئله پراکندگی در شبکه‌های اجتماعی علامت‌دار با ترکیب ویژگی‌های توصیف کاربران در شبکه‌های اجتماعی است که با تاثیر مستقیم گره‌های برتر و تاثیر غیرمستقیم گره‌های معمولی بر پیش‌بینی رتبه‌بندی‌ها ارزیابی می‌شود. اعتماد با یک سیستم فازی مددانی مبتنی بر ویژگی‌های انعکاس شباهت فازی، اعتماد کلی و عدم اعتماد کلی تعیین می‌شود. ارزیابی روش پیشنهادی با معیار صحت بر روی مجموعه داده‌های شبکه‌های اجتماعی Epinions و Slashdot انجام شد. صحت روش پیشنهادی در مجموعه داده‌های Epinions و Slashdot به ترتیب برابر ۰٫۹۹۱ و ۰٫۹۹۸ می‌باشد. نتایج به‌دست آمده نشان می‌دهد روش پیشنهادی نسبت به مشکل پراکندگی داده‌ها در شبکه‌های اجتماعی علامت‌دار قوی عمل می‌کند و این اثربخشی مدل پیشنهادی را بیان می‌نماید.

کلیدواژگان: پیش‌بینی پیوند، شبکه‌های اجتماعی علامت‌دار، اعتماد، عدم اعتماد، محاسبات فازی.

* رایانامه نویسنده مسؤول: a.harounabadi@iauctb.ac.ir

۱- مقدمه

همراه دارند. از این رو، معیارهای انتخاب شده برای تعیین اعتماد نسبت به سایر ویژگی‌ها غالب می‌باشند. ویژگی‌های فوق در روش پیشنهادی، یک راه حل مناسب برای پیش‌بینی پیوند مثبت و منفی در شبکه‌های اجتماعی فراهم می‌کند و به کاربران کمک می‌کند که تا به چه کسانی و به چه اندازه در شبکه‌های اجتماعی اعتماد داشته باشند. از آنجا که در میزان اعتماد و یا عدم اعتماد، قطعیت وجود ندارد، این معیارها براساس روش‌های محاسبات فازی ارائه می‌شود. همچنین در سال‌های اخیر، الگوریتم‌های توصیه‌گر، مورد توجه قرار گرفته‌اند. الگوریتم توصیه‌گر فیلترینگ مشارکتی یکی از پرکاربردترین تکنیک‌های الگوریتم‌های توصیه‌گر است که از داده‌های رتبه‌بندی کاربر برای مدل‌سازی استفاده می‌کند. الگوریتم فیلترینگ مشارکتی، همراه با اطلاعات اجتماعی، با افزودن اطلاعات اجتماعی بین کاربران هنگام مدل‌سازی مشخص می‌شود. با الگوریتم فیلترینگ مشارکتی می‌توان رابطه اعتماد را بر رتبه‌بندی‌های پیش‌بینی شده برای انتخاب گره‌های برتر بکار برد. با تاثیر مستقیم گره‌های برتر بر رتبه‌بندی‌های پیش‌بینی شده می‌توان بر رابطه اعتماد تاثیرگذاری بیشتری داشت [۶].

هر یک از معیارهای پیش‌بینی پیوند به صورت جداگانه و یا تلفیقی در ادبیات تحقیق مورد بررسی، مبتنی بر الگوریتم‌های مختلف مورد بررسی قرار گرفته‌اند. اما ارائه روشی که بتواند این معیارها را به درستی انتخاب نموده تا به گونه‌ای کنار هم قرار گیرند که بیشترین دقت پیش‌بینی را داشته باشد، ضروری می‌باشد. از آنجا که عدم قطعیت در شناسایی رابطه کاربران و امتیازدهی آیت‌ها وجود دارد، ترکیب این معیارها برای پیش‌بینی اعتماد و عدم اعتماد براساس سیستم‌های فازی انجام می‌شود.

بنابراین، نوآوری این تحقیق بر روی موارد ذیل متمرکز می‌گردد:

۱. برای استخراج گره‌های برتر، یک روش انتخاب پیشنهاد می‌شود که گره‌های برتر را براساس دو ویژگی میزان اعتماد و ویژگی مثبت گره‌ها محاسبه می‌کند.
۲. به منظور حل مسئله پراکندگی بین کاربران در شبکه‌ها و افزایش دقت در تشخیص اعتماد، یک روش پیش‌بینی پیوند بین کاربران براساس ترکیب مجموعه از ویژگی‌ها استفاده می‌شود که این ویژگی‌ها عبارتند از: ویژگی‌های انعکاس شباهت فازی اعتماد و عدم اعتماد و مجموعه ویژگی‌های مرتبط با آن و اعتماد کلی و عدم اعتماد کلی با مجموعه ویژگی‌های مرتبط با آن.

پیش‌بینی پیوند به عنوان یکی از زمینه‌های تحقیقاتی پرکاربرد در شبکه‌های اجتماعی شناخته می‌گردد که وظیفه اصلی آن پیش‌بینی پیوندها در آینده‌ای نزدیک بین کاربران و آیت‌ها است. تشکیل این پیوندها و یا عدم پیوندها به معنای ایجاد اعتماد و یا عدم اعتماد بین کاربران می‌باشد. این رفتارها در شبکه‌ها، توجه به شبکه‌های اجتماعی علامت‌دار را بیشتر می‌کند که در طی چندسال اخیر به طور گسترده مورد بررسی قرار گرفته‌اند، زیرا روابط کاربران را بهتر از شبکه‌های اجتماعی بدون علامت نشان می‌دهند. شبکه‌های اجتماعی علامت‌دار رابطه بین دو گره را تعیین می‌کنند که به عنوان پیوند مثبت یا منفی نامگذاری می‌شوند. در واقع در پیش‌بینی پیوند برای روابط علامت‌دار سه نوع ارتباط اجتماعی بین دو کاربر در نظر گرفته می‌شود: مثبت (اعتماد/+)، منفی (عدم اعتماد/(-)) و عدم رابطه (عدم وجود پیوند/صفر). پیوند مثبت نشان‌دهنده وجود یک رابطه قابل اعتماد و پیوند منفی نشان‌دهنده وجود یک رابطه غیرقابل اعتماد است. افزایش علاقه به شبکه‌های اجتماعی علامت‌دار، نیاز به بازنگری در مسائل پیش‌بینی پیوند را افزایش داده است [۱]. محاسبات فازی به صورت تقریبی به جای استدلال دقیق در بیان روابط که فقط اعتماد و یا فقط عدم اعتماد را مدل می‌کند، استفاده می‌شود. برای حل این تضاد برای مقادیر قطعی در روابط اعتماد، از محاسبات فازی استفاده می‌شود.

به منظور بررسی اعتماد در شبکه‌های اجتماعی می‌بایست معیارهایی که ویژگی‌های اجتماعی کاربران را توصیف می‌کنند و بر میزان اعتماد کاربران تاثیرگذاری بسزایی دارند، مورد بررسی قرار گیرند. این ویژگی‌ها باید به گونه‌ای باشند که مشکل پراکندگی بین کاربران در شبکه‌های اجتماعی را به خوبی حل نمایند و همچنین پیش‌بینی پیوند را با دقت بالا و سرعت زیاد انجام دهند. از این رو، ویژگی‌هایی در این تحقیق مورد استفاده قرار می‌گیرند که حائز شرایط باشند. این ویژگی‌ها عبارتند از: تشابه اولویت، تجربه، عمل متقابل، صفات اجتماعی محلی و چگالی اشتراک برای تعیین اعتماد و عدم اعتماد کلی [۲] و [۴] و [۵] و انعکاس شباهت فازی^۱ اعتماد و عدم اعتماد [۳]. ویژگی‌های فوق بر میزان اعتماد و عدم اعتماد تاثیر می‌گذارند زیرا هم رابطه اعتماد و عدم اعتماد را با توجه به رابطه بین دو کاربر و همسایگان آنها بررسی می‌کنند و هم به حل تعارض عقاید کاربران اطمینان دارند. همچنین نسبت به پراکندگی شبکه قوی‌تر عمل می‌کنند و سرعت قابل قبولی را برای اجرا به

^۱ Fuzzy Trust Mirroring

۳. به منظور پیش‌بینی اعتماد و عدم اعتماد در شبکه‌های اجتماعی علامت‌دار، یک سیستم فازی مبتنی بر ویژگی‌های فوق طراحی می‌شود.

ادامه این مقاله به این صورت سازماندهی می‌شود: بخش دوم به کارهای مرتبط با شبکه‌های اجتماعی علامت‌دار می‌پردازد. مدل پیش‌بینی پیوند پیشنهادی در بخش سوم توصیف خواهد شد. بخش چهارم مجموعه داده‌ها و معیارهای ارزیابی مورد استفاده برای تجزیه و تحلیل شبکه‌های اجتماعی علامت‌دار و مطالعه آزمایش‌های روش پیشنهادی را مورد بررسی قرار خواهد داد. نهایتاً، در فصل پنجم، نتیجه‌گیری از یافته‌های کار خود به همراه برخی از جهت‌گیری‌های تحقیقاتی آینده ارائه خواهد شد.

۲- کارهای مرتبط

به عنوان زمینه تحقیقاتی پایه شبکه‌های اجتماعی، پیش‌بینی پیوند به طور روزافزون جامعه تحقیقاتی این زمینه را به خود جذب کرده است. تحقیقات متعددی برای مشکل پیوندهای گمشده و مقابله با مشکل پراکندگی در شبکه‌های اجتماعی علامت‌دار^۱ انجام شده که برخی فقط بر اعتماد تمرکز می‌کنند، برخی دیگر هم بر اعتماد و هم بر عدم اعتماد. هریک از تحقیقات مورد بررسی در ادبیات موضوع، دارای روش پیشنهادی متفاوتی می‌باشند. به طور نمونه برخی از آنها مبتنی بر روش‌های بانظارت و برخی مبتنی بر روش‌های بدون نظارت هستند، برخی از روش‌ها احتمالاتی و برخی مبتنی بر منطق ذهنی هستند و یا برخی از روش‌ها براساس معیارهای تعاملی محلی و برخی براساس معیارهای تعاملی سراسری می‌باشند. با توجه به اینکه الگوریتم‌های پیش‌بینی پیوند براساس معیار شباهت، دقت کمتری را نسبت به برخی از روش‌ها به دلیل دامنه کاربردها ایجاد می‌کند، اخیراً، روش‌های بی‌شماری برای پیش‌بینی پیوند ارائه شده است که در ادامه برخی از آنها مورد بررسی قرار می‌گیرد. پیش‌بینی پیوند در شبکه‌های اجتماعی به چهار روش مبتنی بر شباهت، مبتنی بر احتمالات، مبتنی بر الگوریتم و روش‌های ترکیبی تقسیم می‌شوند که هر یک از این گروه‌ها، خود شامل زیر گروه‌هایی می‌باشد که در آن گروه بررسی می‌گردد [۷].

۲-۱- رویکردهای پیش‌بینی پیوند مبتنی بر شباهت

روش‌های پیش‌بینی پیوند مبتنی بر معیار شباهت به سه گروه: شاخص‌های محلی، شاخص‌های سراسری و شاخص‌های شبه محلی تقسیم می‌شوند. به منظور بررسی معیارهای شباهت، هدف کار ژو و

ما [۸]، پیش‌بینی علامت با دو شاخص: شباهت و اعتبار ترجیحی بود. شاخص اندازه‌گیری شباهت-عدم شباهت برای ویژگی‌های مشترک جفت گره‌ها استفاده می‌شود تا تعادل اجتماعی را نگه دارد و تئوری حالت را حفظ کند. مزیت اندازه‌گیری ویژگی‌های مشترک جفت گره‌ها در مدل پیش‌بینی پیشنهادی منجر می‌شود که به بهترین شکل از داده‌های موجود برای پیش‌بینی علامت پیوندهای از دست رفته استفاده شود. اما اشکال آن است که فقط دو شاخص اندازه‌گیری شباهت-عدم شباهت در اعتبارسنجی مدل، استفاده می‌گردد. شا وراس دی سنا روسا و همکاران [۹]، از خوشه‌بندی محلی برای یافتن گروهی از اشیاء مشابه استفاده نمودند که مزیت آن توانایی بهره‌برداری از ساختارهای همبستگی متعدد بین کاربرانی می‌باشد که ترجیحات خود را برای اشیایی که احتمالاً دارای ویژگی‌های مشابه هستند، بیان می‌کنند. اما این روش حساس به رتبه‌بندی اختصاص یافته موجود به هر خوشه می‌باشد و نیاز به حافظه و زمان بیشتری برای ذخیره خوشه‌ها دارد. در روش‌های مبتنی بر شباهت شبه محلی، قربان زاده و همکاران [۱۰] به بررسی ارتباط بین گره‌ها با معیارهای محلی ترکیبی و رابطه همسایه‌ها و تعداد همسایه‌های مشترک با توجه به مرکزیت و اعتبار همسایه‌های مشترک پرداختند و پیش‌بینی پیوند بین دو گره که همسایگی مشترک ندارند، را بررسی نمودند. ونگ و همکاران [۱۱]، به بهبود فیلترهای مشترک مبتنی بر همسایگی با استفاده از ترکیب معیارهای شباهت ترکیبی پرداختند و به مزایایی مانند عدم درج اطلاعات کاربران یا گزینه‌ها و کاهش زمان و حافظه مصرفی اشاره نمودند، اما مدیریت پویا بودن ترجیحات کاربر امکان پذیر نبود. ایکسیو و همکاران [۱۲]، به ارزیابی اعتماد در شبکه‌های اجتماعی علامت‌دار براساس شبکه عصبی مبتنی بر مکانیسم توجه دوسطحی با ویژگی‌های سراسری کاربران پرداختند. ویژگی‌های برتر کار ایشان، مکمل بودن داده‌های کاربران با یکدیگر، بررسی همزمان اطلاعات همسایگان مورد اعتماد کاربران و تداوم زمانی رفتارهای کاربران می‌باشد.

۲-۲- رویکردهای پیش‌بینی پیوند مبتنی بر احتمالات

روش‌های پیش‌بینی پیوند مبتنی بر معیار احتمالات به چهار گروه: مدل فاکتورسازی تانسور احتمالی، مدل متغیر پنهان احتمالی، مدل مارکوف و مدل‌سازی برچسب پیوند تقسیم می‌شوند. در بررسی روش‌های مبتنی بر احتمالات، ونگ و همکاران [۱۳] بیان نمودند که پیش‌بینی پیوند در شبکه اطلاعات نا همگن براساس یک روش

¹ Sparsity in Signed Social Networks

در بررسی روش‌های مبتنی بر الگوریتم، یاگی و همکاران [۱۸]، هدف از تحقیق خود را پیش‌بینی پیوند با استفاده از ترکیب شبکه عصبی و الگوریتم‌های تکاملی: ژنتیک، بهینه‌سازی ازدحام ذرات و جستجوی حشرات بیان نمودند و مزیت کار ایشان، استفاده از شبکه عصبی به عنوان دسته‌بند برای مجموعه داده پیش‌بینی پیوند و مدیریت توزیع نامتوازن مجموعه داده با تکنیک نمونه‌برداری و میانگین‌گیری به عنوان تابع برازندگی بود. سریلاتا و همکاران [۱۹]، پیش‌بینی پیوند در شبکه‌های اجتماعی با الگوریتم بهینه‌سازی کرم شب‌تاب براساس اطلاعات ویژگی گره‌ها را معرفی نمودند و از پیش‌بینی پیوند مبتنی بر کرم شب‌تاب برای بهبود عملکرد روش شباهت بهره بردند. مزیت کار ایشان، استفاده از گره‌های اجتماعی و ترکیب ویژگی‌های آنها با همدیگر بود. گیو و همکاران [۲۰]، شباهت توپولوژیکی نزدیکترین همسایه‌ها در یک شبکه هدایت شده را برای حل مشکل بررسی جهت لبه و اطلاعات گره شبکه پیشنهاد کردند. در ابتدا، این مطالعه باعث بهبود شاخص سورنسن در شبکه‌های هدایت شده شد و سپس شکل ماتریسی هر شاخص پایه را با جبر ماتریسی بیان کردند. نهایتاً توپولوژی نزدیکترین همسایگان هر شاخص پایه را برای به دست آوردن شاخص شباهت توپولوژیکی نزدیکترین همسایه‌ها استخراج نمودند. نورایی آباده و میرزایی [۲۱]، به بررسی روش یادگیری ماشین برای استخراج ویژگی‌ها در پیش‌بینی ارزش اعتماد شبکه‌های اجتماعی علامت دار پرداختند تا به درک ماهیت پویای شکل‌گیری و تکامل گروه در شبکه‌های اجتماعی برای توصیف بهتر نحوه جمع‌آوری گره‌های اعتماد افراد و تشکیل جوامع دست پیدا نمایند. تیلمن و همکاران [۲۲]، پیش‌بینی پیوند شبکه‌های چندسطحی با سه الگوریتم مبتنی بر تعداد و سطح را بررسی نمودند و از ساختار همبستگی نوع ارتباطات براساس یک آستانه تئوری در شبکه‌های دنیای واقعی بهره بردند، اما تعریف دقیق مقدار آستانه با توجه به مولفه‌های الگوریتم امکان‌پذیر نبود.

۲-۴- رویکردهای پیش‌بینی پیوند مبتنی بر روش‌های

ترکیبی

در بررسی روش‌های ترکیبی که براساس ترکیب روش‌های مذکور می‌باشد، لیو و همکاران [۲۳]، پیش‌بینی احتمالات با شبکه کانولوشن گراف برای شناسایی مجدد افراد را بررسی نمودند و اهدافی از جمله: تشکیل یک زیرگراف نمونه برای هر گره با اطلاعات متنی محلی مبتنی بر یادگیری و مبتنی بر اطلاعات غنی از گره‌های مرکزی، انعطاف‌پذیری و اثربخشی روش پیشنهادی را بیان نمودند، اما برای یادگیری شبکه، استخراج ویژگی و تخمین

یادگیری عمیق گراف بهبود یافته برای یادگیری تعبیه راس پیش‌بینی شده با حداقل اطلاعات ازدست رفته برای کشف جوامع محلی بکار می‌رود و مزایای این روش عبارت است از: کشف جامع با شباهت درون جامعه و بین جامعه، استفاده از ویژگی‌های ذاتی موجودیت‌های مختلف شبکه و یک لایه هسته بهینه‌سازی شده برای معیار شباهت تعبیه شده در راس‌های جفت نودها. طهماسبی و همکاران [۱۴]، در حیطه فعالیت فاکتورسازی ماتریس، مدل فاکتورسازی ماتریس جمعی زمانی و اجتماعی برای سیستم‌های توصیه کننده برحسب تنظیمات ترجیحی کاربر با مرور زمان را بررسی نمودند و هدف آنها جمع اطلاعات رتبه‌بندی کاربران و اطلاعات اعتماد اجتماعی به طور مشترک در روش فاکتورسازی ماتریس و قابلیت توسعه روش پیشنهادی بر روی مجموعه داده‌های بزرگ با میلیون‌ها کاربر و گزینه بود، اما مشکل آن بود که تعاملات کاربران در هر زمان را براساس اعتماد ضمنی و نه براساس اعتماد صریح بررسی کردند. همچنین مشکل کاربران مبتدی (شروع سرد) که هیچ رابطه اعتمادی نداشتند را حل نکردند. ژانگ و همکاران [۱۵]، به بررسی الگوریتم توصیه اجتماعی مبتنی بر تجزیه ماتریس تصادفی در شبکه اجتماعی با ایجاد یک ماتریس برای شبکه اجتماعی و یک ماتریس برای رتبه‌بندی کاربر پرداختند و تابع هدف فاکتورسازی ماتریس را با قاعده گذاری اجتماعی طراحی نمودند. مزیت روش ایشان، ترکیب اطلاعات کاربران و امتیازدهی به آنها در یک ماتریس با الگوریتم فاکتورسازی ماتریس بود، اما ایراد آن، تاثیر محدود بودن جستجو به کاربرانی که با هم در ارتباط هستند، بر روی پیشنهادات جستجو بود. هی و همکاران [۱۶]، به بررسی روش مبتنی بر شباهت با استخراج اطلاعات شباهت گره و اطلاعات ساختار هندسی توپولوژی پیوند و فاکتورسازی ماتریس نیمه غیرمنفی باینری منظم برای حفظ همپوشانی تشخیص جوامع پرداختند و اهداف همپوشانی تشخیص جوامع و تقسیم گره‌ها به خوشه‌های مختلف با پیوندهای مثبت درون خوشه‌ای و پیوندهای منفی بین خوشه‌ای را دنبال کردند، اما با توجه به میلیون‌ها گره در شبکه‌های اجتماعی، زمان مصرفی الگوریتم ایشان بالا بود. نصیری و همکاران [۱۷]، منظم‌سازی نمودار محکم فاکتورسازی ماتریس غیر منفی برای پیش‌بینی پیوند در شبکه‌های مورد بررسی را برای مکمل بودن ساختار توپولوژیکی و ویژگی‌های گره‌ها در پیش‌بینی پیوند ارزیابی نمودند.

۲-۳- رویکردهای پیش‌بینی پیوند مبتنی بر الگوریتم

روش‌های پیش‌بینی پیوند مبتنی بر معیار الگوریتم به سه گروه: فراابتکاری، فاکتورسازی ماتریس و یادگیری ماشین تقسیم می‌شوند.

در پیش‌بینی پیوند دست یافت. همچنین تعیین ساختار همبستگی مشخص برای شبکه‌های چندسطحی در رسیدن به این هدف، بسیار حائز اهمیت می‌باشد [۲۲]. با بررسی کارهای انجام شده در زمینه فاکتورسازی ماتریس مشاهده شد که برای پیش‌بینی پیوند اطلاعات رتبه‌بندی کاربران و اطلاعات اعتماد اجتماعی به طور مشترک در روش فاکتورسازی ماتریس جمع آوری شده است. تعاملات کاربران در هر زمان براساس اعتماد ضمنی بررسی می‌شود و اعتماد صریح نادیده گرفته شده است و حل مشکل کاربران مبتدی (شروع سرد) که هیچ رتبه و هیچ رابطه اعتمادی ندارند، بررسی نشده است [۱۴]. به منظور همپوشانی گره‌ها در تشخیص جوامع، تقسیم گره‌ها به خوشه‌های مختلف با پیوندهای مثبت درون خوشه‌ای و پیوندهای منفی بین خوشه‌ای انجام می‌شود که این فرایند منجر به نادیده گرفتن زمان مصرفی و نهایتاً بالا رفتن زمان مصرفی الگوریتم با توجه به میلیون‌ها گره در شبکه‌های اجتماعی می‌شود. به منظور بهبود پیش‌بینی پیوند و زمان مصرفی می‌توان بررسی همسایگی و نزدیکی گره‌ها در مرتبه اول و دوم را بررسی نمود [۱۶].

در تحقیقات مبتنی بر روش یادگیری ماشین مشخص گردید که روش‌های مبتنی بر یادگیری عمیق، به علت بررسی چندسطحی، نتایج دقیق‌تری فراهم می‌نمایند. ولی به نسبت، زمان مصرفی بیشتری به علت چندسطحی بودن دارند [۱۳]. به منظور افزایش دقت در بهبود پیش‌بینی پیوند، می‌توان از توابع برازندگی مختلف به خصوص توابع فازی استفاده نمود [۱۸]. در کار [۱۹] از روش‌های مبتنی بر شباهت برای پیش‌بینی پیوند استفاده شده است که هم جهت لبه شبکه را برای پیش‌بینی پیوندها در نظر می‌گیرد و هم از اطلاعات گره‌های شبکه استفاده می‌کند.

با بررسی کارهای قبلی، مشخص گردید علی‌رغم تلاش‌ها و تحقیقات بسیار جهت حل مسئله پیوندهای گمشده و پراکندگی شبکه‌ها، بررسی شبکه‌های اجتماعی علامت‌دار برای نشان دادن تعامل افراد با یکدیگر برای پیش‌بینی پیوند مبتنی بر عدم قطعیت و سیستم‌های فازی مبتنی بر ویژگی‌های این تحقیق بیان نشده‌اند. به طور نمونه، در روش‌های مبتنی بر پیش‌بینی دوست، تعداد دوستان مشترک در فرایند پیش‌بینی پیوند تاثیرگذار است. اما این روش معایبی دارد از جمله عدم تطابق علایق که برای گسترده نمودن شبکه کاربران بدون فایده است [۲۲]. در شبکه‌های اجتماعی، با تجزیه و تحلیل اعضای مجاور هر گره، سیستم لبه‌هایی که در آینده‌ای نزدیک به شبکه اضافه می‌شوند، پیش‌بینی می‌کند. این استراتژی بر ساختار شبکه تمرکز می‌کند درحالی‌که نقص این روش در آن است که خصوصیات ذاتی گره‌ها در شبکه نادیده گرفته شده است [۲۰]. با بررسی ادبیات موضوع مشخص گردید برخی از ویژگی‌ها از نظر پردازش در دقت و

شباهت می‌توانستند روش‌های قوی‌تر بکار ببرند. هیو و همکاران [۲۴]، به منظور رد شایعات به موقع و در زمان واقعی به منظور اعتمادسازی، ترکیبی از الگوریتم‌های بهینه سازی جهش قورباغه و بهینه سازی ازدحام ذرات پیشنهاد دادند که برای ایجاد اعتماد از سه ویژگی درجه صمیمیت، قابلیت اطمینان و اعتبار استفاده نمودند و دو زیر الگوریتم برای تکذیب شایعات به موقع و تکذیب سریع شایعات برای مسدود کردن مداوم شایعات در طول تکامل حقایق در هر خوشه را بکار بردند.

با بررسی تحقیقات انجام شده در زمینه ادبیات موضوع مشخص گردید که در کارهای انجام شده برای روش‌های مبتنی بر شباهت، علی‌رغم مزایای آنها، معایبی وجود دارد. به طور نمونه در کارهای [۸] و [۹]، فقط معیارهای محلی مسئله در شبکه‌های اجتماعی مورد بررسی قرار گرفته‌اند. درحالی‌که کاربرد معیارهای محلی و سراسری در کنار یکدیگر در بهبود دقت پیش‌بینی پیوند تاثیر بسزایی خواهد داشت. در کارهای [۱۰] و [۱۲]، مدل شبکه‌های اجتماعی براساس گراف‌های بدون علامت هستند و از نوع گراف‌های جهت دار، علامت‌دار و یا وزنی استفاده نشده است که می‌توان از ترکیب این ویژگی‌ها استفاده نمود. به منظور بررسی دقیق پیوند بین گره‌ها، می‌توان از ویژگی‌های مناسب و همچنین ترکیب ویژگی‌ها [۸] و [۱۱]، استفاده نمود. در بین روش‌های احتمالی برای پیش‌بینی پیوند در کارهای [۱۰] و [۲۳]، مشخص شد می‌توان از سایر الگوریتم‌ها برای آزمایش عملکرد روش پیشنهادی و همچنین اعمال روش‌های قوی برای یادگیری شبکه، استخراج ویژگی و تخمین شباهت برای بهبود پیش‌بینی پیوند استفاده نمود. همچنین با ترکیب اطلاعات کاربران و امتیازدهی به آنها در یک ماتریس واحد، الگوریتم توصیه اجتماعی دوست پیشنهاد می‌شود که فضای جستجو هدف آن محدود می‌باشد. می‌توان فرایندهای جستجو را بر روی کاربران بیشتری که با هم در ارتباط هستند، اعمال نمود تا فضای جستجو و الگوریتم جستجوی جدیدی را بکار برده تا نحوه امتیازدهی به کاربران دقیق گردد [۱۵]. برای روش‌های فراابتکاری مبتنی بر پیش‌بینی پیوند، می‌توان از سایر الگوریتم‌های تکاملی جهت بررسی فرایند بهبود پیش‌بینی پیوند استفاده نمود [۱۹] و [۲۳]. از ویژگی‌های مطرح شده برای الگوریتم‌ها به صورت ترکیبی در مابقی الگوریتم‌های این گروه استفاده می‌شود تا میزان تاثیرگذاری ویژگی‌ها در شرایط متفاوت بررسی گردد. در شبکه‌های چند سطحی به منظور حفظ ساختار همبستگی نوع ارتباطات براساس یک آستانه تئوری در شبکه‌های دنیای واقعی می‌توان استفاده نمود. اما تعیین دقیق مقدار آستانه با توجه به مولفه‌های الگوریتم، بسیار مهم می‌باشد تا براساس آن بتوان به نتایج بهتری

و قوی بودن نسبت به پراکندگی شبکه‌های اجتماعی و صحت بالا را دارد.

- ویژگی‌های محلی و سراسری مبتنی بر خوشه‌بندی^۲ [۲۵]: براساس معیارهای شباهت با تعیین ویژگی‌های جدید مربوط به موقعیت و رفتار دینامیکی گره‌ها انجام می‌شود.
- ویژگی انعکاس شباهت فازی در کار شیرگاهی [۳]، مبتنی بر اطلاعات محلی بوده و نیاز به تصمیم‌گیری سراسری ندارد. از این‌رو، روشی توزیع شده است و علاوه بر دقت بالا، سرعت قابل قبولی دارد. با ترکیب ویژگی‌های اثرگذار و ارائه راهکار سیستم فازی، مقیاس‌های، صحت، دقت و فراخوانی بهبود می‌یابد.

جدول ۱ خلاصه‌ای از نمادها و علائم را نشان می‌دهد.

جدول ۱. خلاصه‌ای نمادها و علائم

علائم	توصیف	علائم	توصیف
$t(u, v)$	وزن یال از گره u به گره v	$d(u, v)$	فاصله اقلیدسی بین بردارهای u و v
$ F_{u_i} $	آیتم‌های مورد علاقه کاربر u	$ NF_{u_i} $	آیتم‌های غیرعلاقه توسط کاربر u
N_T	تعداد کل گره‌ها در اعتماد مستقیم به گره i	M	ماتریس مجاورت با سطرها و ستون‌های برای نام گره‌ها
T_{ji}	مقدار اعتماد مستقیم از گره i به گره j	C_v	تعداد همسایگان گره v
α_{ij}	مقدار اعتماد مستقیم گره i به گره j	θ_{u_i}	برداری براساس بردار رتبه‌بندی کاربر u_i
$ S_j^- $	قدرت برخوردهای منفی از کاربر j	$ S_j^+ $	قدرت برخوردهای مثبت از کاربر j
b_u	بایاس کاربر	b_j	بایاس آیتم
q_j	بردار ویژگی پنهان d بعدی آیتم j	p_u	بردار ویژگی پنهان d بعدی کاربر u
μ	میانگین رتبه‌بندی کلی	α	رابطه وزنی بین اولویت گره و اولویت برتر
Experts	گره‌های برتر انتخاب شده	$W_{u,e}$	وزن اعتماد گره u به گره برتر e
Rep_i	اعتبار گره i	$ I_{u_i} $	بردار رتبه‌بندی کاربر u_i

سرعت برای پیش‌بینی پیوند در برابر پراکندگی شبکه نسبت به برخی دیگر بهتر عمل می‌کنند. همچنین برخی از این ویژگی‌ها به صورت جداگانه غالباً در سایر تحقیقات مورد بررسی قرار گرفته‌اند. به این ترتیب، ترکیب ویژگی‌ها می‌تواند در بهبود پیش‌بینی پیوند در شبکه‌های اجتماعی موثر واقع شود.

۳- روش پیشنهادی

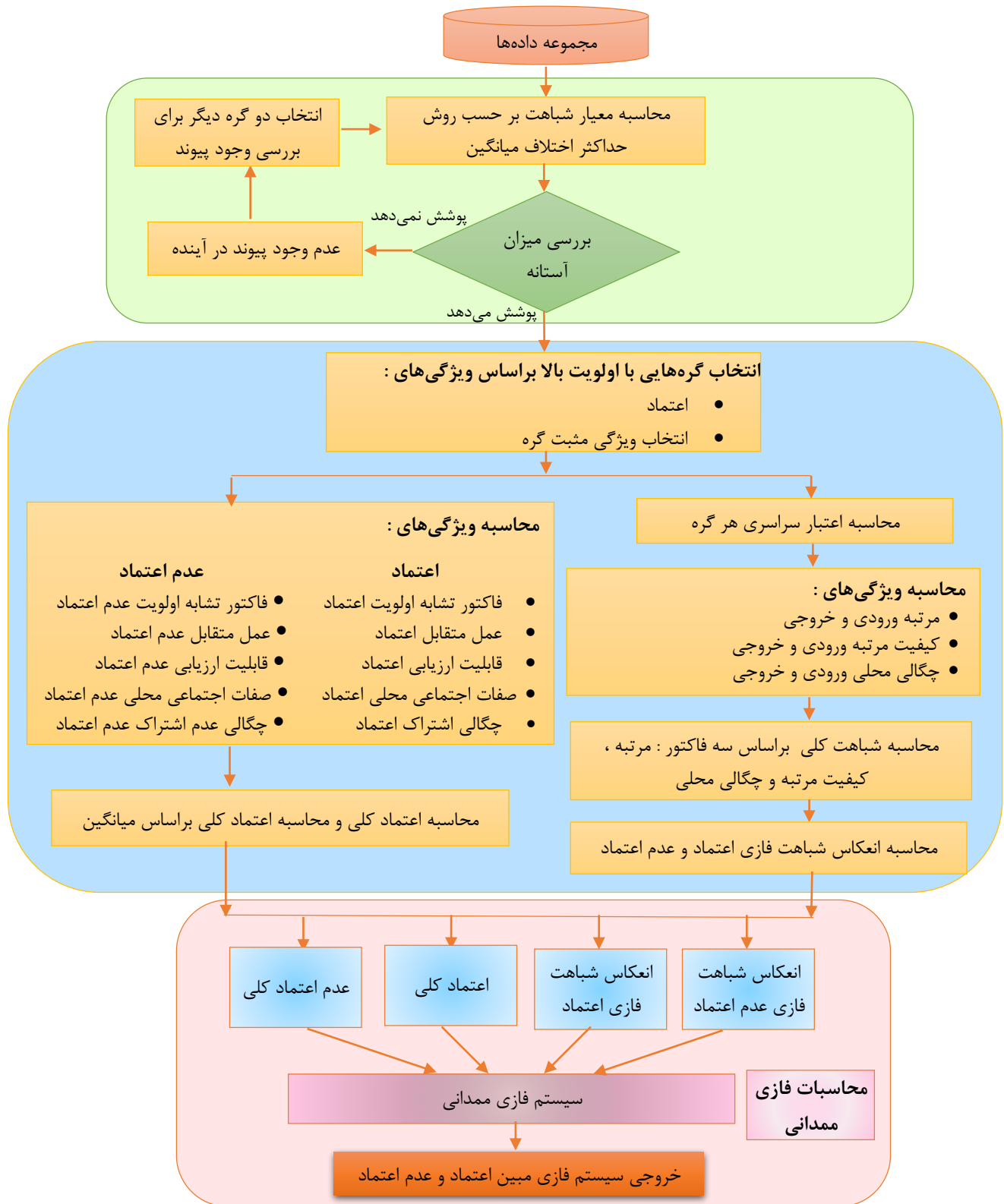
در این تحقیق مدل پیشنهادی متفاوتی برای پیش‌بینی پیوند در شبکه‌های اجتماعی علامت‌دار براساس اطلاعات گره‌ها و قدرت پیوندها بین کاربران برای شبکه‌های اجتماعی علامت‌دار ارائه خواهد شد. مبنای اصلی برای روش پیشنهادی، مدل محاسباتی فازی است که مبتنی بر قدرت پیوندهای محاسبه شده از نظر مقادیر اعتماد-عدم اعتماد و انعکاس شباهت فازی اعتماد و عدم اعتماد بین کاربران شبکه است و به اصطلاح پیش‌بینی پیوند در شبکه‌های اجتماعی علامت‌دار با اعتماد و عدم اعتماد مبتنی بر فازی^۱ نامیده می‌شود. الگوریتم روش پیشنهادی در شکل ۱ نشان داده شده است. روش پیش‌بینی پیوند پیشنهادی، با بررسی احتمال وجود پیوند بین کاربران و آیتم‌ها در آینده براساس روش‌های مبتنی بر شباهت از جمله روش معیار میانگین واگرایی، میزان شباهت‌هایشان را بررسی می‌کند. نهایتاً سیستم فازی ممدانی با چهار ورودی اعتماد کلی، عدم اعتماد کلی و انعکاس شباهت فازی اعتماد و عدم اعتماد طراحی می‌شود که خروجی سیستم فازی مبین اعتماد و یا عدم اعتماد با پیش‌بینی پیوند مثبت یا منفی می‌باشد.

مهم‌ترین مزیت روش پیشنهادی، استفاده از سیستم‌های فازی طراحی شده مبتنی بر عدم قطعیت در بررسی میزان اعتماد و عدم اعتماد کاربران نسبت به یکدیگر در شبکه‌های اجتماعی به همراه ترکیب ویژگی‌های موثر در تعیین میزان اعتماد می‌باشد. سایر مزایای ویژگی‌های موثر مورد استفاده عبارتند از:

- فاکتور تشابه اولویت، تجربه و عمل متقابل در کار گریدها و همکاران [۴] و [۵]. این ویژگی‌ها هم اطلاعات محلی و هم اطلاعات سراسری کاربران را در مورد پیوندهای متقابل بین کاربران و نظریه تعادل اجتماعی براساس اعتماد و عدم اعتماد بررسی می‌کند.
- ویژگی صفات اجتماعی محلی در کار اکیلال و همکاران [۲]، خصوصیات اجتماعی بسیاری وجود دارند که بر اعتماد تاثیر می‌گذارند، علت انتخاب این ویژگی، سادگی و سهولت درک آنها می‌باشد. علاوه بر این، مزایای دیگری مانند انعطاف‌پذیری

² Local and Global Clustering (LGC)

¹ link prediction in signed social networks with fuzzy model - based trust and distrust (LPSSN-FTD)



شکل ۱. بلوک دیاگرام روش پیشنهادی

کاربران مورد اعتماد تعیین می‌شود. مقدار $Global_u$ و $Activity_u$ با رابطه (۳) و (۴) محاسبه می‌شود [۶]:

$$Global_u = \frac{k - \min}{\max - \min} \quad (3)$$

$$Activity_u = \frac{N_u - \min}{\max - \min} \quad (4)$$

که در آن $Global_u, Activity_u \in [0,1]$ ، \max و \min به ترتیب حداکثر و حداقل مقدار اعتماد کاربر، k درجه اعتماد کاربر u و N_u تعداد آیتم‌های رتبه‌بندی کاربر u برای مقادیر بیشتر از ۳ را نشان می‌دهد، اگر مقدار رتبه‌بندی هر کاربر برای هر آیتم بیشتر از ۳ باشد، نگرش مثبت ایجاد شده و کاربر در فرایند رتبه‌بندی فعال تر می‌شود.

- **تعریف ۲:** با تعریف ۱، می‌توان هر گره را به یک گره اولیه براساس رابطه (۵) تبدیل نمود:

$$Possibility_u = \frac{1}{2} (Global_u + Activity_u) \quad (5)$$

که $Possibility_u \in [0,1]$ ، به صورت نزولی مرتب شده و چند مقدار اول به عنوان مجموعه گره‌های برتر اولیه در نظر گرفته می‌شود.

- **تعریف ۳:** براساس تعریف ۲، هر کاربر می‌تواند به گره برتر اولیه تبدیل شود. ابتدا تعداد گره برتر اولیه تعیین می‌گردد، گره‌های برتر اولیه براساس تعداد دفعاتی که مورد اعتماد هستند تعیین و رتبه‌بندی می‌شوند. کاربران در n تای بالا به عنوان برتر انتخاب می‌شوند. ویژگی‌های گره‌های برتر عبارتند از: ۱- از اطلاعات رتبه‌بندی و اعتماد برای تعریف گره استفاده می‌شود. ۲- باتوجه به ویژگی‌های گره‌های برتر، تعداد دفعات اعتماد و نگرش مثبت رتبه‌بندی تعیین می‌شود. ۳- گره‌های مورد اعتماد گره‌های برتر اولیه، کاملاً بررسی می‌شوند. رتبه‌بندی پیش‌بینی برای گره‌های برتر در رابطه (۶) تعریف می‌شود [۶]:

$$\hat{r}_{u,j} = \frac{\sum_{e \in Experts} W_{u,e} q_j^T p_e}{\sum_{e \in Experts} |W_{u,e}|} \quad (6)$$

کاربر مورد اعتماد به طور غیرمستقیم با تغییر بردار ویژگی کاربر بر رتبه‌بندی‌های پیش‌بینی تأثیر می‌گذارد، پس بردار ویژگی کاربر با بردار ویژگی دوست مورد اعتماد نزدیک می‌شود، یعنی:

$$\sum_u \sum_{v \in T_u} W_{u,v} \|p_u - p_v\|_F^2$$

رابطه (۷) است [۶]:

$$\hat{r}_{u,j} = u + b_u + b_j + \alpha q_j^T p_u + (1 - \alpha) \frac{\sum_{e \in Experts} W_{u,e} q_j^T p_e}{\sum_{e \in Experts} |W_{u,e}|} \quad (7)$$

۳-۱- معیار میانگین واگرایی مبتنی بر محاسبه اعتماد - عدم اعتماد

مقداردهی اولیه، براساس اطلاعات موجود در مورد کاربران است که میزان شباهت کاربران را براساس اولویت‌هایشان بررسی می‌کند. برای این منظور، می‌توان از معیارهای مبتنی بر شباهت استفاده نمود که در این تحقیق از روش معیار میانگین واگرایی^۱ بهره گرفته می‌شود. معیار میانگین واگرایی برای محاسبه شباهت در رتبه‌بندی کاربران براساس عادات رتبه‌بندی به منظور پیوند آینده بین کاربر u_i و کاربر u_j استفاده می‌گردد. MMD بین یک جفت کاربر می‌تواند با رابطه (۱) محاسبه می‌شود [۲۶]:

$$Similar(u_i, u_j) = \frac{1}{1 + (\frac{1}{r} \sum_{i=1}^r \{(\theta_{u_i} - \theta_{u_j})^2 - \frac{1}{|I_{u_i}|} - \frac{1}{|I_{u_j}|}\})} \quad (1)$$

پس وجود پیوند آینده بین کاربر u_i و کاربر u_j با چک کردن شباهت بین آنها پیش‌بینی خواهد شد. اگر شباهت بین دو کاربر بیشتر یا مساوی با مقدار آستانه مطابق رابطه (۱) باشد، پس می‌توان پیوند آینده را پیش‌بینی کرد؛ در غیر این صورت، اگر شباهت کمتر از مقدار آستانه باشد، پیش‌بینی پیوند با اطلاعات موجود امکان پذیر نیست و از این رو هیچ پیش‌بینی‌ای وجود نخواهد داشت.

مقدار آستانه با رابطه (۲) تعریف می‌شود:

$$Similar(u_i, u_j) = \begin{cases} \geq \text{مقدار آستانه} & \text{future link can be predicted} \\ < \text{مقدار آستانه} & \text{prediction not possible} \end{cases} \quad (2)$$

اگر مقدار آستانه بالاتر باشد، کاربران تشابه بیشتری دارند (و بالعکس) و به این ترتیب، صحت پیش‌بینی پیوند گمشده بین کاربران افزایش می‌یابد. با این حال، از آنجا که مجموعه داده‌های دنیای واقعی بسیار پراکنده هستند، مقدار آستانه بالا منجر می‌شود که شانس یافتن کاربران مشابه کاهش یابد.

۳-۲- انتخاب گره‌های برتر

در این بخش، رابطه اعتماد براساس بردار ویژگی تعیین می‌شود تا گره‌های برتر تولید شوند. در واقع، رابطه اعتماد در تعیین گره‌هایی با اولویت بیشتر تأثیر می‌گذارد [۶].

- **تعریف ۱:** درجه اعتماد $Global_u$ و نگرش مثبت $Activity_u$ در شبکه‌های اجتماعی، اگر گره‌ها بارها مورد اعتماد واقع شوند، احتمال برتر شدن آنها بیشتر می‌شود. درجه اعتماد کاربر u با تعداد

¹ Mean measure of divergence (MMD)

۳-۳- محاسبه انعکاس شباهت فازی اعتماد و عدم

اعتماد

در این مرحله برای تعیین اعتماد و عدم اعتماد بین دو گره، از سیستم فازی استفاده می‌شود. این بخش از روش پیشنهادی انعکاس شباهت فازی مبتنی بر اعتماد و عدم اعتماد^۱ نامیده می‌شود که مبتنی بر ویژگی‌هایی از جمله مرتبه ورودی و خروجی، کیفیت مرتبه ورودی و خروجی و چگالی محلی ورودی و خروجی است که در ادامه شرح داده می‌شوند. سپس، ترکیبی از ویژگی‌های ذکر شده بیان می‌گردد تا میزان شباهت بین دو گره بدست آید. نهایتاً، مقدار انعکاس شباهت فازی اعتماد و عدم اعتماد براساس میزان تشابه محاسبه می‌شود. محدوده مجموعه‌های فازی، قواعد و ساختار سیستم‌های فازی براساس مقادیر محاسبه شده ویژگی‌های مطرح شده در گره‌ها تعیین می‌گردد. دامنه متغیرهای ورودی و خروجی مجموعه‌های فازی براساس بخش‌هایی از مجموعه داده‌های Epinions تعیین می‌شود تا توزیع نسبتاً مساوی از داده‌ها در این مجموعه‌ها وجود داشته باشد. تمام توابع مجموعه‌های فازی برای متغیرهای ورودی و خروجی پارامترها به صورت دوزنقه ای یا مثلثی هستند و سه مجموعه فازی کم، متوسط و زیاد برای ویژگی‌های ورودی و ویژگی‌های خروجی در نظر گرفته می‌شود [۳].

ابتدا اعتبار هر گره براساس رابطه (۸) محاسبه می‌شود:

$$Rep_i = \begin{cases} 0 & N_T = 0 \\ \frac{\sum_{j \neq i, T_{ji} \neq 0} T_{ji}}{N_T} & N_T \neq 0 \end{cases} \quad (8)$$

ویژگی‌های موثر در محاسبه انعکاس شباهت فازی به شرح زیر معرفی می‌گردند:

- مرتبه بیرونی، پارامتری است که می‌گوید تا چه حد یک کاربر در یک شبکه با اطلاعات سایر کاربران ارتباط دارد [۳].

$$D_{out}(i) = \frac{1}{1 - Rep_i} \sum_{j \neq i} Rep_j \alpha_{ij} \quad (9)$$

- مرتبه درونی، پارامتری است که اطلاعات یک کاربر را از سایر کاربران دریافت می‌کند. مرتبه درونی تعداد پیوندهای ورودی با کاربر کانونی نسبت به مقدار حداکثر آن است [۳].

$$D_{in}(i) = \frac{1}{1 - Rep_i} \sum_{j \neq i} Rep_j \alpha_{ji} \quad (10)$$

- کیفیت رتبه بیرونی و کیفیت رتبه درونی میزان دستیابی به کاربران را که از کاربران دیگر می‌رسد یا رسیدن به دیگر کاربران را در دو مرحله می‌سنجد. بنابراین، این پارامترها میزان ارتباط یک

کاربر با کاربران دیگر را نشان می‌دهد که خود دارای مرتبه‌های بالایی هستند [۳]. کیفیت رتبه بیرونی با رابطه (۱۱) تعریف می‌شود:

$$Q_{out}(i) = \frac{1}{1 - Rep_i} \sum_{j \neq i} Rep_j (\alpha_{ji} - D_{out}(i)) \quad (11)$$

$$(D_{out}(j) - \frac{1}{1 - Rep_i} \sum_{k \neq i} Rep_k D_{out}(k))$$

- کیفیت رتبه درونی با رابطه (۱۲) تعریف می‌شود:

$$Q_{in}(i) = \frac{1}{1 - Rep_i} \sum_{j \neq i} Rep_j (\alpha_{ji} - D_{in}(i)) \quad (12)$$

$$(D_{in}(j) - \frac{1}{1 - Rep_i} \sum_{k \neq i} Rep_k D_{in}(k))$$

- چگالی محلی مرتبه بیرونی، میزان انتقال اطلاعات کاربر به همسایگان متصل را اندازه‌گیری می‌کند [۳].

$$LD_{out}(i) = \frac{\sum_{j \neq i} \sum_{k \neq i, k \neq j} Rep_j Rep_k \alpha_{ij} \alpha_{ik} \alpha_{jk}}{\sum_{j \neq i} \sum_{k \neq i, k \neq j} Rep_j Rep_k \alpha_{ij} \alpha_{ik}} \quad (13)$$

- چگالی محلی مرتبه درونی، میزان اطلاعات بدست آمده کاربر از همسایگان متصل را اندازه‌گیری می‌کند.

$$LD_{in}(i) = \frac{\sum_{j \neq i} \sum_{k \neq i, k \neq j} Rep_j Rep_k \alpha_{ji} \alpha_{ki} \alpha_{jk}}{\sum_{j \neq i} \sum_{k \neq i, k \neq j} Rep_j Rep_k \alpha_{ji} \alpha_{ki}} \quad (14)$$

پس از محاسبه شباهت بین گره‌ها، براساس سیستم ممدانی فازی با شش ورودی از هر یک از ویژگی‌ها: مرتبه ورودی و خروجی، کیفیت مرتبه ورودی و خروجی و چگالی محلی ورودی و خروجی به صورت جداگانه، یک خروجی با مقدار تشابه کلی محاسبه می‌شود. انعکاس شباهت فازی براساس مقدار شباهت کلی با توجه به آگاهی خاص هر گره از مقدار اعتبار سراسری آن، با رابطه (۱۵) محاسبه می‌شود:

$$T_{mir}(u, v) = \begin{cases} Rep(u) & \text{if } sim(u, v) \geq 0.7 \\ Rep(u) * sim(u, v) & \text{if } sim(u, v) < 0.7 \end{cases} \quad (15)$$

نهایتاً نتایج حاصل از روش انعکاس شباهت فازی مبتنی بر اعتماد و عدم اعتماد محاسبه می‌گردد. هدف از انعکاس شباهت فازی، تصمیم‌گیری بر این موضوع است که آیا اعتماد یا عدم اعتماد وجود دارد یا خیر؟. که براساس رابطه (۱۶) محاسبه می‌شود:

$$\begin{cases} Trust & \text{if } T_{mir}(u, v) \geq 0.6 \\ Distrust & \text{if } T_{mir}(u, v) \leq 0.3 \\ Null & \text{otherwise} \end{cases} \quad (16)$$

۳-۴- محاسبه فاکتورهای اعتماد کلی و عدم اعتماد

کلی

در این مرحله ابتدا برای تعیین اعتماد کلی و عدم اعتماد کلی بین دو گره، از پنج ویژگی: تشابه اولویت، تجربه، عمل متقابل، صفات اجتماعی محلی و چگالی اشتراک استفاده می‌شود که به شرح زیر

¹ TD-Trust Mirroring

صفات اجتماعی محلی عبارتند از: زودباوری، بدگمانی، شایستگی، عدم شایستگی و عمل متقابل. این ویژگی ترکیبی از مجموعه ویژگی‌های مرتبط با آن است که با رابطه (۲۳) محاسبه می‌شود:

(۲۳)

$$t(u_i, u_j) = \frac{Mf^+(u_i, u_j) + mf^-(u_i, u_j) + t(u_j, u_i) f^=(u_i, u_j)}{f^+(u_i, u_j) + f^-(u_i, u_j) + f^=(u_i, u_j)}$$

که در آن $f^+(u_i, u_j)$ براساس حاصل جمع زودباوری و شایستگی، $f^-(u_i, u_j)$ براساس حاصل جمع بدگمانی و عدم شایستگی و $f^=(u_i, u_j)$ براساس حاصل جمع عمل متقابل گره u_i نسبت به سایر گره‌ها و عمل متقابل u_j نسبت به سایر گره‌ها محاسبه می‌شود. برای محاسبه مقدار اعتماد و عدم اعتماد براساس این ویژگی‌ها از رابطه (۲۴) استفاده می‌گردد:

$$t(u_i, u_j) = \begin{cases} > 0 & \text{مبتنی بر اعتماد} \\ < 0 & \text{مبتنی بر عدم اعتماد} \end{cases} \quad (24)$$

• چگالی اشتراک

چگالی تعداد دوستان مشترک گره‌های u و v در شبکه اجتماعی است [۲۵]. چگالی اشتراک مبتنی بر اعتماد در رابطه (۲۵) و مبتنی بر عدم اعتماد در رابطه (۲۶) محاسبه می‌شود.

$$I_{u_i, u_j}^+ = D_{C_{u_i}^+ \cap C_v^+} \quad (25)$$

$$I_{u_i, u_j}^- = D_{C_{u_i}^- \cap C_v^-} \quad (26)$$

به منظور محاسبه چگالی داده‌ها، از ضریب خوشه‌بندی به صورت رابطه (۲۷) استفاده می‌شود [۲۵].

$$D_v = \frac{\sum_{i \in C_v} (\sum_{j \in C_v} (M_{ij}))}{(|C_v| * (|C_v| - 1)) / 2} \quad (27)$$

۴- ارزیابی روش پیشنهادی

این بخش شامل معرفی مجموعه داده‌های مورد استفاده برای ارزیابی روش پیشنهادی در بخش ۴-۱، معرفی الگوریتم‌های مورد مقایسه در بخش ۴-۲، و همچنین معیارهای مورد استفاده برای ارزیابی کارایی و عملکرد مدل‌های پیشنهادی در بخش ۴-۳ می‌باشد. علاوه بر این، آزمایش‌های مختلفی بر روی مجموعه داده‌های شبکه‌های اجتماعی در بخش ۴-۴ توصیف می‌شود. نهایتاً در بخش ۴-۵، به بحث و کنکاش روش پیشنهادی می‌پردازد.

۴-۱- توصیف مجموعه داده‌ها

مجموعه داده Epinions شامل هفت جدول تحت عنوان: شباهت، اعتماد، کاربر، نظرات، تجارت، آیت‌ها و دسته‌بندی می‌باشد که

می‌باشند. سپس براساس روابط میانگین وزنی مقدار اعتماد کلی و عدم اعتماد کلی بین دو گره محاسبه می‌شود. این بخش از روش پیشنهادی، اعتماد کلی و عدم اعتماد کلی در شبکه‌های اجتماعی علامت دار^۱ نامیده می‌شود.

• تشابه اولویت

تشابه اولویت نشان می‌دهد که اعتماد کلی ایجاد شده در بین کاربران شبکه براساس شباهت‌ها و تمایزات اجتماعی مانند علاقه به گزینه‌های مشترک یا عنوان‌های^۲ مشترک می‌باشد. تشابه اولویت مبتنی بر اعتماد در رابطه (۱۷) و مبتنی بر عدم اعتماد در رابطه (۱۸) محاسبه می‌شود [۵].

$$Pref_Sim_Trust_{(u_i, u_j)} = \frac{1}{2} \left[\frac{|F_{u_i} \cap F_{u_j}|}{F_{u_i}} + \frac{|NF_{u_i} \cap NF_{u_j}|}{NF_{u_i}} \right] \quad (17)$$

$$Pref_Sim_Distrust_{(u_i, u_j)} = \frac{1}{2} \left[\frac{|F_{u_i} \cap NF_{u_j}|}{F_{u_i}} + \frac{|NF_{u_i} \cap F_{u_j}|}{NF_{u_i}} \right] \quad (18)$$

• تجربه

باعث می‌شود یک کاربر از هر چیزی یا همه چیز مربوط به برخوردهای گذشته خود با سایر کاربران تاثیر بگیرد. تجربه مبتنی بر اعتماد در رابطه (۱۹) و مبتنی بر عدم اعتماد در رابطه (۲۰) محاسبه می‌شود.

$$Trust_Exp(u_i, u_j) = \frac{|S_j^+|}{\max\{|S_i|, |S_j|\}} * \frac{|S_i^+|}{\max\{|S_i|, |S_j|\}} \quad (19)$$

$$Distrust_Exp(u_i, u_j) = \frac{|S_j^-|}{\max\{|S_i|, |S_j|\}} * \frac{|S_i^-|}{\max\{|S_i|, |S_j|\}} \quad (20)$$

• عمل متقابل

اگر دو نفر به یکدیگر اعتماد داشته باشند، نفع و اگر بی‌اعتماد باشند، انتقام افزایش می‌یابد. عمل متقابل بین کاربر u_i و u_j مبتنی بر اعتماد در رابطه (۲۱) و مبتنی بر عدم اعتماد در رابطه (۲۲) محاسبه می‌شود [۵].

$$Trust_Rec(u_i, u_j) = Agr(u_i, u_j) \left(1 - Disagr(u_i, u_j) \right) * Rel(u_i, u_j) \quad (21)$$

$$Distrust_Rec(u_i, u_j) = Disagr(u_i, u_j) \left(1 - Agr(u_i, u_j) \right) * Rel(u_i, u_j) \quad (22)$$

در اینجا، $Rel(u_i, u_j)$ بیانگر قابلیت اطمینان است که بدون داشتن اطلاع یا عدم اطمینان در ارزش متقابل اعتماد بین کاربر u_i و کاربر u_j است. $gr(u_i, u_j)$ مقدار توافق بین کاربر u_i و کاربر u_j است. $Disagr(u_i, u_j)$ اختلاف بین کاربر u_i و کاربر u_j است.

• صفات اجتماعی محلی

² Topics

¹ Signed social networks based on total trust and distrust (SSN-TTD)

اطلاعات محلی-سراسری برای غلبه بر مشکل پراکندگی در شبکه‌های اجتماعی علامت‌دار پیشنهاد شده است.

- قابلیت ارزیابی-شایستگی-عمل متقابل^۴: پیش‌بینی اعتماد و عدم اعتماد را در شبکه‌های اجتماعی علامت‌دار براساس فاکتورهای قابلیت ارزیابی، شایستگی و عمل متقابل انجام می‌دهد.
- ویژگی‌های محلی و سراسری مبتنی بر خوشه‌بندی^۵: یک الگوریتم خوشه‌بندی زیرفضایی برای گروه‌بندی اشیاء اجتماعی است که براساس معیارهای شباهت با تعیین ویژگی‌های جدید مربوط به موقعیت و رفتار دینامیکی گره‌ها انجام می‌شود.

۴-۳- معیارهای ارزیابی

برای اعتبارسنجی کارایی مدل پیشنهادی، از معیارهایی که در ادامه معرفی می‌شوند، استفاده می‌شود که برای محاسبه صحت و مسائل دسته‌بندی مناسب هستند. در ابتدا متغیرهای مورد استفاده در معیارها به صورت زیر تعریف می‌شود.

- صحت^۶: عبارتست از نسبت نمونه‌های صحیح پیش‌بینی شده به کل نمونه‌های موجود:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (28)$$

در اینجا، TP: مثبت درست^۷، TN: منفی درست^۸، FN: منفی غلط^۹ و FP: مثبت غلط^{۱۰} را نشان می‌دهد.

- دقت^۹: عبارتست از نسبت میزان نمونه‌های مثبت پیش‌بینی شده که به درستی تشخیص داده شده‌اند به کل نمونه‌های مثبت:

$$Precision = \frac{TP}{TP + FP} \quad (29)$$

- فراخوانی^{۱۰}: عبارتست از نسبت میزان نمونه‌هایی که به درستی جز نمونه‌های مثبت پیش‌بینی شده به کل نمونه‌های مثبت واقعی:

$$Precision = \frac{TP}{TP + FN} \quad (30)$$

- F-Measure: میانگین هارمونیک بین صحت و فراخوانی را نشان می‌دهد. معمولا مقدار β مثبت بوده و نشان‌دهنده اهمیت فراخوانی نسبت به صحت می‌باشد که معمولا با توجه به اهداف تجربی در اکثر کارها این مقدار را برابر یک در نظر می‌گیرند:

$$F - Measure = \frac{(1 + \beta^2) * Precision * Recall}{\beta^2 * Recall * Precision} \quad (31)$$

ویژگی‌های نظرات در مورد آیتم‌های موجود را برای محاسبه معیارها استفاده می‌کند. آیتم‌ها می‌توانند به صورت محصول یا خدماتی باشند که هر کدام به یک دسته منحصر به فرد تعلق دارند. امتیاز در رتبه‌بندی [۱-۵] می‌باشد که ۱ نشان‌دهنده کمترین و ۵ نشان‌دهنده بالاترین امتیاز است که می‌تواند توسط یک کاربر به یک آیتم داده شود. این مجموعه داده از ۲ فایل تشکیل شده است: ۱- تعداد کاربران، آیتم‌ها، نظرات و رتبه‌بندی‌های ارائه شده توسط کاربران به آیتم‌ها، ۲- وضعیت‌های اعتماد ارائه شده توسط کاربران که در جدول ۲ نشان داده شده است [۲۷].

مجموعه داده Slashdot یک وب سایت خبری اجتماعی است که بر تعاملات و بحث‌های فناوری و جامعه تمرکز دارد. در Slashdot، به جای تشکیل رتبه‌بندی کاربر به آیتم‌ها، رتبه‌بندی‌هایی به صورت کاربران با برچسب گذاری آنها به عنوان "دوست" (رتبه بالا) یا "دشمن" (رتبه پایین) ارائه می‌شوند. هیچ رتبه خنثی در Slashdot وجود ندارد و کاربران می‌توانند دیگران را به عنوان دوست یا دشمن اضافه کنند که به صورت پیوندهای مثبت و پیوندهای منفی در توپولوژی شبکه نمایش داده می‌شوند. آیتم‌ها همان خود کاربران شبکه هستند که رتبه‌بندی را دریافت می‌کنند و کاربران افرادی هستند که حداقل یک امتیاز را می‌دهند که از این ویژگی‌ها به عنوان محاسبه معیارها استفاده می‌شود. تعداد کاربران و آیتم‌ها در جدول ۲ نشان داده شده است [۲۸].

جدول ۲. اطلاعات مجموعه داده

	تعداد آیتم‌ها	تعداد کاربران	تعداد مثلثات
Epinions	۵۰۸۸۳۷	۷۵۸۷۹	۱۶۲۴۴۸۱
Slashdot	۹۴۸۴۶۴	۸۲۱۶۸	۶۰۲۵۹۲

مقایسه‌های روش پیشنهادی و کارهای پیشین بر روی داده‌های یکسان از مجموعه داده‌ها انجام گردیده است.

۴-۲- الگوریتم‌های مورد ارزیابی

برای مقایسه عملکرد و کارایی روش پیشنهادی با سایر روش‌ها، الگوریتم‌های ادامه که برای شبکه‌های اجتماعی علامت‌دار طراحی شده‌اند، معرفی می‌گردند.

- پیش‌بینی پیوند مبتنی بر اطلاعات محلی و سراسری^۱: [۵]: یک مدل محاسباتی فازی مبتنی بر اعتماد و عدم اعتماد است که در دو مدل پیش‌بینی پیوند مختلف مبتنی بر اطلاعات محلی و

⁶ True Negatives (TN)

⁷ False Negatives (FN)

⁸ False Positives (FP)

⁹ Precision

¹⁰ Recall

¹ Local and Global Information based Link Prediction (LGILP)

² Gullibility, Competence, and Reciprocity (GCR)

³ Local and Global Clustering (LGC)

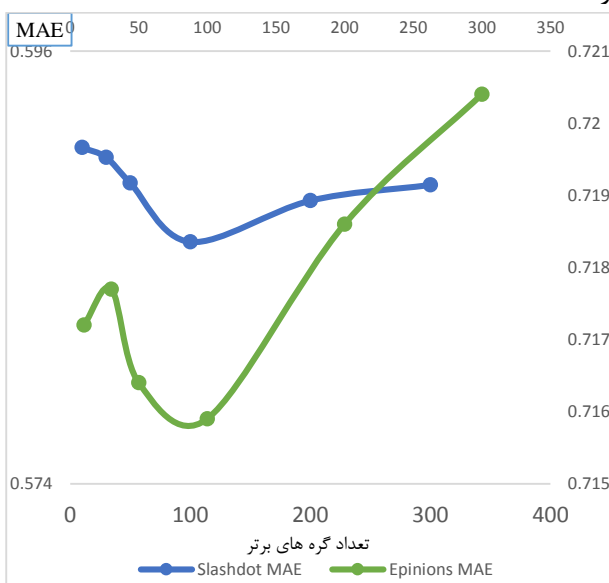
⁴ Accuracy

⁵ True Positives (TP)

موضوع، تعداد گره‌های برتر ۱۰۰ تعیین می‌شود که هم در کارایی موثر هست و هم زمان الگوریتم را نسبت به تعداد بیشتر گره‌های برتر چندان افزایش نمی‌دهد. اگر چه ممکن است مقدار تاثیر بهبود MAE در فرایند روش پیشنهادی بسیار ناچیز می‌باشد، ولی در هر صورت تفکیک نمودن گره‌های برتر و معمولی برای افزایش صحت فرایند تاثیرگذار می‌باشد.

۴-۴-۲- روش انعکاس شباهت فازی اعتماد و عدم اعتماد

در این مرحله با انتخاب ۱۰۰ گره برتر در ابتدای بررسی الگوریتم، نتایج حاصل از روش انعکاس شباهت فازی مبتنی بر اعتماد و عدم اعتماد با ترکیب ویژگی‌های: مرتبه ورودی و خروجی، کیفیت مرتبه ورودی و خروجی و چگالی محلی ورودی و خروجی محاسبه می‌شود. این ۱۰۰ گره برتر ارزش اعتماد بالاتری نسبت به سایر گره‌ها دارند که منجر به افزایش صحت نتایج می‌شود. ویژگی‌های انتخاب شده برای انعکاس شباهت فازی در برنامه‌های کاربردی که نیاز به اعتماد پویا دارند، می‌تواند تقریب قابل قبولی برای اعتماد و عدم اعتماد ارائه دهد.



شکل ۲. تاثیر تعداد گره‌های برتر در رتبه‌بندی گره‌ها با معیار MAE

۴-۴-۳- روش اعتماد کلی و عدم اعتماد کلی

همان طور که در بخش ۴-۴-۱ ذکر شد، تعداد گره‌های برتر برابر ۱۰۰ انتخاب شد. همچنین با توجه به ادبیات موضوع در بررسی اعتماد کلی و عدم اعتماد کلی مشخص شد که با افزایش تعداد کاربران، افزایش صحت در پیش‌بینی اعتماد کلی و عدم اعتماد کلی وجود دارد. در این مرحله، نتایج حاصل از معیارهای مورد بررسی

خطای مطلق میانگین^۱: برای انعکاس میزان خطای پیش‌بینی شده استفاده می‌شود. هرچه مقدار آن کمتر باشد، کارایی پیش‌بینی بیشتر می‌شود:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (32)$$

که در آن y_i و \hat{y}_i ، برچسب کلاس واقعی و برچسب کلاس پیش‌بینی شده را برای نمونه i به ترتیب نشان می‌دهند. M تعداد نمونه‌های پیش‌بینی شده است.

• ریشه میانگین مربعات خطا^۲: برای اندازه‌گیری انحراف بین مقدار واقعی و مقدار پیش‌بینی شده استفاده می‌شود:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (33)$$

۴-۴-۴- ارزیابی نتایج

در این بخش، نتایج آزمایشات بخش‌های مختلف الگوریتم پیشنهادی و نتیجه نهایی ارائه می‌شود. در ابتدا از هر مجموعه داده ۳۰۰۰ نمونه به صورت تصادفی انتخاب می‌شود و آزمایشات بر روی نمونه‌های انتخاب شده صورت می‌گیرد. در جداول این بخش، مقادیر در رتبه اول به صورت پررنگ و زیرخط دار، مقادیر در رتبه دوم به صورت پررنگ و مقادیر در رتبه سوم به صورت خطدار نشان داده می‌شوند.

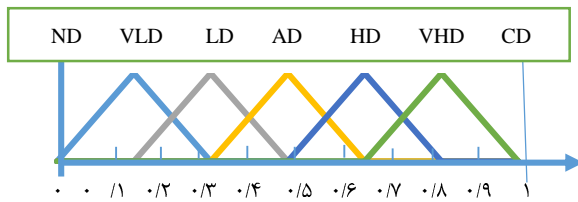
۴-۴-۴-۱- میزان تاثیر انتخاب گره‌های برتر

به طور قطع، استفاده از فرایند انتخاب گره‌های برتر در بررسی اعتماد و عدم اعتماد، کارایی الگوریتم را بهبود می‌بخشد. به این منظور دو ویژگی مهم تاثیرگذار می‌بایست تنظیم شود: ۱- تنظیم پارامتر α و ۲- تعیین تعداد گره‌های برتر. تنظیم پارامتر α ، در رابطه (۷) در بخش ۳-۲، برای کنترل نسبت ترجیح خود و ترجیح دوستان قابل اعتماد می‌باشد. با توجه به نتایج تجربی و بررسی نتایج تحقیق [۶] مقدار α برابر ۰٫۷، تنظیم می‌شود. تعداد گره‌های برتر بر صحت روش پیشنهادی تاثیر می‌گذارد. نتایج حاصل از انتخاب گره‌های برتر در رتبه‌بندی گره‌ها در شکل ۲ برای معیار MAE اندازه‌گیری می‌شود، تعداد گره‌های برتر، برابر ۱۰، ۳۰، ۵۰، ۱۰۰، ۲۰۰ و ۳۰۰ تنظیم می‌شود و نتایج برای دو مجموعه داده Epinions و Slashdot نشان داده می‌شود.

با توجه به نتایج بدست آمده در شکل ۲ می‌توان گفت با کاهش و یا افزایش تعداد گره‌های برتر، مقدار MAE افزایش پیدا می‌کند و همچنین زمان آموزش آنها نیز بیشتر می‌گردد. با توجه به این دو

² Root Mean Square Error (RMSE)

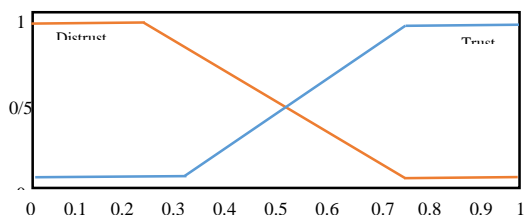
¹ Mean Absolute Error (MAE)



شکل ۴. توابع عضویت انعکاس شباهت فازی عدم اعتماد و عدم اعتماد کلی

تابع عضویت برای متغیر خروجی مبتنی بر اعتماد و عدم اعتماد با دو حد آستانه تعریف می‌شود که براساس آزمون و خطا بدست آمده و بهترین مقدار برای حد آستانه می‌باشد، زیرا در بهبود نتایج تاثیرگذاری بیشتری دارند. تابع عضویت برای متغیر خروجی در شکل ۵ نشان داده شده است.

اگر مقدار قوانین حاصل از خروجی سیستم فازی کمتر از ۰,۳۵ باشد، بین دو گره، عدم اعتماد وجود دارد و اگر مقدار فوق بین ۰,۳۵ تا ۰,۶۵ باشد، احتمال اعتماد و عدم اعتماد برای هر دو گره یکسان است و اگر مقدار ذکر شده بیشتر از ۰,۶۵ باشد، اعتماد بین دو گره وجود دارد. با توجه به توابع عضویت تعریف شده و تعداد ورودی‌ها برای سیستم فازی ممدانی، تعداد قوانین ۴۸۰۲ می‌باشد که با ۱۹۶ قانون می‌توان کل داده‌ها را پوشش داد. تعدادی نمونه از قوانین فازی در جدول ۳ نشان داده شده است.



شکل ۵. تابع عضویت برای خروجی

جدول ۳. تعدادی نمونه از قوانین فازی مدل پیشنهادی

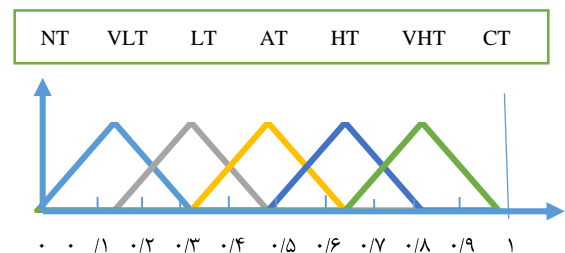
آنگاه (نتیجه اعتماد بین دو گره)	اگر			نتیجه فازی
	عدم اعتماد کلی	اعتماد کلی	انعکاس شباهت فازی عدم اعتماد	
عدم اعتماد	عدم اعتماد متوسط	اعتماد بالا	عدم اعتماد خیلی پایین	اعتماد کامل
اعتماد	عدم اعتماد پایین	اعتماد بالا	عدم اعتماد خیلی پایین	اعتماد خیلی بالا
عدم اعتماد	بدون عدم اعتماد	عدم اعتماد	بدون عدم اعتماد	عدم اعتماد
اعتماد	عدم اعتماد خیلی پایین	اعتماد بالا	بدون عدم اعتماد	اعتماد کامل
عدم اعتماد	عدم اعتماد خیلی پایین	اعتماد متوسط	عدم اعتماد متوسط	اعتماد متوسط
اعتماد	عدم اعتماد خیلی بالا	اعتماد خیلی بالا	عدم اعتماد متوسط	اعتماد خیلی بالا

براساس ترکیب ویژگی‌های اعتماد کلی و عدم اعتماد کلی محاسبه می‌شود. ترکیب این ویژگی‌ها، مسئله پراکندگی بین کاربران در شبکه‌های اجتماعی را با بررسی هم رابطه اعتماد و هم رابطه عدم اعتماد بین کاربران و همسایگان آنها به خوبی حل می‌کند.

۴-۴-۴- تعیین اعتماد و عدم اعتماد کاربران

در این مرحله، کارایی و عملکرد نهایی الگوریتم روش پیشنهادی مبتنی بر اعتماد و عدم اعتماد با سیستم فازی ممدانی که چهار ورودی و یک خروجی دارد، بررسی می‌شود. برای فازی نمودن ورودی از توابع مثلثی و برای دیفازی نمودن از روش مرکز ثقل استفاده می‌شود.

همچنین برای سیستم استنتاج فازی ممدانی از روش حداقل-حداکثر ممدانی استفاده می‌گردد. برای ورودی سیستم فازی ممدانی: انعکاس شباهت فازی اعتماد و عدم اعتماد در بخش ۳-۲ و اعتماد کلی و عدم اعتماد کلی در بخش ۳-۳ و برای خروجی سیستم فازی ممدانی اعتماد و یا عدم اعتماد در نظر گرفته می‌شود. توابع عضویت فازی مثلثی برای متغیرهای زبانی ورودی استفاده می‌شود و مقادیر آنها در بازه [۰, ۱] نرمال می‌شوند. بر این اساس، برای هر ورودی سیستم فازی: اعتماد کلی، عدم اعتماد کلی، انعکاس شباهت فازی اعتماد و انعکاس شباهت فازی عدم اعتماد، یک مجموعه فازی هفت‌تایی مثلثی تعریف می‌شود. مجموعه فازی هفت‌تایی مثلثی اعتماد کلی و انعکاس شباهت فازی اعتماد عبارت است از: بدون اعتماد (NT)، اعتماد خیلی پایین (VLT)، اعتماد پایین (LT)، اعتماد متوسط (AT)، اعتماد بالا (HT)، اعتماد خیلی بالا (VHT) و اعتماد کامل (CT) که در شکل ۳ نشان داده شده است. به طور مشابه، عدم اعتماد کلی و عدم اعتماد در انعکاس شباهت فازی به این صورت فازی می‌شود: بدون عدم اعتماد (ND)، عدم اعتماد خیلی پایین (VLD)، عدم اعتماد پایین (LD)، اعتماد متوسط (AD)، عدم اعتماد بالا (HD)، عدم اعتماد خیلی بالا (VHD) و عدم اعتماد کامل (CD) که در شکل ۴ نشان داده شده است.



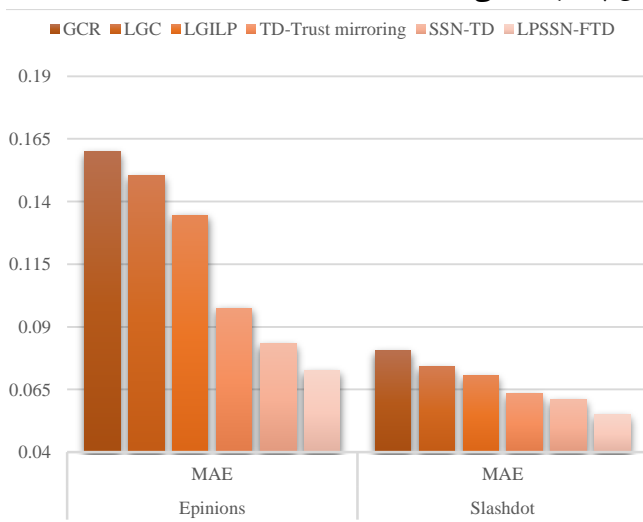
شکل ۳. توابع عضویت انعکاس شباهت فازی اعتماد و اعتماد کلی

جدول ۵. مقایسه روش پیشنهادی با سایر روش‌ها برای عدم اعتماد

کاربران در دو مجموعه داده Epinions و Slashdot

	Epinions				Slashdot			
	صحت	دقت	فراخوانی	F-Measure	صحت	دقت	فراخوانی	F-Measure
LGILP	۰.۹۲۴	۰.۹۲۴	۰.۹۱۶	۰.۹۲۷	۰.۹۴۸	۰.۹۴۸	۰.۹۳۷	۰.۹۴۸
TD-Trust mirroring	۰.۹۶۹	۰.۹۶۹	۰.۹۶۴	۰.۹۸۲	۰.۹۹۶	۰.۹۹۱	۰.۹۷۲	۰.۹۹۱
GCR	۰.۸۶۲	۰.۸۴۳	۰.۸۵۴	۰.۸۴۹	۰.۸۸۷	۰.۸۸	۰.۸۸۴	۰.۸۷۵
LGC	۰.۶۹۹	۰.۶۸۳	۰.۶۷۴	۰.۶۹۵	۰.۸۰۱	۰.۷۹۱	۰.۷۹۶	۰.۸۰۹
SSN-TTD	۰.۹۷۱	۰.۹۷۳	۰.۹۵۳	۰.۹۳۸	۰.۹۹۱	۰.۹۸۷	۰.۹۷۶	۰.۹۹۵
LPSSN-FTD	۰.۹۸۷	۰.۹۷۱	۰.۹۷۸	۰.۹۹۴	۰.۹۹۷	۰.۹۹۴	۰.۹۸۹	۰.۹۹۸

در شکل ۶ نتایج حاصل از مقایسه مقدار MAE و در شکل ۷ نتایج حاصل از مقایسه مقدار RMSE روش پیشنهادی با سایر روش‌ها برای اعتماد کاربران بر روی دو مجموعه داده Epinions و Slashdot نشان داده شده است. همچنین در شکل ۸ نتایج حاصل از مقایسه مقدار MAE و در شکل ۹ نتایج حاصل از مقایسه مقدار RMSE روش پیشنهادی با سایر روش‌ها برای عدم اعتماد کاربران بر روی دو مجموعه داده Epinions و Slashdot نشان داده شده است. نتایج حاکی از آن است که MAE و RMSE روش پیشنهادی نسبت به سایر روش‌ها کمتر بوده و این نشان‌دهنده قابل قبول بودن ساختار روش پیشنهادی می‌باشد.



شکل ۶- مقایسه نتایج MAE روش پیشنهادی با سایر روش‌ها برای اعتماد کاربران در دو مجموعه داده Epinions و Slashdot

نتایج حاصل از جدول ۴ و جدول ۵ حاکی از آن است که روش TD-Trust Mirroring نسبت به روش‌های: LGILP, GCR و LGC در تمامی معیارها و برای هر دو مجموعه داده عملکرد بهتری دارد، پس کارایی قابل قبولی را ارائه می‌دهد. با توجه به اینکه تعداد گره‌های مورد اعتماد و گره‌های مورد عدم اعتماد با احتمال بیشتری مشخص می‌شوند، پس برای افزایش صحت روش TD-Trust Mirroring در تشخیص اعتماد و عدم اعتماد موثر می‌باشند.

همچنین در جدول ۴ و جدول ۵، نتایج نشان می‌دهد که ترکیب ویژگی‌ها در کنار هم براساس انتخاب گره‌های برتر برای روش SSN-TTD نسبت به سایر روش‌های معرفی شده: LGILP, GCR, LGC و TD-Trust Mirroring، برای هر دو مجموعه داده Epinions و Slashdot بهتر و قوی‌تر عمل می‌کند.

نتایج حاصل از میزان اعتماد و عدم اعتماد روش پیشنهادی بین دو کاربر در جدول ۴ و جدول ۵ با عنوان LPSSN-FTD نشان داده شده است. نتایج حاکی از آن است که روش پیشنهادی نسبت به سایر روش‌ها از نظر کارایی برای دو مجموعه داده Epinions و Slashdot، عملکرد بهتری دارد. در جدول ۴ و جدول ۵، میزان صحت، دقت، فراخوانی و F-Measure روش پیشنهادی نسبت به سایر روش‌ها به صورت پررنگ و زیرخط دار نشان داده شده و این حاکی از بهبود نتایج روش پیشنهادی برای هر دو مجموعه داده Epinions و Slashdot می‌باشد. نتایج نشان می‌دهد که الگوریتم پیشنهادی نسبت به سایر روش‌های مورد بررسی برای پراکندگی شبکه قوی‌تر عمل می‌کند.

جدول ۴. مقایسه روش پیشنهادی با سایر روش‌ها برای اعتماد کاربران

در دو مجموعه داده Epinions و Slashdot

	Epinions				Slashdot			
	صحت	دقت	فراخوانی	F-Measure	صحت	دقت	فراخوانی	F-Measure
LGILP	۰.۹۲۴	۰.۹۲۴	۰.۹۱۱	۰.۹۴۲	۰.۹۴۱	۰.۹۴۱	۰.۹۲۱	۰.۹۴۸
TD-Trust mirroring	۰.۹۵۳	۰.۹۵۷	۰.۹۵۹	۰.۹۵۹	۰.۹۴۹	۰.۹۴۸	۰.۹۵۱	۰.۹۴۱
GCR	۰.۸۹۷	۰.۸۸۴	۰.۸۴۷	۰.۸۷۹	۰.۹۲۲	۰.۹۱۹	۰.۹۰۸	۰.۹۱۲
LGC	۰.۷۱۳	۰.۶۹۱	۰.۷۰۲	۰.۷۱۹	۰.۸۲۹	۰.۸۰۸	۰.۸۱۲	۰.۸۳۳
SSN-TTD	۰.۹۵۹	۰.۹۵۹	۰.۹۴۸	۰.۹۴۸	۰.۹۹۸	۰.۹۹۷	۰.۹۷۳	۰.۹۹۹
LPSSN-FTD	۰.۹۹۱	۰.۹۷۸	۰.۹۶۹	۰.۹۹۶	۰.۹۹۸	۰.۹۹۸	۰.۹۸۱	۰.۹۹۹

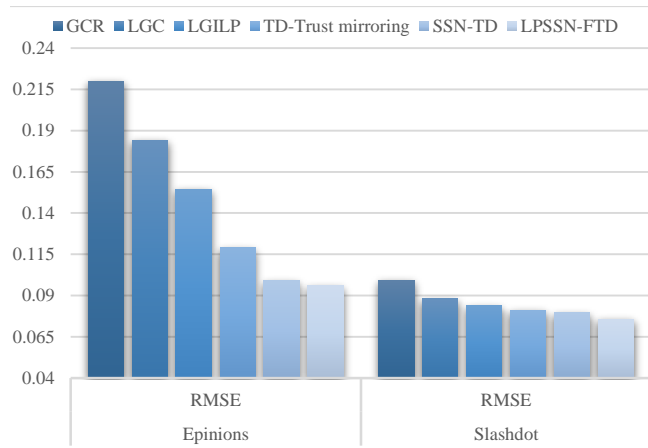
۴-۵- تحلیل نتایج مورد ارزیابی

در این تحقیق، ویژگی‌هایی از نظر صحت در برابر پراکندگی شبکه و سرعت برای الگوریتم پیشنهادی مبتنی بر پیش‌بینی اعتماد و عدم اعتماد با استفاده از محاسبات فازی مورد بحث واقع شد. از آنجا که کاربران در شبکه‌های اجتماعی برای اعتماد و عدم اعتماد به یکدیگر دائماً در حال تغییر و فعالیت هستند، ارائه دقیق‌تر، سریع‌تر و قوی‌تر الگوریتم‌های پیش‌بینی پیوند مبتنی بر اعتماد و عدم اعتماد اهمیت پیدا می‌کند و ضروری می‌باشد. از آنجا که معیارهای سراسری نیاز به پردازش زمان زیادی برای تشخیص اعتماد و عدم اعتماد دارند و معیارهای محلی، زمان پردازش کمتری نیاز دارند و از نظر عملکرد معیارهای سراسری نسبت به معیارهای محلی دارای مقدار صحت بسیار بیشتر می‌باشند، از این رو تلفیقی از این ویژگی‌ها برای تشخیص اعتماد و عدم اعتماد بکار گرفته شد.

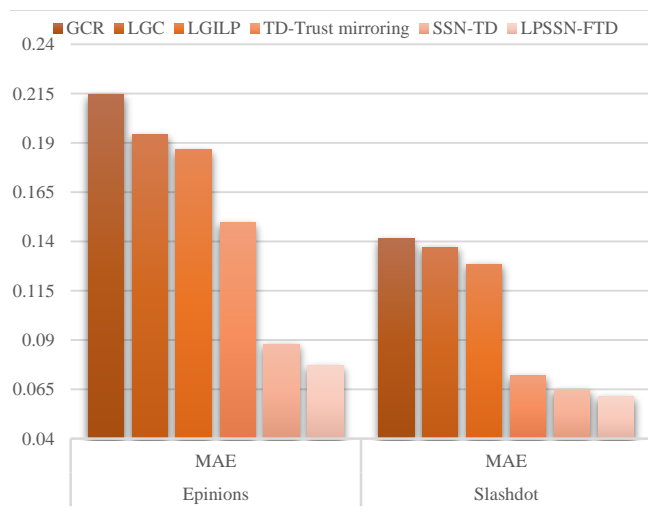
با توجه به ویژگی‌های ذهنی و شخصیتی کاربران، کنار هم قرار دادن این ویژگی‌ها که از نظر اعتماد و عدم اعتماد کاملاً شناخته شده می‌باشند، بر میزان اعتماد و عدم اعتماد به طور قطع تاثیر مثبت می‌گذارد. علت استفاده از این ویژگی‌ها به دلایل زیر می‌باشد:

۱. ویژگی‌های شناخته شده تاثیر بسزایی در ایجاد اعتماد دارند.
۲. ویژگی‌های شناخته شده به راحتی فرموله می‌شوند.
۳. ویژگی‌ها همه به صورت مستقل مورد ارزیابی قرار می‌گیرند و سپس براساس روش میانگین وزنی برای محاسبه اعتماد و عدم اعتماد ترکیب می‌شوند.
۴. استفاده از گره‌های برتر، باعث می‌شود که تا حد زیادی شبکه در مقابل مسئله پراکندگی حفظ گردد و در نتیجه بسیاری از محاسبات برای ویژگی‌ها کاهش یابد.

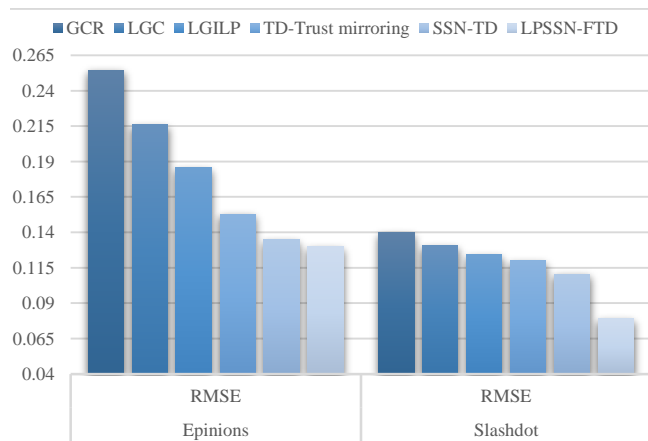
همانطور که در بخش قبل نشان داده شد، با توجه به نتایج جدول ۴ و جدول ۵ می‌توان گفت نتایج حاصل از روش پیشنهادی در دو مجموعه داده Epinions و Slashdot، نسبت به سایر روش‌های معرفی شده در بخش ۴-۲ یعنی LGILP، Trust mirroring و GCR کارایی بهتری دارد و در تمامی معیارها یعنی: دقت، صحت و F-Measure، عملکرد بهتری را ارائه می‌دهد. با بررسی نتایج در شکل‌های ۶، ۷، ۸ و ۹ می‌توان گفت که عملکرد روش پیشنهادی در هر دو مجموعه داده به طور رضایت بخشی پایدار است. در واقع، مقادیر MAE و RMSE برای روش پیشنهادی تحت تأثیر تعداد گره‌های برتر، انعکاس شباهت فازی مبتنی بر اعتماد و عدم اعتماد و همچنین اعتماد کلی و عدم اعتماد کلی منجر به پایداری شبکه‌های اجتماعی و افزایش صحت در پیش‌بینی پیوند مثبت و منفی می‌شود. علاوه بر این، در مقایسه با نتایج چهار الگوریتم دیگر،



شکل ۷- مقایسه نتایج RMSE روش پیشنهادی با سایر روش‌ها برای اعتماد کاربران در دو مجموعه داده Epinions و Slashdot



شکل ۸- مقایسه نتایج MAE روش پیشنهادی با سایر روش‌ها برای عدم اعتماد کاربران در دو مجموعه داده Epinions و Slashdot



شکل ۹- مقایسه نتایج RMSE روش پیشنهادی با سایر روش‌ها برای عدم اعتماد کاربران در دو مجموعه داده Epinions و Slashdot

fuzzy system”, International Journal Machine Learning & Cybernetics, Springer. Vol.9, 2018, PP.1153–1168.

- [4] V.Kant, and K.K.Bharadwaj, Fuzzy computational models of trust and distrust for enhanced recommendations, International Journal of Intelligent Systems, Vol.28, 2013, PP.332–365.
- [5] N.Girdhar, S.Minz, and K.K.Bharadwaj, “Link prediction in signed social networks based on fuzzy computational model of trust and distrust”, Soft Computing, Vol.23, 2019, PP.12123–12138.
- [6] Z.Duan, W.Xu, Y.Chen, and L.Ding, “ETBRec: a novel recommendation algorithm combining the double influence of trust relationship and expert users”, Applied Intelligence, Vol.52, 2022, PP.282–294.
- [7] N.D.Nur, A.H.Sitil, S.Muntadher, S.Firdaus, and N.Anuar, “Applications of link prediction in social networks: A review”, Journal of Network and Computer Applications, Vol.166, 2020, PP.1–31.
- [8] X.Zhu, and Y.Ma, “Sign Prediction on Social Networks Based Nodal Features”, Journal of Complexity, Vol.2020, 2020, PP.1–11.
- [9] R.E.Veras De Sena Rosa, F.A.S.Guimarães, R.d.S.Mendonça and V.F.d.Lucena, “Improving Prediction Accuracy in Neighborhood-Based Collaborative Filtering by Using Local Similarity”, IEEE Access, Vol.8, 2020, PP.142795–142809.
- [10] H.Ghorbanzadeh, A.Sheikhahmadi, M.Jalili, and S.Sulaimany, “A Hybrid Method of Link Prediction in Directed Graphs”, Expert Systems with Applications, Vol.165, 2020, PP.1–13.
- [11] D.Wang, Da-wei, Y.Yih and M.Ventresca, “Improving neighbor-based collaborative filtering by using a hybrid similarity measurement”, Expert Systems with Applications, Vol.160, 2020, PP.1–23.
- [12] Y.Xu, Z.Feng, X.Zhou, M.Xing, H.Wu, X.Xue, Sh.Chen, Ch.Wang and L.Qi, “Attention-based neural networks for trust evaluation in online social networks”, Information Sciences, Vol.630, 2023, PP.507–522.
- [13] X.Wang, Y.Chai, H.Li, and D.Wu, “Link prediction in heterogeneous information networks: An improved deep graph convolution approach”, Decision Support Systems, Vol.141, 2021, PP.113448–113460.
- [14] H.Tahmasbi, M.Jalali, and H.Shakeri, “TSCMF: Temporal and social collective matrix factorization model for recommender systems”, Journal of Intelligence Information Systems, Vol.56, 2021, PP.169–187.
- [15] T.Zhang, W.Li, L.Wang, and J.Yang, “Social recommendation algorithm based on stochastic gradient matrix decomposition in social network”, Journal of Ambient Intelligence and Humanized Computing, Vol.11, 2020, PP.601–608.
- [16] C.He, H.Liu, Y.Tang, S.Liu, X.Fei, Q.Cheng, and H.Li, “Similarity preserving overlapping community detection in signed networks”, Future Generation Computer Systems, Vol.116, 2021, PP.275–290.
- [17] E.Nasiri, K.Berahmand, and Y.Li, “Robust graph regularization nonnegative matrix factorization for link prediction in attributed networks”, Multimedia Tools and Applications, Vol.82, 2023, PP.3745–3768.
- [18] R.I.Yaghi, H.Faris, I.A.Jarah, A.M.Al-Zoubi, A.A.Heidari, and S.Mirjalili, “Link Prediction Using Evolutionary Neural Network Models”, Evolutionary Machine Learning Techniques, Vol.32, 2020, PP.85–112.
- [19] P.Srilatha, R.Manjula, and C.P.Kumar, “Link Prediction on Social Attribute Network Using Lévy Flight Firefly Optimization”, Advances in Artificial Intelligence and Data Engineering, Vol.1133, 2021, PP.1299–1309.
- [20] F.Guo, W.Zhou, Z.Wang, Ch.Ju, Sh.Ji, Q.Lu, “A link prediction method based on topological nearest-neighbors similarity in

روش پیشنهادی در همه موارد بهتر عمل می‌کند.

۵- نتیجه‌گیری

هدف این تحقیق پیش‌بینی پیوند در شبکه‌های اجتماعی علامت‌دار برای بررسی میزان اعتماد و عدم اعتماد بین کاربران شبکه می‌باشد. روابط کاربران در شبکه‌های اجتماعی علامت‌دار مشابه روابط کاربران در دنیای واقعی بوده و از نظر اعتماد و عدم اعتماد روابط یکسانی بین کاربران وجود ندارد. در واقع، اینکه کاربران در دنیای مجازی به اندازه دنیای واقعی چقدر به دیگران اعتماد و یا عدم اعتماد داشته باشند، ضروری می‌باشد. بعضی از روش‌های پیش‌بینی پیوند بخصوص در شبکه‌های اجتماعی بدون علامت، از ناتوانی در مدیریت صحیح اعتماد بین ارتباط کاربران، ضعف پیش‌بینی عدم اعتماد، کارایی پایین و عدم عملکرد مناسب در شبکه‌های پراکنده رنج می‌برند. از این رو، یک مدل پیش‌بینی پیوند مبتنی بر اعتماد و عدم اعتماد براساس محاسبات فازی در شبکه‌های اجتماعی علامت‌دار پیشنهاد می‌شود که از قوانین فازی برای تعیین اعتماد و عدم اعتماد بین کاربران استفاده می‌کند. نتایج آزمایش‌های انجام شده بر روی مجموعه داده‌های Epinion و Slashdot حاکی از آن است که روش پیشنهادی به خوبی می‌تواند میزان اعتماد و عدم اعتماد کاربران را تشخیص دهد و از دقت بالایی برخوردار است.

کارهای آینده عمدتاً برای حل مسائل پیش‌بینی پیوند در شبکه‌های اجتماعی علامت‌دار براساس ویژگی‌های مختلف اطلاعات کاربران و استفاده از روش‌های متفاوت برای تعیین میزان صحت پیش‌بینی خواهد بود. بسیاری از مدل‌های موثر پیش‌بینی پیوند برای دستیابی به نتایج پیش‌بینی پیوند مثبت و منفی استفاده می‌شود. سعی خواهد شد چنین اطلاعاتی برای افزایش اطلاعات پیوند پیش‌بینی شده تولید شود و موضوعاتی مانند ویژگی‌های زمانی، انتشار اعتماد، ویژگی‌های اجتماعی و ویژگی‌های پویا بررسی گردد. همچنین استفاده از سایر ویژگی‌های محلی و سراسری در تشخیص اعتماد منجر به قوی شدن و دقیق شدن بررسی اعتماد و عدم اعتماد در شبکه‌های اجتماعی علامت‌دار می‌شود.

مراجع

- [1] X.Li, H.Fang, and J.Zhang, “FILE: A novel framework for predicting social status in signed networks”, Thirty-Second AAAI Conference on Artificial Intelligence, AAAI18 - Artificial Intelligence and the Web, Vol.32, No.1, 2018, PP.330–337.
- [2] K.Akilal, M.Omar, and H.Slimani, “Characterizing and using gullibility, competence, and reciprocity in a very fast and robust trust and distrust inference algorithm for weighted signed social networks”, Knowledge-Based Systems, Vol.192, 2020, PP.1–11.
- [3] H.Shirgahi, M.Mohsenzadeh, and H.H.S.Javadi, “A new method of trust mirroring estimation based on social networks parameters by

- time rumor refutations in Online Social Networks", *Expert Systems with Applications*, Vol.212, 2023, PP.118638-118670.
- [25] S.Ghasemi, and A.Zarei., "Improving link prediction in social networks using local and global features: a clustering-based approach, *Progress in Artificial Intelligence*", Vol.11, 2022, PP.79–92.
- [26] Suryakant, and T.Mahara, A New Similarity Measure Based on Mean Measure of Divergence for Collaborative Filtering in Sparse Environment, *Procedia Computer Science*, Vol.89, 2016, PP.450–456.
- [27] <https://snap.stanford.edu/data/soc-Epinions1.html>, Last Visited (01, October. 2022).
- [28] <http://snap.stanford.edu/data/soc-Slashdot0902.html>, Last Visited (01, October.2022).
- [29] J.Golbeck, "Combining provenance with trust in social networks for semantic content filtering", *International Provenance and Annotation Workshop*, Vol.4145, 2006, PP.101–108.
- directed networks", *Journal of Computational Science*, Vol.69, 2023, PP.102002-102016.
- [21] M.Nooraie.Abadeh, M.Mirzaie, "A differential machine learning approach for trust prediction in signed social networks", *Supercomput*, Vol.79, 2023, PP.9443–9466.
- [22] R.E.Tillman, P.Vamsi, Ch.Jiahao, R.Prashant and M.Veloso, "Heuristics for Link Prediction in Multiplex Networks", *In Proceedings of ECAI'2020, 24th European Conference on Artificial Intelligence*, Vol.325, 2020, PP.1938-1945.
- [23] H.Liu, Z.Zhenzhen, B.Fan, H.Zeng, Y.Zhang, and G.Jiang, "PrGCN: Probability prediction with graph convolutional network for person re-identification", *Neurocomputing*, Vol.423, 2021, PP.57-70.
- [24] X.Hu, X.Xiong, Y.Wu, M.Shi, P.Wei, and Ch.Ma, "A Hybrid Clustered SFLA-PSO algorithm for optimizing the timely and real-

Fault Diagnosis and Detection in Photovoltaic Systems Using Neural Network VGG16

Samaneh Azimi¹, Mohammad Manthouri^{2*}, Mahdi Akhbari³

¹ Department of Electrical Engineering, Shahed University, Tehran, Iran

² Department of Electrical Engineering, Shahed University, Tehran, Iran

³ Department of Electrical Engineering, Shahed University, Tehran, Iran

Received: 14 March 2023, Revised: 17 October 2023, Accepted: 27 November 2023

Paper type: Research

Abstract

Fault detection in photovoltaic (PV) arrays is necessary to increase the output power and also the useful life of a PV system. The presence of conditions such as partial shade, high impedance faults, and the maximum power point detector (MPPT) system make the fault detection of PV in environmental conditions more challenging. The literature identified and classified defects just in few scenarios. In this study two-dimensional scalograms are generated from PV system data. The VGG16 as a pre-trained convolutional neural network is used for feature extraction. Finally, to identify and classify faults in the PV system a fully connected neural network is trained. Unlike the previous methods proposed in the literature on the subject of defect detection and classification, various defective cases with MPPT combination are considered in this research. It has been shown that the proposed method including pre-trained CNN performs better than the existing methods and achieves an error detection accuracy of 83.375%.

Keywords: Photovoltaic array, Maximum power point tracking, Fault classification, Convolutional neural network, Scalograms

* Corresponding Author's email: mmanthouri@shahed.ac.ir

تشخیص و شناسایی خطا در سیستم‌های فتوولتائیک با استفاده از شبکه عصبی عمیق VGG16

سمانه عظیمی^۱، محمد منثوری^{۲*}، مهدی اخباری^۳

^۱ گروه قدرت، دانشکده فنی مهندسی، دانشگاه شاهد، تهران، ایران

^۲ گروه کنترل، دانشکده فنی مهندسی، دانشگاه شاهد، تهران، ایران

^۳ گروه قدرت، دانشکده فنی مهندسی، دانشگاه شاهد، تهران، ایران

تاریخ دریافت: ۱۴۰۱/۱۲/۲۳ تاریخ بازبینی: ۱۴۰۲/۰۷/۲۵ تاریخ پذیرش: ۱۴۰۲/۰۹/۰۶

نوع مقاله: پژوهشی

چکیده

تشخیص خطا در آرایه‌های فتوولتائیک (PV) برای افزایش توان خروجی و همچنین طول عمر مفید یک سیستم PV ضروری است. وجود شرایطی مانند سایه جزئی، خطاهای امیدانس بالا و وجود سامانه ردیاب نقطه حداکثر توان (MPPT)، تشخیص خطا را در شرایط محیطی به چالش می‌کشد. بیشتر تحقیقات انجام‌شده در این زمینه فقط در چند سناریو از عیوب به شناسایی و طبقه‌بندی پرداخته‌اند و اکثراً نیاز به تجهیزات پیشرفته (تصویرساز حرارتی، پهپادها) برای اکتساب داده دارند که روش اکتساب داده این پژوهش به وسیله اسکالوگرام‌های دوبعدی این نیاز را مرتفع کرده است. این پژوهش با استفاده از شبکه‌ی عصبی کانولوشنی عمیق از پیش آموزش داده شده VGG16 و با بهره‌گیری از ویژگی‌های استخراج شده به وسیله اسکالوگرام‌های دوبعدی تولیدشده از داده‌های سیستم PV، به شناسایی و طبقه‌بندی خطا در سیستم PV با استفاده از یک شبکه عصبی کاملاً متصل می‌پردازد. برخلاف روش‌های قبلی پیشنهادشده در ادبیات موضوع تشخیص و طبقه‌بندی عیوب، موارد مختلف معیوب همراه با ترکیب MPPT در مطالعه ما در نظر گرفته شده است. در این تحقیق نشان داده شده است که روش پیشنهادی شامل CNN از پیش آموزش دیده تنظیم شده، از روش‌های موجود بهتر عمل می‌کند و به دقت تشخیص خطای ۸۳/۳۷۵ درصد دست پیدا می‌کند.

کلیدواژه‌گان: آرایه فتوولتائیک، ردیاب نقطه حداکثر توان، طبقه‌بندی خطا، شبکه عصبی کانولوشنی VGG16، اسکالوگرام.

* رایانامه نویسنده مسؤول: mmanthouri@gmail.com

۱- مقدمه

آثار نامطلوب منابع انرژی فسیلی بر سلامت انسان‌ها، تغییرات اقلیمی و اکوسیستم به دنبال رشد صنعتی در سراسر جهان و افزایش مصرف جهانی این نوع منابع لزوم گسترش بهره‌گیری از منابع انرژی‌های تجدیدپذیر را افزایش می‌دهد زیرا این منابع تمیز، پایدار و سازگار با محیط‌زیست هستند. یکی اقتصادی‌ترین منابع انرژی تجدیدپذیر تجاری، انرژی فتوولتائیک خورشیدی است. مزایای متعدد این صنعت نظیر عملکرد عاری از صدا و منبع انرژی با قابلیت دسترسی جهانی سبب توسعه سیستم‌های فتوولتائیک شده است.

بهره‌گیری از سیستم‌های خورشیدی به دلیل کاهش هزینه‌های پنل‌ها و سیاست دولت‌ها به نفع مصرف‌کنندگان به سرعت در حال افزایش است [۱]. تولید برق به میزان ۵۸۰۰ تراوات ساعت تا سال ۲۰۲۵ با توجه به نرخ بهره‌گیری از سیستم‌های فتوولتائیک قابل پیش‌بینی است [۲]. علیرغم مزایا و در دسترس بودن رایگان، صنعت PV به دلیل حضور در شرایط محیطی سخت و وابستگی خطاها به این شرایط با چالش‌هایی از جمله کاهش قدرت خروجی، قابلیت اطمینان و آسیب‌پذیری مواجه است [۳]. با توجه به حضور سیستم‌های PV در محیط بیرونی، امکان وقوع خطاهایی از جمله مدار باز (OC)، خط به خط (LL)، خط به زمین (LG)، نقطه داغ (HS)، افت سیم‌کشی و اثرات زیست‌محیطی (سایه زدن یکنواخت و غیریکنواخت، رطوبت، جمع شدن برف و گردوغبار) در این سیستم‌ها افزایش می‌یابد. این خطاها ممکن است سبب کاهش کارایی و طول عمر PVS شود [۴] و [۵].

پایش سیستم‌های فتوولتائیک به منظور تشخیص به موقع خطاها جهت اطمینان از طول عمر تجهیزات، حداکثر بازده توان و اطمینان از عملکرد قابل اعتماد این سیستم‌ها در مقایسه بزرگ امری ضروری است. روش‌های مختلف نظارت و تشخیص خطا در سرعت، پیچیدگی و سنسورهای موردنیاز و توانایی شناسایی تعداد خطاها متفاوت است. بهره‌گیری از دستگاه‌های محافظتی در قانون ۶۹۰ NCE National Electric Code مانند دستگاه محافظت در برابر خطای زمین (GFPD)^۱، دستگاه محافظت از جریان بیش‌ازحد (OCPD)^۲ و مدار قطع‌کننده خطا قوس (AFCI)^۳ برای بررسی و کاهش انواع خطاها [۶-۹] خود با چالش‌هایی نظیر تابش کم، وجود دیوده‌های مسدودکننده، ردیاب نقطه حداکثر توان (MPPT)^۴ و

ویژگی‌های غیرخطی آرایه‌های PV که مانع از عملکرد صحیح دستگاه‌های محافظت‌کننده‌ی خطا در شرایط مختلف می‌شوند مواجه هستند. از این رو، ممکن است خطاها برای مدت طولانی تداوم‌یافته و قابل‌شناسایی نباشد [۱۰]. بنابراین بهره‌گیری از روش‌های تشخیص عیوب پیشرفته، کارآمد و سازگار با چالش‌های سیستم‌های فتوولتائیک جهت تشخیص به موقع عیوب، افزایش راندمان نیروگاه و طول عمر مفید سیستم امری اجتناب‌ناپذیر است. علاوه بر طرح‌های حفاظتی، روش‌های پیشرفته تشخیص عیب ارائه شده است تا با تشخیص به موقع خطاها، سیستم‌های فتوولتائیک از حفاظت قابل اطمینان برخوردار شوند. اثربخشی روش‌های حفاظت پیشرفته به دلیل خطاهای امپدانس بالا، شرایط آب و هوایی و زمانی که MPPT در حال کار است چالش‌برانگیز است [۱۰]. همچنین روش‌های تشخیص خطا عموماً به تشخیص تعداد محدود از خطاها پرداخته‌اند. به طور خاص [۱۱] به تشخیص فقط یک خطای زمین پرداخته است. همچنین روش پیشنهادی در [۱۲] فقط شامل دو سناریوی خطا در آرایه PV (LLF) و (PS) می‌باشد. طبقه‌بندی چهارخطا (LLF, OCF, SF, BDF)^۵ در [۱۳] مورد بررسی قرار گرفته است.

برای پرداختن به چالش‌های فوق و کمبودهای کار موجود در تشخیص خطا، این مقاله با بهره‌گیری از رویکرد شبکه‌های عصبی کانولوشنی دوبعدی عمیق CNN و با بهره‌گیری از تصاویر اسکالوگرام‌های دوبعدی PV به منظور شناسایی و طبقه‌بندی شش خطای ایجادشده و هم‌زمانی هر شش خطا در سیستم PV پرداخته است. برخلاف تحقیقات جاری، موارد مختلف خطا (از جمله هم‌زمانی انواع خطا، شرایط سایه جزئی، خطای تخریب، خطای زمین، خطا در جعبه اتصال، خطای خط به خط، خطای مدار باز و تجمیع خطاها) با ترکیب MPPT در نظر گرفته شده است. نتایج نشان می‌دهند که روش پیشنهادی این مقاله Fine-Tuned VGG-16 CNN از روش‌های موجود از دقت طبقه‌بندی خطای بالاتری برخوردار است و همچنین قادر به شناسایی تعداد بیشتر خطا چه در سطح آرایه و چه در سطح مازول فتوولتائیک است. همچنین برای استخراج تصاویر به‌عنوان ورودی شبکه عصبی کانولوشنی عمیق نیاز به تجهیزات جانبی مانند تجهیزات تصویربرداری حرارتی ندارد بنابراین روش پیشنهادی نیاز به داده‌های ورودی تصویری برای بهره‌گیری انواع روش تشخیص خطا مبتنی بر یادگیری عمیق را مرتفع

⁴ Maximum Power Point Tracker

⁵ Line To Line Fault, Open Circuit Fault, Shading Fault

¹ Ground Fault Protection Device

² Over Current Protection Device

³ Fault Circuit Interrupter

می‌سازد. لذا قادر به بهره‌گیری مستمر در طول شبانه‌روز در یک سیستم خورشیدی بزرگ می‌باشد.

۲- پیشینه پژوهش

روش‌های تشخیص خطای متعددی برای شناسایی و تشخیص خطا در سیستم‌های PV پیشنهاد شده‌اند و طبقه‌بندی‌های متفاوتی برای انواع روش تشخیص خطا با توجه به رویکرد، نیاز حسگر، توانایی تشخیص خطاها، پیچیدگی یکپارچه‌سازی، دقت، هزینه پیاده‌سازی، چالش‌های حفاظتی و محدودیت‌های دستگاه‌های حفاظتی معمولی ارائه شده است.

در تشخیص عیب مبتنی بر یادگیری ماشین، الگوریتم‌های یادگیری ماشین برای یادگیری رابطه بین پارامترهای ورودی و خروجی یک سیستم PV آموزش داده می‌شوند. داده‌های آموزشی را می‌توان به صورت تجربی یا با کمک مدل‌های PV دقیق جمع‌آوری کرد. با این حال، داده‌های غیرعادی در هنگام وقوع خطا نیز برای آموزش و پیش‌بینی دقیق مورد نیاز است [۷].

در ادامه تعدادی از منابعی که با استفاده از روش‌های یادگیری عمیق به تشخیص خطا پرداخته‌اند، بررسی شده و در انتها به صورت خلاصه در جدول ۱ ارائه شده است.

در [۱۴] توسعه یک شبکه عصبی مصنوعی چند خروجی (ANN) را برای تشخیص و شناسایی خطا در سمت DC یک سیستم فتوولتائیک (PVS) ارائه شده است. روش توسعه یافته با استفاده از خروجی طبقه‌بندی، فقط قادر به تشخیص خطاهای مدار باز، اتصال کوتاه و عدم تطابق است.

در [۱۵] خطای خط به خط، خطای زمین و خطای اتصال کوتاه با استفاده از شبکه‌ی عصبی (Bi-LSTM) قابل شناسایی است.

تغییرات در سطح ماژول مانند سوختگی، لایه‌لایه شدن و شکستگی با استفاده از تصاویر هوایی ثبت شده توسط پهپادها و با بهره‌گیری از روش ترکیبی شبکه‌ی عصبی کانولوشنی و (KNN) در [۱۶] تشخیص داده شده است.

در [۱۷] سه نوع خطای زمین، خرابی دیود بای پس و خطای سایه جزئی با استفاده از CNN مبتنی بر ResNet قابل تشخیص است.

در [۱۸] ماژول‌های دارای شکستگی و لایه‌لایه شدن با استفاده از تصاویر هوایی که توسط اینترنت اشیا ارسال و در فضای ابری ثبت

شده‌اند، با استفاده از CNN تشخیص داده می‌شوند.

در مقاله [۱۹] با استفاده از تصاویر ترموگرافی و با بهره‌گیری از شبکه‌ی عصبی CNN ماژول‌های شکسته و دچار لایه‌لایه شدن تشخیص داده شد.

در [۲۰] برای تشخیص ترک بر روی ماژول از تصاویر الکترو لومینسانس^۱ و شبکه عصبی عمیق استفاده شده است.

الگوریتم تشخیص خطای مدار باز، خطای اتصال کوتاه و خطا عدم تطابق با ترکیب شبکه‌ی عصبی CNN و SVM در [۲۱] بیان شده است که با دقت خوبی این سه خطا طبقه‌بندی شده‌اند.

جدول ۱. خلاصه پیشینه‌ی تحقیق

منبع	روش تشخیص خطا	خطاهای تشخیص داده شده
[۱۴]	ANN	خطاهای مدار باز، اتصال کوتاه و عدم تطابق
[۱۵]	Bi-LSTM	خطای خط به خط، خطای زمین و خطای اتصال کوتاه
[۱۶]	CNN & KNN	سوختگی، لایه‌لایه شدن و شکستگی
[۱۷]	CNN	خطای زمین، خرابی دیود بای پس و خطای سایه جزئی
[۱۸]	CNN	شکستگی و لایه‌لایه شدن
[۱۹]	CNN	شکستگی و لایه‌لایه شدن
[۲۰]	DL	تشخیص ترک
[۲۱]	CNN & SVM	خطاهای مدار باز، اتصال کوتاه و عدم تطابق

۳- انواع خطا در آرایه‌ی PV

عواملی که باعث کاهش خروجی یک ماژول یا آرایه PV می‌شوند به عنوان خطا در نظر گرفته می‌شوند. خطاها ممکن است موقت یا دائمی باشند. خطاها در یک سیستم PV می‌توانند به دلایل زیادی رخ دهند، برخی از عواملی که باعث اتلاف انرژی در سیستم‌های PV می‌شوند عبارتند از: تغییر سطح تابش خورشیدی و دمای عملکرد، عدم تطابق در پارامترهای ماژول PV، خطا در ردیابی نقطه حداکثر توان، اثرات سایه، پیری، تجمع برف و گردوغبار روی ماژول‌های PV و خطاهای فاجعه‌بار [۲۲-۲۵]. لذا بینش عمیق برای (۱) تشخیص هر رخداد خطا، (۲) درک ضرورت تشخیص عیب و (۳) تجزیه و تحلیل چالش‌های حفاظتی ضروری است [۲۶]. از این رو برای درک بهتر، تجزیه و تحلیل دقیقی از خطاهای مختلف در این بخش ارائه شده است.

¹ Electroluminescence

۳-۱- خطای تخریب

مهم‌ترین حالت‌های خطای تخریب عبارت‌اند از: شکست جعبه اتصال، شکستن شیشه، اتصال سلولی معیوب، قاب شل و لایه‌لایه شدن. این حالت‌های تخریب در آب‌وهوای گرم و مرطوب بیشتر گزارش شده است [۷]. ورود آب یا رطوبت در یک ماژول PV با شیشه شکسته منجر به تنش خوردگی قابل توجهی می‌شود و باعث تغییر در مقاومت‌های پارازیتی می‌شود [۲۷].

۳-۲- خطای زمین

این پدیده به‌طور تصادفی توسط یک مدار اتصال کوتاه الکتریکی رخ می‌دهد. در واقع، این خطا یک مسیر با امپدانس کم غیر عمدی در میان یکی از هادی حامل جریان و زمین است [۲۸]. خطای زمین اغلب نگرانی‌های ایمنی را در بردارد.

۳-۳- خطای جعبه اتصال

قابلیت اطمینان در جعبه اتصال یکی از حیاتی‌ترین مسائل برای PVM در طول تست تأیید و عملیات در این زمینه است. جعبه اتصال محافظی برای سیم‌کشی از رشته‌ها به پایانه‌های خارجی است. این حالت خرابی می‌تواند ناشی از تثبیت ضعیف جعبه اتصال به صفحه پشتی، نفوذ رطوبت، خوردگی اتصالات، سیم‌کشی ضعیف منجر به قوس داخلی، نصب ضعیف یا به دلیل تخریب حرارتی باشد [۲۹].

۳-۴- خطای خط به خط

خطای خط به خط یک اتصال با مقاومت کم بین دونقطه با پتانسیل مختلف در یک شبکه الکتریکی یا سیستم است. از عوامل ایجادکننده این خطا می‌توان به شکست الکتریکی کابل‌ها، اتصال کوتاه اتفاقی بین هادی‌های انتقال جریان، عایق کم بین اتصالات رشته در جعبه اتصال و آسیب مکانیکی اشاره نمود [۶].

۳-۵- خطای مدارباز

این خطا زمانی اتفاق می‌افتد که یکی از مسیرهای حمل جریان در مسیرهای سری با بار شکسته یا باز شود. از عوامل به وجود آورنده این خطا می‌توان به اتصال ضعیف بین سلول‌ها، قطع و وصل کردن کانکتورها در جعبه اتصال و پارگی سیم‌ها اشاره نمود. وجود خطاهایی نظیر نقطه سوز و خرابی در دیودها نیز از عوامل ایجاد خطای مدارباز به شمار می‌روند [۲۸].

۳-۶- سایه و نیم‌سایه

به‌طور کلی دو نوع سایه در سیستم‌های فتوولتائیک تعریف می‌شود. نوع اول سایه سخت که در صورتی رخ می‌دهد که ماژول‌های PV با مواد جامد، به‌عنوان مثال سایه ساختمان‌ها یا گردوغبار، سایه‌دار شده نوع دوم سایه نرم است که ناشی از عناصر موجود در هوا مانند دوده رخ می‌دهد. سایه سخت منجر به کاهش ولتاژ می‌شود؛ زیرا سایه‌دار شدن تعدادی از ماژول‌هایی سری منجر به عدم دریافت تابش یکسان در ماژول‌ها شده و این عامل سبب کاهش جریان اتصال کوتاه ماژول‌های سایه‌دار می‌شود. ماژول‌هایی که سایه‌دار نشده‌اند جریان اتصال کوتاه بالاتری دارند و می‌خواهند جریان تولیدی بیشتر را عبور دهد سری بودن این ماژول‌ها سبب می‌شود که ماژول‌های سایه‌دار به نقطه‌های اتسپات خود نزدیک شوند. درحالی‌که سایه نرم روی جریان تأثیر می‌گذارد. در صورت عدم رفع سایه شاهد کاهش توان تولیدی و آسیب سلول‌های سایه‌دار خواهیم بود [۶].

۴- شبیه‌سازی آرایه فتوولتائیک

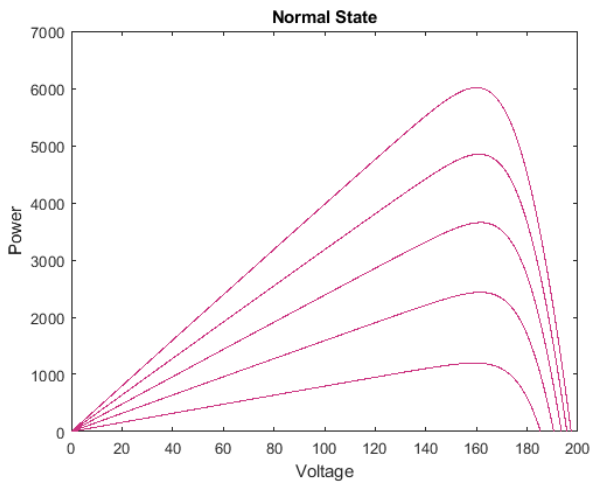
با توجه به رشد سریع در منابع انرژی تجدیدپذیر به‌خصوص درزمینه‌ی سیستم‌های خورشیدی، طراحان PV به یک ابزار قابل‌اطمینان و انعطاف‌پذیر برای پیش‌بینی تولید برق توسط سیستم‌های PV در اندازه‌های مختلف نیاز دارند. مدل PV خورشیدی به‌طور بی‌وقفه به‌روز می‌شود تا به محققان برای درک بهتر عملیات کمک کند. بیشتر مدل‌های ریاضی بر اساس روابط ولتاژ-جریان مبتنی بر مدل تک دیودی یا دو دیودی گسترش یافته‌اند.

۴-۱- مدل ریاضی ماژول فتوولتائیک

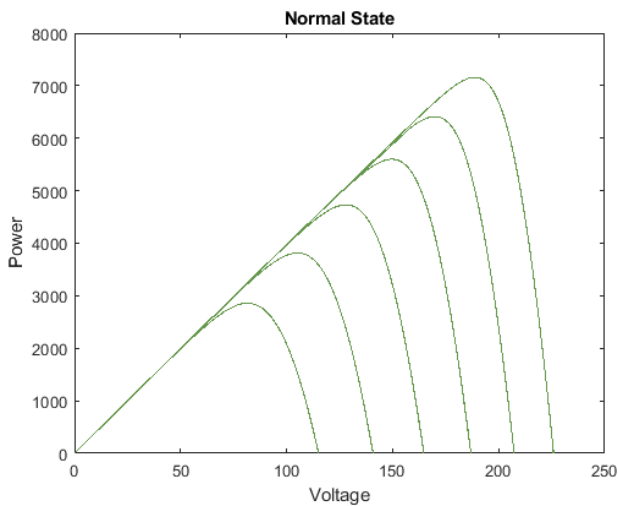
ماژول فتوولتائیک از سلول‌های فتوولتائیک تشکیل شده است که به‌طور موازی به هم متصل می‌شوند. این سلول اساساً یک اتصال p-n است که از نیمه رسانایی مانند سیلیکون ساخته شده است که امکان تبدیل انرژی نور به برق را فراهم می‌کند. همان‌طور که در شکل ۱ نشان داده شده است. یک سلول PV ایده آل به‌صورت مدار تک دیودی مدل می‌شود [۳۰].

مدار معادل که شامل منبع فتوجریان، دیود، مقاومت موازی بیان‌کننده جریان شنت و مقاومت سری است که مقاومت داخلی در برابر جریان را توصیف می‌کند. معادله مشخصه I-V یک سلول PV با روابط (۱) تا (۴) تعیین می‌شود [۳۰].

ولتاژ مدار باز ۳۹/۵

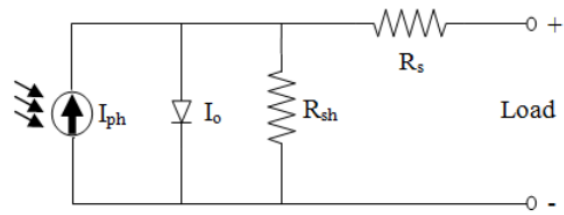


شکل ۲. نمودار P_V در حالت تابش متغیر-دما ثابت



شکل ۳. نمودار P_V در حالت تابش ثابت-دما متغیر

شکل ۴ پیکربندی سیستم PV را در حالت‌های خطای مختلف را نشان می‌دهد. سیستم فتوولتائیک طراحی شده از یک آرایه ۴×۵ با ۲۰ ماژول ۳۰۰ وات با توان خروجی ۶ کیلووات تشکیل شده است. شکل ۵ مدل سیمولینک کل سیستم PV را با استفاده از مدل تک دیود مرجع مطابق با مرجع [۳۰] نشان می‌دهد که شامل مبدل تقویت‌کننده DC-DC با MPPT برنامه‌ریزی شده با الگوریتم اغتشاش و مشاهده (P&O) می‌باشد.



شکل ۱. مدار معادل سلول PV [۲۸]

جریان نوری:

$$I_{ph} = [I_{SC} + I_{co}(T_c - T_{rf})] \frac{I_r}{I_{rf}} \quad (1)$$

جریان اشباع معکوس در دمای مرجع:

$$I_{rs} = I_{sc} / [\exp\left(\frac{qV_{oc}}{N_s K A T_c}\right) - 1] \quad (2)$$

جریان اشباع دیود:

$$I_{s=I_{rs}} \left[\frac{T_c}{T_{rf}} \right]^3 \exp\left[\frac{qE_g}{AK} \left(\frac{1}{T_{rf}} - \frac{1}{T_c} \right) \right] \quad (3)$$

جریان ماژول PV:

$$I = N_p I_{ph} - N_p I_s \left[\exp\left\{ \frac{qV + IR_s}{N_s K A T_c} \right\} - 1 \right] - \frac{V + IR_s}{R_p} \quad (4)$$

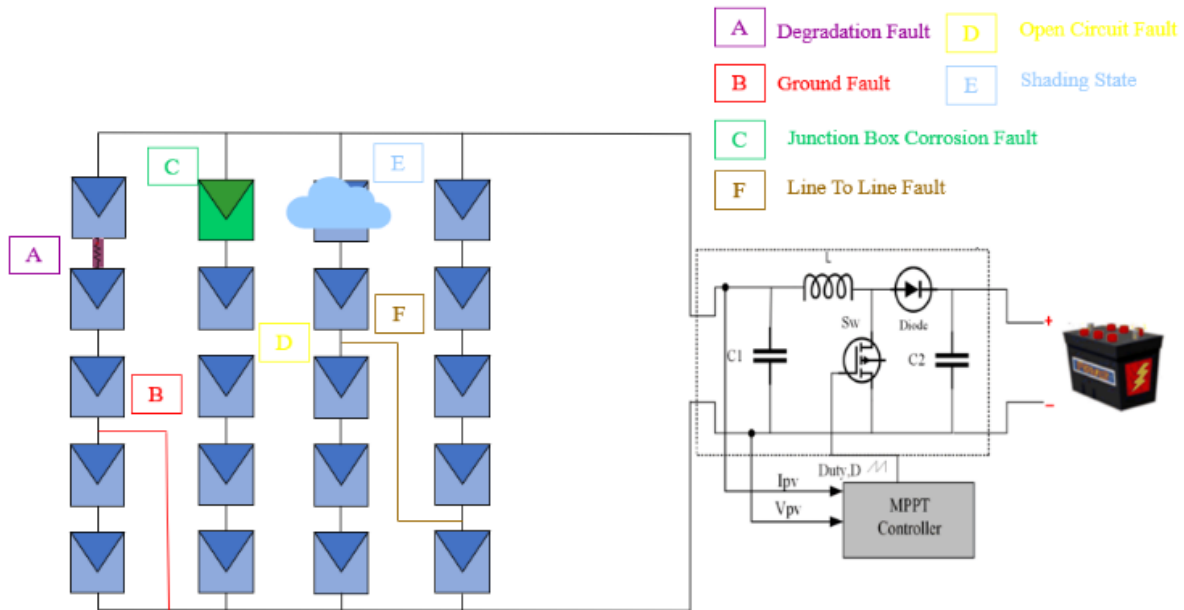
۲-۴-۲ مدل سازی در محیط شبیه‌ساز متلب

پس از مدل کردن سیستم فتوولتائیک در محیط شبیه‌ساز متلب برای اطمینان از عملکرد صحیح سیستم مطابق جدول ۲ مقادیر پارامترهای ماژول PV تحت شرایط تست در حالت استاندارد استخراج شد.

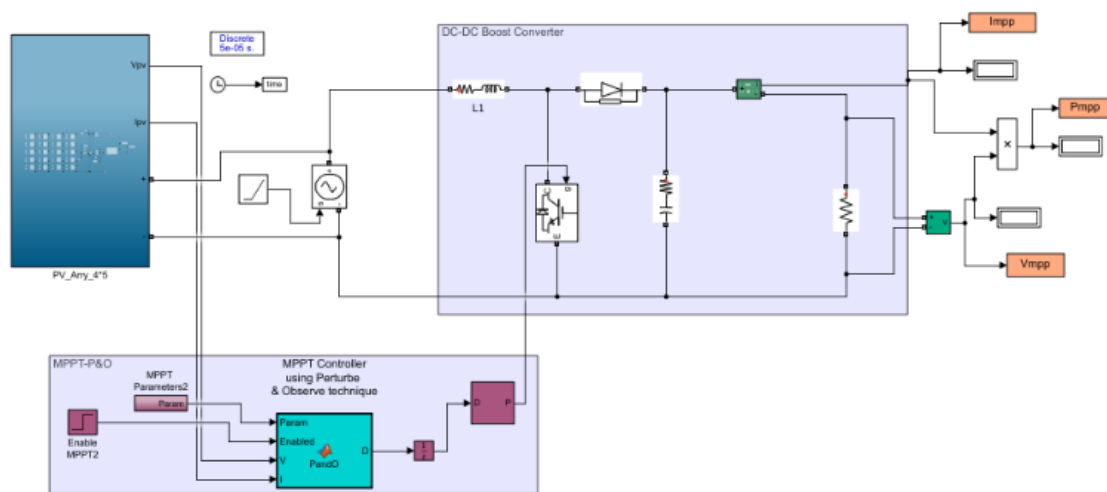
شکل ۲ و ۳ منحنی مشخصه ولتاژ-توان را در دو حالت تابش متغیر-دما ثابت و تابش ثابت-دما متغیر در شرایط عادی نشان می‌دهد.

جدول ۲. پارامترهای ماژول PV تحت شرایط تست در حالت استاندارد

پارامتر	مشخصات ماژول PV
ماکزیمم ولتاژ	۳۱/۹۸
ماکزیمم جریان	۹/۴۰۵
ماکزیمم توان	۳۰۰
جریان اتصال کوتاه	۱۰/۰۵



شکل ۴. انواع خطا در آرایه‌ی خورشیدی



شکل ۵. مدل سیستم PV در محیط شبیه‌ساز متلب

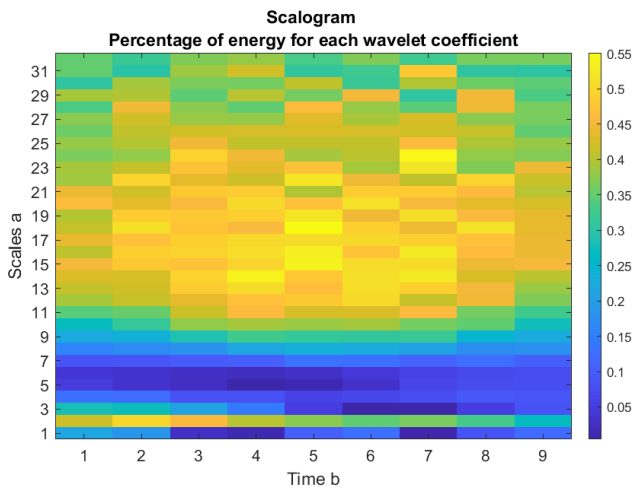
که خطرات ایمنی را نیز دربردارد. داده‌برداری با روش اسکالوگرام که در زمینه‌ی پزشکی از آن استفاده شده علاوه بر افزایش دقت داده‌برداری سایر چالش‌های بیان شده را نیز مرتفع می‌کند

حالت‌های بدون خطا، خطای تخریب، خطای زمین، خطا در جعبه‌ی اتصال، خطای خط به خط، خطای سایه جزئی، خطای مدار باز و تجمیع خطاها، هشت حالت برای ارزیابی روش پیشنهادی در نظر گرفته شده است. داده‌های جمع‌آوری شده شامل تابش (G)، دما (T)، جریان اتصال کوتاه (Isc)، ولتاژ مدار باز (Voc)، جریان فتوولتائیک (Ipv)، ولتاژ فتوولتائیک (Vpv)، جریان MPP (Imp)، ولتاژ MPP (Vmp) و توان MPP (Pmp) است. که ۹ داده مستخرج از خروجی

برای بررسی روش عیب‌یابی پیشنهادی از مدل شبیه‌سازی ارائه شده در شکل ۵ برای جمع‌آوری مجموعه داده تحت شرایط عملیاتی معیوب و غیر معیوب استفاده شده است.

۵- اکتساب داده

روش‌های اکتساب داده اکثراً نیاز به تجهیزات پیشرفته مانند تصویرساز حرارتی، پهپادها و ... دارند که علاوه بر تحمیل هزینه بصورت دوره‌ای داده‌برداری از نیروگاه صورت می‌گیرد که امکان تشخیص خطا در طول این دوره‌ها را با چالش روبه‌رو می‌کند، مخصوصاً برای سه خطای مهم مدار باز، اتصال کوتاه و خطای زمین



شکل ۶. تصاویر اسکالوگرام ورودی شبکه‌ی CNN

از نظر مفهوم شبیه به تبدیل فوریه است اما از خانواده موجک‌ها به‌عنوان توابع پایه خود به‌جای توابع سینوسی و کسینوسی استفاده می‌کند. بیان ریاضی CWT در معادله ۵ آورده شده است که در آن $\psi(t)$ یک نمونه اولیه موجک و $x(t)$ سیگنال اصلی است [۳۱].

$$CWT\{x(t); a, b\} = \int x(t)\psi_{a,b}^*(t)dt \quad (5)$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right)$$

۷- شبکه عصبی کانولوشنی

شبکه عصبی کانولوشنی کلاس خاصی از الگوریتم‌های یادگیری عمیق است. معماری اصلی یک CNN شامل سه لایه اصلی: لایه کانولوشن، لایه ادغام و لایه کاملاً متصل است. عملیات کانولوشن دوبعدی یک نقشه ویژگی بهینه از داده‌های ورودی را با استفاده از تعدادی هسته فیلتر آموخته‌شده تولید می‌کند. لایه ادغام به‌منظور کاهش ابعاد ویژگی داده‌ها عمل می‌کند. در نهایت لایه کاملاً متصل برای طبقه‌بندی استفاده شده است [۳۲].

CNN با استفاده از روش بهینه‌سازی پس انتشار خطا، خطای طبقه‌بندی را بر روی پارامترهای شبکه (به‌عنوان مثال، هسته‌ها/وزن‌های فیلتر) به حداقل می‌رساند. شبکه آموزش داده می‌شود تا ویژگی‌های متمایز را از داده‌های زیربنایی بیاموزد و با استفاده از طبقه‌بندی کننده‌ای مانند لایه softmax منجر به طبقه‌بندی این داده‌ها می‌شود [۱۰].

آرایه PV و ۳ مقدار از مبدل تقویت‌کننده که به‌طور مستقیم یا غیرمستقیم متأثر از وقوع خطا در سیستم PV. تغییر می‌کنند. در مجموع ۱۱۵۲۰ نمونه داده‌ی برچسب‌دار جمع‌آوری شد. هر کلاس دارای ۱۴۴۰ نمونه است که از شبیه‌سازی آرایه PV با ترکیب‌های چندگانه از موقعیت‌های زیر که در جدول‌های ۳ و ۲ نشان داده شده‌اند، استخراج شده است.

جدول ۳. برچسب داده مجموعه داده‌ها

برچسب داده	زیرسیستم
G, T, Isc, Voc, Ipv, Vpv, Imp, Vmp, Pmp	آرایه PV
Imax, Vmax, Pmax	مبدل بوست ^۱

جدول ۴. مقادیر پارامترهای مجموعه داده

پارامترها	مقدار هدف
تابش	۱۰۰-۱۲۰۰ W/m ² تغییر با گام ۲۰
دما	۰-۶۰ °C تغییر با گام ۵

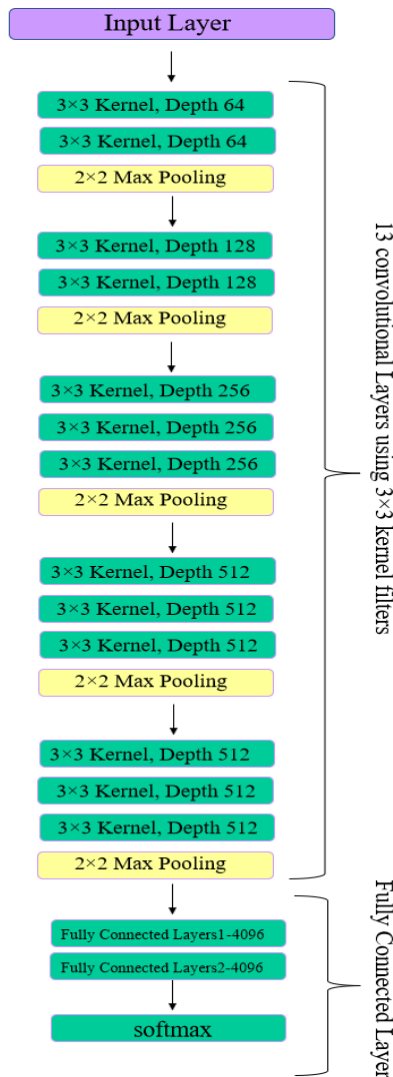
۶- تولید اسکالوگرام دوبعدی^۲

شبکه عصبی کانولوشنی وظایف طبقه‌بندی را مستقیماً از روی تصاویر انجام می‌دهد. امروزه به‌طور گسترده CNN ها در حوزه زیست پزشکی با تبدیل سیگنال‌ها از حوزه زمان (1-D) به حوزه فرکانس (2-D) با استفاده از تبدیل موجک استفاده می‌شوند. داده‌های مستخرج با اعمال تبدیل موجک پیوسته (CWT) سبب تولید اسکالوگرام (تصویر دوبعدی) که در شکل ۶ ارائه شده است می‌شوند. اسکالوگرام یک نمایش بصری سیگنال‌ها بر اساس نمایش زمان-فرکانس با استفاده از تبدیل موجک (WT) است. لازم به ذکر است در این پژوهش به‌منظور افزایش در دقت شبکه عصبی کانولوشنی استفاده شده که به‌طور مستقیم در ارتباط با تعداد داده‌های تصویری ورودی می‌باشد. تصاویر اسکالوگرام با دو وضوح متفاوت استخراج شده است. پس از ایجاد نمودارهای اسکالوگرام برای نمونه داده‌ها در متلب تمام تصاویر اسکالوگرام به‌اندازه استاندارد ۲۲۴ × ۲۲۴ برش داده می‌شوند تا نیاز لایه تصویر ورودی VGG_16 از پیش آموزش دیده را برآورده کنند.

³ Error-Backpropagation

¹ Boost Converter

² 2-D Scalogram



شکل ۷. ساختار شبکه از پیش آموزش داده شده VGG-16

تبدیل موجک پیوسته (CWT) پردازش شده است تا اسکالوگرام‌های دوبعدی تولید شود که به‌طور مؤثر ویژگی‌های فرکانس زمانی داده‌های یک‌بعدی سیستم PV را ضبط کند و داده‌های دوبعدی تولید کند. تصاویر اسکالوگرام اکنون به شکلی هستند که می‌توانند به‌آسانی همراه با یک شبکه عصبی کانولوشنی دوبعدی (2-D CNN) از پیش آموزش دیده برای استخراج ویژگی (در این پژوهش از CNN لایه Sotmax صورت گرفته است. همچنین VGG-16 از پیش آموزش دیده را به‌دقت تنظیم شده است تا یک طبقه‌بندی کننده ۸ طرفه ایجاد کنیم که ویژگی‌ها را استخراج می‌کند و همچنین طبقه‌بندی را در یک چارچوب مشترک انجام می‌دهد. شرح هر مرحله به‌تفصیل در بخش‌های بعدی بیان شده است.

آموزش CNN از ابتدا به مجموعه داده‌های عظیمی برای آموزش نیاز دارد. بنابراین، استفاده از یک شبکه از پیش آموزش دیده توسط مطالعات مختلف توصیه شده است. به‌عبارت‌دیگر، یک شبکه CNN از پیش آموزش دیده بر روی یک مجموعه داده معیار بزرگ که به‌راحتی در دسترس است (مثلاً مجموعه داده ImageNet برای تشخیص اشیا) به‌عنوان یک استخراج‌کننده ویژگی برای هدف قرار دادن یک دامنه متفاوت (مانند خطاهای سیستم PV که در آن تعداد زیادی نمونه آموزشی ممکن است در دسترس نباشد) استفاده شود [۱۰].

۷-۱- شبکه از پیش آموزش داده شده VGG-16

در این پژوهش از VGG-16 CNN از پیش آموزش دیده به‌عنوان الگوریتم استخراج و طبقه‌بندی ویژگی استفاده شده است. ساختار VGG-16 در شکل ۷ نشان داده شده است که شامل ۱۳ لایه کانولوشن و ۲ لایه کاملاً متصل و یک طبقه‌بندی کننده SoftMax است [۳۳].

۸- روش عیب‌یابی پیشنهادی مبتنی بر شبکه عصبی کانولوشنی از پیش آموزش دیده شده

شکل ۸ فلوجارت روش پیشنهادی را نشان می‌دهد. ابتدا، مجموعه‌ای از دیتاست اعمال می‌شود که از مجموع ۱۱۶۴۸ نمونه با ۱۴۵۶ نمونه در هر کلاس تشکیل شده است. این مجموعه داده به‌عنوان یک آزمون ثابت برای ارزیابی رویکرد این پژوهش ارائه شده است تا یک مقایسه معنادار را ایجاد کند. سپس این داده‌های یک‌بعدی با استفاده از تبدیل موجک پیوسته (CWT) پردازش شده است تا اسکالوگرام‌های دوبعدی تولید شود که به‌طور مؤثر ویژگی‌های فرکانس زمانی داده‌های یک‌بعدی سیستم PV را ضبط کند و داده‌های دوبعدی تولید کند. تصاویر اسکالوگرام اکنون به شکلی هستند که می‌توانند به‌آسانی همراه با یک شبکه عصبی کانولوشنی دوبعدی (2-D CNN) از پیش آموزش دیده برای استخراج ویژگی (در این پژوهش از Pretrain VGG-16 CNN استفاده شده است) و طبقه‌بندی بعدی از طریق لایه Sotmax صورت گرفته است. همچنین VGG-16 از پیش آموزش دیده را به‌دقت تنظیم شده است تا یک طبقه‌بندی کننده ۸ طرفه ایجاد کنیم که ویژگی‌ها را استخراج می‌کند و همچنین طبقه‌بندی را در یک چارچوب مشترک انجام می‌دهد. شرح هر مرحله به‌تفصیل در بخش‌های بعدی بیان شده است.

۸-۱- استخراج ویژگی‌های CNN از قبل آموزش دیده

یک ویژگی مهم CNN ها این است که آن‌ها به‌طور خودکار ویژگی‌های طبقه‌بندی‌شده را از داده‌های آموزشی برچسب‌گذاری شده یاد می‌گیرند که طبقه‌بندی را تسهیل می‌کند. همان‌طور که پیش‌تر بیان شد، آموزش مجدد یک CNN از ابتدا نه تنها از نظر محاسباتی گران است، بلکه به مقادیر زیادی داده نیز نیاز دارد. با این وجود، کارهای مختلف در جامعه یادگیری ماشین با موفقیت قدرت تعمیم شبکه‌های عمیق را که از قبل بر روی یک مجموعه داده عظیم انجام شده است، نشان داده‌اند که با توجه به طبقه‌بندی در مجموعه‌های داده دیگر، حتی از حوزه‌های مختلف، عملکرد خوبی دارد. بنابراین، در این پژوهش یک VGG-16 از پیش آموزش داده‌شده بر روی یک مجموعه داده طبقه‌بندی تصویر در مقیاس بزرگ را برای طبقه‌بندی خطا در آرایه‌های PV اعمال شد. به‌طور خاص، فعال‌سازی‌های نورون‌ها در دومین لایه کاملاً متصل «fc7» آغاز می‌شود که در شکل ۸ نشان داده شده است. به عبارت دیگر برای استخراج بردارهای ویژگی از مدل VGG-16، وزن‌های هر ۵ بلوک کانولوشن منجمد شده و خروجی به دست آمده به طبقه‌بندی جدید داده می‌شود. سپس ویژگی‌های از پیش آموزش دیده VGG-16 CNN به یک لایه طبقه‌بندی ۸ کلاس مطابق با شکل ۸ منتقل شد.

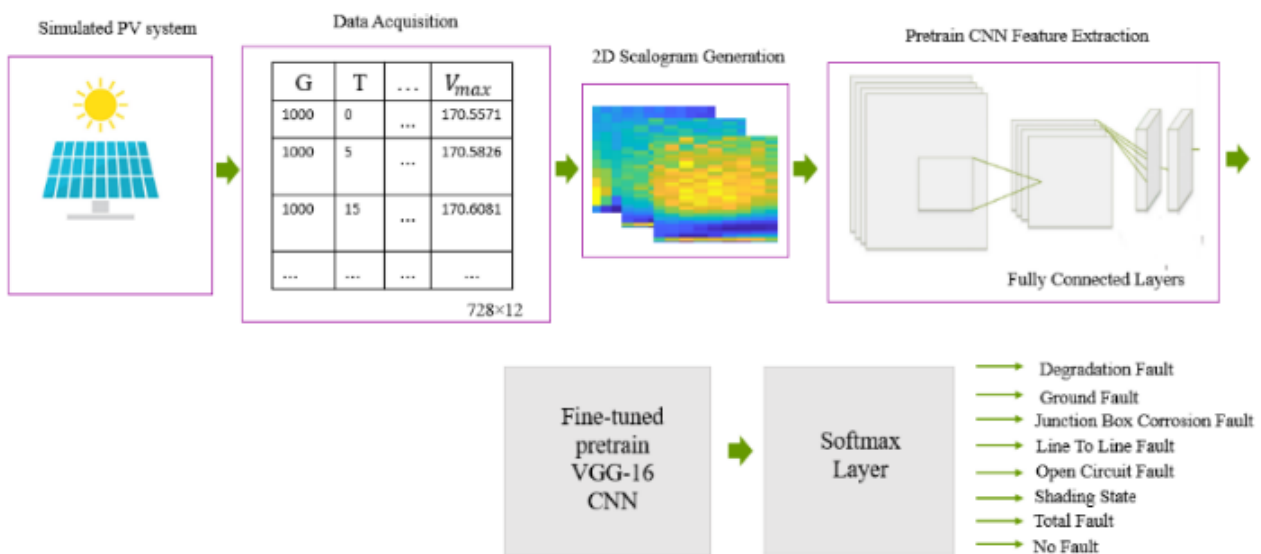
۸-۱-۱ VGG-16 CNN دقیق تنظیم شده

یکی دیگر از تکنیک‌های پرکاربرد برای استفاده مجدد از مدل استخراج ویژگی، تنظیم دقیق است که در شکل ۹ نشان داده شده است. تنظیم دقیق شامل خارج کردن از حالت انجماد چندلایه

بالایی یک پایه مدل منجمد است که برای استخراج ویژگی استفاده می‌شود و آموزش به‌طور مشترک هم بخش جدید اضافه‌شده مدل (در این مورد، طبقه‌بندی کننده کاملاً متصل) و هم تمام لایه‌های بالایی مجدد آموزش داده می‌شود [۳۴].

تنظیم دقیق لایه‌های بالای پایه کانولوشن یک شبکه از پیش آموزش داده‌شده تنها زمانی امکان‌پذیر است که یک طبقه‌بندی کننده قبل از آن آموزش داده‌شده باشد در غیر این صورت سیگنال خطای منتشرشده از طریق شبکه در طول آموزش بسیار بزرگ خواهد بود و ویژگی‌هایی که قبلاً توسط لایه‌های کانولوشنی VGG-16 که به خوبی استخراج شده‌اند از بین می‌روند؛ بنابراین مراحل تنظیم دقیق شبکه به شرح زیر است:

۱. داده‌های تصویری را به یک Pretrain VGG-16 CNN که یک لایه طبقه‌بندی به آن اضافه می‌کنیم جهت استخراج ویژگی‌ها اعمال می‌کنیم.
۲. در مرحله دوم Pretrain VGG-16 CNN را منجمد می‌کنیم.
۳. لایه طبقه‌بندی اضافه‌شده را آموزش می‌دهیم.
۴. بلوک پنجم VGG-16 که شامل دولایه کانولوشن و یک لایه ادغام است را باز کرده است.
۵. در این مرحله لایه باز شده شبکه Pretrain VGG-16 CNN و لایه طبقه‌بندی مجدد آموزش می‌دهیم.



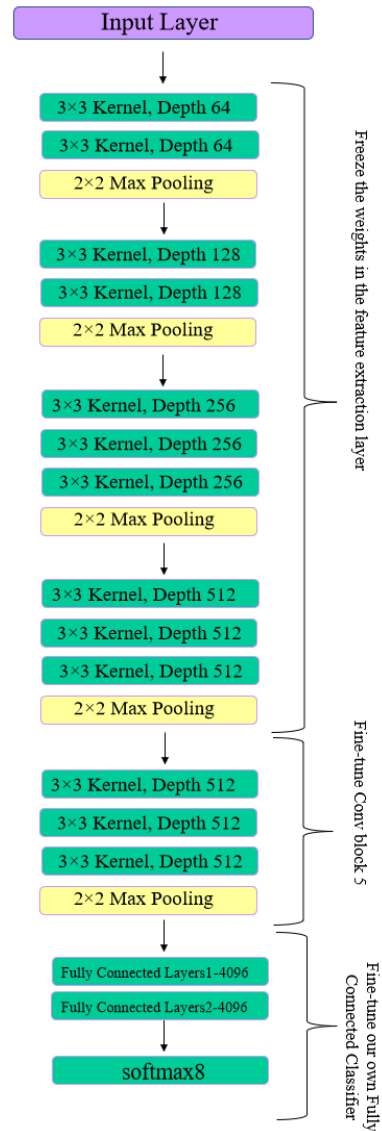
شکل ۸. فلوجارت رویکرد CNN پیشنهادی

استفاده شد. نرخ یادگیری^۵ انتخاب شده ۰/۰۰۰۱ بود. CNN دقیق با ۸۰ درصد داده‌های تصادفی انتخاب شده با ۱۰۰ دوره آموزش^۶ داده شد (یک دوره یک دور کامل آموزش در کل مجموعه داده است). دقت^۷ و ضرر^۸ دو پارامتر مورد استفاده برای ارزیابی عملکرد هستند که در شکل ۱۰ نشان داده شده است. دقت داده‌های آموزش به تدریج افزایش می‌یابد و به حدود ۹۸/۷۳ درصد می‌رسد. با بررسی، می‌توان مشاهده کرد که ضرر به آرامی کاهش می‌یابد و به حدود ۱/۲۷ همگرا می‌شود.

۱۰- تست و ارزیابی

آزمایش بر روی ۲۰ درصد داده‌ها انجام شد و بهترین میانگین دقت و تلفات (بیش از شش اجرای تصادفی)، در هشت کلاس (بدون خطا، خطای تخریب، خطای زمین، خطا در جعبه اتصال، خطای خط به خط، خطای مدارباز و تجمیع خطاها) مشاهده شد. ماتریس درهم ریختگی^۹ در شکل ۱۰ ارائه شده است که دقت هر کلاس (حساسیت) را در امتداد پایین‌ترین ردیف و دقت هر کلاس را در امتداد سمت راست‌ترین ستون نشان می‌دهد. سلول پایین سمت راست، دقت کلی را نشان می‌دهد که در تمام هشت کلاس به‌طور میانگین محاسبه شده است.

برای محاسبه دقت (حساسیت)، دقت یا ویژگی از یک ماتریس درهم ریختگی، ابتدا باید چهار پارامتر TP، FP، TN و FN که به ترتیب بیانگر تعداد مثبت‌های درست، مثبت‌های کاذب، منفی‌های درست و منفی‌های کاذب است را با توجه به یک کلاس خاص از آن محاسبه کنیم. به‌عنوان مثال، ماتریس ارائه شده در شکل ۹ را در نظر بگیرید و فرض کنید که تحلیل برای خطای تخریب است. TP (تعداد نمونه‌های خطای تخریب که به‌درستی به‌عنوان خطای تخریب طبقه‌بندی شده‌اند) ۲۳۵ است؛ و TN (تعداد نمونه‌های نامربوط که به‌درستی به‌عنوان نامربوط طبقه‌بندی شده‌اند) ۴۹ است، FN (تعداد نمونه‌های خطای تخریب که به‌اشتباه به‌عنوان خطای دیگر طبقه‌بندی شده‌اند) با جمع کردن اعداد در ستون اول (به‌استثنای ردیف اول) عدد ۵۶ حاصل می‌شود.



شکل ۹. تنظیم دقیق آخرین بلوک کانولوشن شبکه VGG16

۹- آموزش و بهینه‌سازی پارامترها^۱

عملکرد الگوریتم‌های DL را می‌توان با تنظیم هایپر پارامترها بهبود بخشید. در این مطالعه چندین ترکیب از مقادیر هایپر پارامتر مورد آزمایش قرار گرفت و بهترین ترکیب اتخاذ شد که در ادامه توضیح داده شده است. حداقل اندازه دسته^۲ روی ۱۰ تنظیم شد تا دقت آموزش افزایش یابد. اندازه دسته به تعداد نمونه‌های آموزشی استفاده شده برای یک تکرار اشاره دارد. نزول گرادیان تصادفی^۳ با مونتوم^۴ تنظیم شده روی ۰/۸۵ برای تنظیم دقیق CNN دوبعدی

⁶ Epochs

⁷ Accuracy

⁸ Loss

⁹ Confusion Matrix

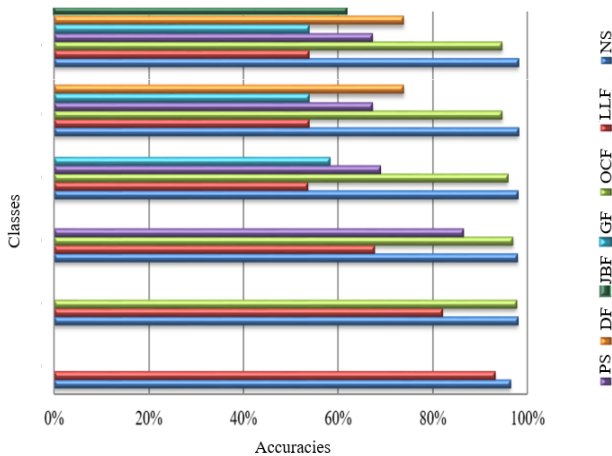
¹ Training and Hyper-Parameters

² Batch Size

³ Stochastic Gradient Descent

⁴ Momentum

⁵ Learning Rate



شکل ۱۱. دقت طبقه‌بندی با افزایش تعداد کلاس خطاها

۱۱- مقایسه pre-train VGG-16 CNN و سایر روش‌ها

در جدول ۵ مقایسه‌ای بین دقت روش‌های تشخیص خطا و تعداد خطاهای قابل تشخیص مبتنی بر الگوریتم‌های یادگیری عمیق صورت گرفته شده است که تأیید می‌کند روش پیشنهادی این مقاله بالاترین درصد دقت با احتساب تعداد بیشتر خطای تشخیصی را در سیستم PV فراهم می‌کند

جدول ۵. مقایسه دقت طبقه‌بندی روش‌های تشخیص خطا مبتنی بر یادگیری عمیق

خطای تشخیصی	روش تشخیصی	دقت طبقه‌بندی خطا
سایه جزئی، خطای تخریب، خطای زمین، خطا در جعبه اتصال، خطای خط به خط، خطای مدار باز و هم‌زمانی خطاها	pre-train CNN VGG-16	۸۳/۳۷۵٪
خطاهای مدار باز، اتصال کوتاه و عدم تطابق	ANN [14]	۷۳/۶٪
خطای خط به خط، خطای زمین و خطای اتصال کوتاه	Bi-LSTM [15]	۸۰/۱۷۴٪
سوختگی، لایه‌لایه شدن و شکستگی	CNN & KNN [16]	۹۶/۹۵٪
خطای زمین، خرابی دیود بای پس و خطای سایه جزئی	CNN [17]	۷۶/۸۲٪
شکستگی و لایه‌لایه شدگی	CNN[18]	۸۰/۸۴٪
شکستگی و لایه‌لایه شدگی	CNN[19]	۸۳/۹۱٪
تشخیص ترک	DL[20]	۹۴/۸۹٪
خطاهای مدار باز، اتصال کوتاه و عدم تطابق	CNN & SVM [21]	۹۱/۲۳٪

کلاس خرابی	DF	GF	JBF	LLF	NS	OCF	PS	TF	
DF	235	8	5	9	0	3	11	13	82.2%
GF	5	196	13	7	1	21	6	17	76.0%
JBF	6	18	232	17	0	9	7	10	82.5%
LLF	8	25	16	190	1	14	10	17	75.1%
NS	0	3	0	0	278	1	0	2	97.8%
OCF	6	8	9	27	3	253	5	3	86.6%
PS	10	22	10	15	2	8	241	28	84.8%
TF	21	11	27	28	4	12	6	233	82.0%
	80.7%	80.3%	76.8%	79.5%	98.6%	86.6%	82.8%	81.2%	83.375 %
	DF	GF	JBF	LLF	NS	OCF	PS	TF	

DF: Degradation Fault (خطای تخریب) NS: Normal State (بدون خطا)
 GF: Ground Fault (خطای زمین) OCF: Open Circuit Fault (خطای مدار باز)
 JF: Junction Box Corrosion Fault (خطای جعبه اتصال) PS: Partial shading (سایه جزئی)
 LLF: Line To Line Fault (خطای خط به خط) TF: Total Fault (همزمانی تمام خطاها)

شکل ۱۰. ماتریس درهم‌ریختگی، نتایج دقت و ضرر طبقه‌بندی خطا برای VGG-16 CNN از پیش آموزش‌دیده دقیق تنظیم‌شده

توجه داشته باشید که $TP + FN$ برابر است با تعداد کل نمونه‌های مربوطه (تعداد نمونه‌های تست خطای کلاس تخریب)، ثابت است، یعنی ۲۹۱.

بر اساس بحث فوق، دقت یا حساسیت برای خطای کلاس تخریب (پایین‌ترین ستون) برابر است با:

$$TP/(TP + FN) = 235/(235 + 56) = 80.7\%$$

همچنین دقت یا حساسیت برای خطای کلاس تخریب (راست‌ترین ستون) برابر است با:

$$TP/(TP + FP) = 235/(235 + 49) = 82.2\%$$

در نهایت میانگین دقت ماتریس درهم‌ریختگی نشان‌دهنده دقت تشخیص خطا به روش شبکه عصبی کانولوشنی از پیش آموزش‌دیده شده است که برابر است با ۸۳/۳۷۵٪.

۱۰-۱- افزایش کلاس‌های خطا

عملکرد روش پیشنهادی (VGG-16 با تنظیم دقیق) با افزایش تعداد کلاس‌ها را می‌توان در شکل ۱۱ تجزیه و تحلیل کرد. خطاهایی با شدت جریان خطای کم مانند خطاهای LL به سختی از سایر خطاها تشخیص داده می‌شوند. دقت خطای LL زمانی که سایه جزئی به عنوان کلاس چهارم اضافه می‌شود تا حد قابل توجهی کاهش می‌یابد. در طبقه‌بندی دو کلاس، دقت برای دو کلاس بیش از ۹۰٪ است، اما با افزایش تعداد کلاس‌ها، عملکرد کاهش می‌یابد.

مراجع

- [1] A. Dhoke, R. Sharma, and T. K. Saha, "A technique for fault detection, identification and location in solar photovoltaic systems," *Solar Energy*, vol. 206, pp. 864-874, 2020.
- [2] A. Jäger-Waldau, "Snapshot of photovoltaics—February 2020," *Energies*, vol. 13, no. 4, p. 930, 2020.
- [3] C. Buerhop, D. Schlegel, M. Niess, C. Vodermayr, R. Weißmann, and C. Brabec, "Reliability of IR-imaging of PV-plants under operating conditions," *Solar Energy Materials and Solar Cells*, vol. 107, pp. 154-164, 2012.
- [4] P. B. Quater, F. Grimaccia, S. Leva, M. Mussetta, and M. Aghaei, "Light Unmanned Aerial Vehicles (UAVs) for cooperative inspection of PV plants," *IEEE Journal of Photovoltaics*, vol. 4, no. 4, pp. 1107-1113, 2014.
- [5] J. Tsanakas and P. Botsaris, "An infrared thermographic approach as a hot-spot detection tool for photovoltaic modules using image histogram and line profile analysis," *International Journal of Condition Monitoring*, vol. 2, no. 1, pp. 22-30, 2012.
- [6] D. S. Pillai and N. Rajasekar, "An MPPT-based sensorless line-line and line-ground fault detection technique for PV systems," *IEEE Transactions on Power Electronics*, vol. 34, no. 9, pp. 8646-8659, 2018.
- [7] D. S. Pillai and N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in PV systems," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 18-40, 2018.
- [8] Y. Zhao, J.-F. De Palma, J. Mosesian, R. Lyons, and B. Lehman, "Line-line fault analysis and protection challenges in solar photovoltaic arrays," *IEEE transactions on Industrial Electronics*, vol. 60, no. 9, pp. 3784-3795, 2012.
- [9] B. P. Kumar, G. S. Ilango, M. J. B. Reddy, and N. Chilakapati, "Online fault detection and diagnosis in photovoltaic systems using wavelet packets," *IEEE Journal of Photovoltaics*, vol. 8, no. 1, pp. 257-265, 2017.
- [10] F. Aziz, A. U. Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A novel convolutional neural network-based approach for fault classification in photovoltaic arrays," *IEEE Access*, vol. 8, pp. 41889-41904, 2020.
- [11] K. AbdulMawjoed, S. S. Refaat, and W. G. Morsi, "Detection and prediction of faults in photovoltaic arrays: A review," in *2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018)*, 2018, pp. 1-8: IEEE.
- [12] R. Hariharan, M. Chakkarapani, G. S. Ilango, and C. Nagamani, "A method to detect photovoltaic array faults and partial shading in PV systems," *IEEE Journal of Photovoltaics*, vol. 6, no. 5, pp. 1278-1285, 2016.
- [13] M. Catelani, L. Ciani, D. Galar, and G. Patrizi, "Optimizing maintenance policies for a yaw system using reliability-centered maintenance and data-driven condition monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6241-6249, 2020.
- [14] S. Voutsinas, D. Karolidis, I. Voyiatzis, and M. Samarakou, "Development of a multi-output feed-forward neural network for fault detection in Photovoltaic Systems," *Energy Reports*, vol. 8, pp. 33-42, 2022.
- [15] Z. Mustafa, A. S. Awad, M. Azzouz, and A. Azab, "Fault identification for photovoltaic systems using a multi-output deep learning approach," *Expert Systems with Applications*, vol. 211, p. 118551, 2023.

در جدول ۵ در منابع [۱۶] و [۲۰-۲۱] دقت تشخیص خطای بالاتری گزارش شده است که به دلیل تشخیص تعداد خطای کمتر و همچنین داده‌برداری در عدم حضور MPPT است که قبلاً نیز بیان شد که وجود MPPT چالشی مهم در تشخیص خطا است.

روش پیشنهادی در منبع [۱۵] بر روی داده‌های بدست آمده در این مقاله شبیه‌سازی شد تا در شرایط یکسان مقایسه ارائه شود، نتایج شبیه‌سازی با روش Bi-LSTM درصد دقت کمتری برای تشخیص خطاها دارد که در جدول ۶ ارائه شده است.

۱۲- نتیجه‌گیری

در این پژوهش، یک پیکربندی تشخیص خطا مبتنی بر شبکه عصبی کانولوشنی برای طبقه‌بندی خطای آرایه PV پیشنهاد شده است؛ که در آن چندلایه آخر یک VGG-16 از پیش آموزش دیده تنظیم شده است تا خروجی هشت طرفه ایجاد کند. برای طبقه‌بندی روش پیشنهادی ویژگی‌های اساسی یک آرایه PV را برای توسعه الگوریتم طبقه‌بندی خطا بدون استفاده از تجهیزات پیشرفته (تصویرساز حرارتی) اندازه‌گیری می‌کند. علاوه بر این، ارزیابی کمی خطاهای آرایه PV بر اساس روش‌های استخراج و طبقه‌بندی ویژگی برای هشت مورد (بدون خطا، خطای سایه جزئی، خطای تخریب، خطای زمین، خطا در جعبه اتصال، خطای خط به خط، خطای مدارباز و تجمع خطاها) تحت شرایط شدید (که در آن این تعداد خطاها به صورت هم‌زمان در گذشته شناسایی نشده‌اند) انجام شده است. اهم دستاوردهای به دست آمده در این تحقیق عبارت‌اند از: اولاً، یک رویکرد جدید برای طبقه‌بندی خطای آرایه PV با استفاده از تولید اسکالوگرام دوبعدی با وضوح متفاوت که سبب افزایش داده و به تبع آن افزایش تعداد داده‌های آموزش که سبب افزایش دقت طبقه‌بندی خطا شد، ارائه گردید و به دنبال آن با بهره‌گیری از شبکه از پیش آموزش داده شده VGG-16 که به نسبت روش مشابه با شبکه AlexNet که دقت تشخیص خطا بالاتری (۸۳/۳۷۵٪) را ارائه می‌کند، پیشنهاد شده و مورد آزمایش قرار گرفت.

جدول ۶. مقایسه دقت طبقه‌بندی روش پیشنهادی

خطای تشخیصی	روش تشخیصی	دقت طبقه‌بندی خطا
خطای سایه جزئی، خطای تخریب، خطای زمین، خطا در جعبه اتصال، خطای خط به خط، خطای مدارباز و همزمانی خطاها	pre- CNN train VGG-16	۸۳/۳۷۵
خطای سایه جزئی، خطای تخریب، خطای زمین، خطا در جعبه اتصال، خطای خط به خط، خطای مدارباز و همزمانی خطاها	Bi-LSTM [15]	۸۲/۱۷۴

- [25] S. R. Madeti and S. Singh, "Online modular level fault detection algorithm for grid-tied and off-grid PV systems," *Solar Energy*, vol. 157, pp. 349-364, 2017.
- [26] M. Dhimish, V. Holmes, B. Mehrdadi, and M. Dales, "Multi-layer photovoltaic fault detection algorithm," *High voltage*, vol. 2, no. 4, pp. 244-252, 2017.
- [27] A. Dhoke, R. Sharma, and T. K. Saha, "PV module degradation analysis and impact on settings of overcurrent protection devices," *Solar Energy*, vol. 160, pp. 360-367, 2018.
- [28] A. Triki-Lahiani, A. B.-B. Abdelghani, and I. Slama-Belkhodja, "Fault detection and monitoring systems for photovoltaic installations: A review," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2680-2692, 2018.
- [29] Y.-Y. Hong and R. A. Pula, "Methods of photovoltaic fault detection and classification: A review," *Energy Reports*, vol. 8, pp. 5898-5929, 2022.
- [30] D. Revati and E. Natarajan, "IV and PV characteristics analysis of a photovoltaic module by different methods using Matlab software," *Materials Today: Proceedings*, vol. 33, pp. 261-269, 2020.
- [31] A. Narin, "Detection of focal and non-focal epileptic seizure using continuous wavelet transform-based scalogram images and pre-trained deep neural networks," *Irbm*, vol. 43, no. 1, pp. 22-31, 2022.
- [32] J. Garcia, S. Muller, E. Caicedo, T. Bastos Filho, and A. Souza, "Non-fatigating brain computer interface based on SSVEP and ERD to command an autonomous car," *Advances in Data Science and Adaptive Analysis*, vol. 1, pp. 1-11, 2018.
- [33] S. Tammina, "Transfer learning using vgg-16 with deep convolutional neural network for classifying images," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 10, pp. 143-150, 2019.
- [34] F. Chollet, *Deep learning with Python*. Simon and Schuster, 2021.
- [16] S. N. Venkatesh and V. Sugumaran, "Machine vision based fault diagnosis of photovoltaic modules using lazy learning approach," *Measurement*, vol. 191, p. 110786, 2022.
- [17] A. Mellit, "An embedded solution for fault detection and diagnosis of photovoltaic modules using thermographic images and deep convolutional neural networks," *Engineering Applications of Artificial Intelligence*, vol. 116, p. 105459, 2022.
- [18] S. Naveen Venkatesh *et al.*, "Photovoltaic Module Fault Detection Based on Deep Learning Using Cloud Computing," *Scientific Programming*, vol. 2023, 2023.
- [19] D. Manno, G. Cipriani, G. Ciulla, V. Di Dio, S. Guarino, and V. L. Brano, "Deep learning strategies for automatic fault diagnosis in photovoltaic systems by thermographic images," *Energy Conversion and Management*, vol. 241, p. 114315, 2021.
- [20] A. Sohail, N. U. Islam, A. U. Haq, S. U. Islam, I. Shafi, and J. Park, "Fault detection and computation of power in PV cells under faulty conditions using deep-learning," *Energy Reports*, vol. 9, pp. 4325-4336, 2023.
- [21] S. Voutsinas, D. Karolidis, I. Voyiatzis, and M. Samarakou, "Development of a machine-learning-based method for early fault detection in photovoltaic systems," *Journal of Engineering and Applied Science*, vol. 70, no. 1, p. 27, 2023.
- [22] S. R. Madeti and S. Singh, "A comprehensive study on different types of faults and detection techniques for solar photovoltaic system," *Solar Energy*, vol. 158, pp. 161-185, 2017.
- [23] V. Kongphet, A. Migan-Dubois, C. Delpha, D. Diallo, and J.-Y. Lechenadec, "Photovoltaic Fault Detection and Diagnosis: Which Level of Granularity for PV Modeling?," in *2020 Prognostics and Health Management Conference (PHM-Besançon)*, 2020, pp. 180-186: IEEE.
- [24] M. Köntges *et al.*, "Review of failures of photovoltaic modules," 2014.

The Relevance, Importance and Dependence of Critical Infrastructures of The Islamic Republic of Iran from a Cyber-Perspective

Abouzar Solat Rafiee¹, Hossain Gharaee Gharakhani^{2*}, Fatemeh Saghafi³, Mohammad Malekinia⁴

¹ Department of Management and Accounting, South Tehran Branch, Islamic Azad University, Tehran, Iran

² Associate Professor, ICT Security faculty, ICT Research Institute (ITRC), Tehran, Iran

³ Associate Professor, Faculty of Management, University of Tehran, Tehran, Iran

⁴ Associate Professor, Department of Management and Accounting, South Tehran Branch, Islamic Azad University, Tehran, Iran

Received: 25 June 2023, Revised: 30 September 2023, Accepted: 18 October 2023

Paper type: Research

Abstract

In recent years, cyber-attacks on the critical infrastructure of countries have increased dramatically. The types of critical infrastructure and their dependencies based on national requirements are different from one country to another. Disruption of the mission or services of a critical infrastructure has a cascading effect on other infrastructures and makes them face serious problems in service delivery. In various studies, approaches Different methods have been adopted to model these dependencies. The important point is not to generalize those models to other countries due to the national requirements of each country. In this research, by forming 11 focus groups consisting of senior and middle managers of each infrastructure field, the network analysis method based on the DEMATEL technique was used, and the most effective and influential critical infrastructure from a cyber-perspective on other critical infrastructures was identified, and the relationship between critical infrastructures and Their prioritization was determined from a cyber-perspective. The results of this research can be useful in the design of the national warning sharing system in order to calculate the national situational awareness in the cyber field and other researches based on the dependence of critical infrastructures.

Keywords: Critical Infrastructure, Critical Infrastructure dependency, DANP.

* Corresponding Author's email: gharaee@itrc.ac.ir

ارتباط، اهمیت و وابستگی زیرساخت‌های حیاتی جمهوری اسلامی ایران از منظر سایبری

ابوذر صولت رفیعی^۱، حسین قرائی گرکانی^{۲*}، فاطمه ثقفی^۲، محمد ملکی نیا^۴
^۱ دانشجوی دکتری مدیریت فناوری اطلاعات دانشگاه آزاد اسلامی واحد تهران جنوب، تهران، ایران
^۲ دانشیار پژوهشگاه ارتباطات و فناوری اطلاعات، تهران، ایران
^۳ دانشیار دانشکده‌گان مدیریت دانشگاه تهران، تهران، ایران
^۴ دانشیار دانشکده مدیریت و اقتصاد دانشگاه آزاد اسلامی واحد تهران جنوب، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۴/۰۴ تاریخ بازبینی: ۱۴۰۲/۰۷/۰۸ تاریخ پذیرش: ۱۴۰۲/۰۷/۲۶
نوع مقاله: پژوهشی

چکیده

سال‌های اخیر حملات سایبری به زیرساخت‌های حیاتی کشورها به طور چشمگیری افزایش یافته است. انواع زیرساخت حیاتی و وابستگی‌های آنها مبتنی بر مقتضیات ملی، از کشوری به کشور دیگر متفاوت است، اختلال در مأموریت یا خدمات یک زیرساخت حیاتی بصورت آبخاری در دیگر زیرساخت‌ها اثر گذاشته و آنها را با مشکلات جدی در امر خدمات‌رسانی روبرو می‌نماید، در مطالعات مختلف رویکردهای متفاوتی جهت مدلسازی این وابستگی‌ها اتخاذ شده است نکته مهم عدم تعمیم آن مدل‌ها برای دیگر کشورها به واسطه مقتضات ملی هر کشور است. در این پژوهش با تشکیل ۱۱ گروه کانونی متشکل از مدیران عالی و میانی هر حوزه زیرساختی از روش تحلیل شبکه مبتنی بر تکنیک DEMATEL استفاده شد و تاثیرگذارترین و تاثیرپذیرترین زیرساخت حیاتی از منظر سایبری بر دیگر زیرساخت‌های حیاتی شناسایی شد و ارتباط بین زیرساخت‌های حیاتی و اولویت‌بندی آنها از منظر سایبری مشخص شد. نتایج این پژوهش می‌تواند در طراحی سامانه ملی اشتراک‌گذاری هشدار به منظور احصاء آگاهی وضعیتی ملی در حوزه سایبری و دیگر پژوهش‌های متکی بر وابستگی زیرساخت‌های حیاتی مفید واقع شود.

کلیدواژگان: زیرساخت‌های حیاتی، وابستگی زیرساخت‌های حیاتی، DANP.

۱- مقدمه

ایالات متحده تحت تأثیر قرار گیرد، شهروندان از آب آشامیدنی سالم و بهداشت محروم خواهند شد. علاوه بر این، بیمارستان‌ها نمی‌توانند کار کنند، شیلنگ‌های آتش‌نشانی کار نمی‌کنند و مدارس، ادارات و تأسیسات دولتی تحت تأثیر قرار خواهند گرفت. در صورتی که هر بخش زیرساخت حیاتی دیگری مورد هدف قرار گیرد، اثرات دومینوی مشابهی رخ خواهد داد.^۱

همچنین تعداد حملات سایبری‌ای که در سال‌های اخیر متوجه سرویس‌ها و زیرساخت‌های ایرانی شده به میزان قابل توجهی رسیده است. اما نکته‌ای که در این میان قابل توجه است این است که حملاتی که از دو سال قبل تا به امروز انجام شده فقط یک حمله سایبری معمولی نیستند بلکه می‌توان گفت ایران در حوزه امنیت سایبری از مهرماه سال ۹۹ وارد یک جنگ سایبری شده است.^۲

مخاطره در زیرساخت‌های حیاتی صرف نظر از منبع انسانی یا طبیعی می‌تواند تأثیر و پیامدهای نگران‌کننده‌ای بر رفاه عمومی داشته باشد که به افراد، مشاغل، دولت و همچنین محیط زیست گسترش می‌یابد. جای تعجب نیست که برخی از نویسندگان پیشنهاد می‌کنند که «شکست این زیرساخت‌ها... یکی از مهم‌ترین آسیب‌پذیری‌های جامعه مدرن است» [۴].

در جدول ۲ به برخی تعاریف و دیدگاه‌ها در خصوص زیرساخت‌های حیاتی پرداخته شده است.

کشورهای مدرن به فضای سایبری، به ویژه، به فناوری اطلاعات، ارتباطات داده، دستگاه‌های تلفن همراه هوشمند و سایر خدمات متصل به سطح جهانی و محاسباتی وابسته هستند. این وابستگی شامل عملیات دولتی، دفاع ملی، زیرساخت‌های حیاتی و رونق اقتصادی است. با این حال، فضای مجازی در معرض اختلالات تصادفی و حملات مخرب از منابع مختلف است [۱]. همچنین ایالات متحده و سایر کشورهای مدرن برای حمایت از جوامع خود به مجموعه وسیعی از زیرساخت‌های حیاتی نیز وابسته هستند. اما زیرساخت‌ها نیز به روش‌های متعددی به یکدیگر وابسته هستند، اما در کلی‌ترین مفهوم، شبکه‌ای از وابستگی‌های متقابل را تشکیل می‌دهند به طوری که اختلال مداوم در یک زیرساخت می‌تواند عملیات در زیرساخت‌های دیگر را تخریب یا متوقف کند [۲، ۳].

بخش‌های زیرساختی حیاتی شامل تولید و انتقال انرژی، آب و فاضلاب، مراقبت‌های بهداشتی و غذا و کشاورزی و ... است. نه تنها هر یک از این بخش‌ها برای عملکرد مناسب جوامع مدرن حیاتی هستند، بلکه به یکدیگر وابسته هستند و حمله به یکی می‌تواند تأثیر مستقیمی بر دیگران داشته باشد. به عنوان مثال، سیستم مالی به شبکه برق بستگی دارد بانک‌ها نمی‌توانند برای مدت طولانی بدون برق کار کنند [۱]. در صورتی که تولید آب آشامیدنی یا تصفیه فاضلاب در نتیجه تهدیدات مداوم بر سیستم‌های آب و فاضلاب

جدول ۱. برخی حملات سایبری به زیرساخت‌های حیاتی ایالات متحده آمریکا

ردیف	سال	عنوان	حمله از	حمله به	واقعه	آثار	اقدام	زیرساخت‌های آسیب پذیر/ دیده	مرجع
۱	فوریه ۲۰۲۱	تضمین آب اولدزمار	ارتباطات و فناوری اطلاعات	آب	سیستم‌های کامپیوتری یک تصفیه‌خانه آب در اولدزمار، فلوریدا را هک شد	هکر سطح هیدروکسید سدیم کارخانه را به طور موقت به سطح خطرناکی تغییر داد.	این تغییر بلافاصله شناسایی شد و با جداسازی منابع ذخیره آسیب دیده از مسمومیت مردم محلی جلوگیری کرد	آب سلامت محیط زیست	Lynngas2021
۲	۲۰۲۱ می	خط لوله دولتی	ارتباطات و فناوری	انرژی	حمله باج افزار به بزرگترین خط لوله توزیع سوخت در ایالات متحده	قیمت‌ها افزایش یافت و عرضه سوخت برای مصرف‌کنندگان برای چندین هفته مختل شد.	باج به اپراتورهای باج افزار پرداخت شد.	انرژی مالی	Turton2021
۳	ژوئن ۲۰۲۱	شرکت JBS بزرگترین	ارتباطات و فناوری اطلاعات	غذایی	باج‌افزاری حیرت‌انگیز	کارخانه‌های فرآوری مرغ و گوشت گاو و سازمان‌های وابسته تعطیل شدند. قطع تغییر عرضه مواد غذایی بازار مصرف به طور موقت	نامعلوم	زنجیره تامین غذایی مالی	Collier2021

² <https://cert.ir>

¹ <https://www.gartner.com/en/articles/why-critical-infrastructure-attacks-are-everyone-s-problem-especially-now>

جدول ۲. تعاریف و دیدگاه‌ها در خصوص زیرساخت‌های حیاتی

مضامین و دیدگاه زیرساخت حیاتی	سال	مرجع	
زیرساخت‌های حیاتی آنقدر حیاتی هستند که ناتوانی یا نابودی آنها تأثیری تضعیف‌کننده بر دفاع یا امنیت اقتصادی ایالات متحده خواهد داشت. [۵]	1996	Clinton	۱
زیرساخت‌های حیاتی شامل: سیستم‌ها و دارایی‌ها، اعم از فیزیکی یا مجازی، آنقدر برای ایالات متحده حیاتی هستند که ناتوانی یا نابودی این گونه سیستم‌ها و دارایی‌ها می‌تواند تأثیر تضعیف‌کننده‌ای بر امنیت، امنیت اقتصادی ملی، سلامت یا ایمنی عمومی ملی یا هر ترکیبی از آنها داشته باشد.	2001	US Congress	۲
زیرساخت‌های حیاتی شامل آن دسته از امکانات، شبکه‌ها، خدمات و دارایی‌های فیزیکی و فناوری اطلاعات است که در صورت اختلال یا تخریب، تأثیر جدی بر سلامت، ایمنی، امنیت یا رفاه اقتصادی شهروندان یا عملکرد مؤثر دولت‌ها خواهد داشت [۶].	2004	European Council	۳
زیرساخت‌های حیاتی شامل آن دسته از تأسیسات، سیستم‌ها یا مجموعه‌هایی می‌شود که ناتوانی یا تخریب آنها می‌تواند تأثیر تضعیف‌کننده‌ای بر امنیت ملی، حاکمیت، اقتصاد و رفاه اجتماعی یک کشور داشته باشد [۷].	2015	India (NCISIPC)	۴
زیرساخت حیاتی هر چیزی است که اختلال در آن ثبات را در جامعه تضعیف کند و در نتیجه امنیت ملی را تهدید کند. تأکید می‌شود که زیرساخت سیستمی است که اشیاء مختلف، پیوندهای بین آنها را ترکیب می‌کند و انواع خاصی از فعالیت‌های انسانی را فراهم می‌کند [۸].	2023	Israel	۵
زیرساخت‌های حیاتی اساس زندگی اجتماعی و فعالیت‌های اقتصادی مردم است که توسط کسب‌وکارهایی شکل می‌گیرد که خدماتی را ارائه می‌دهند که جایگزینی آنها با دیگران بسیار دشوار است، اگر عملکرد آن به حالت تعلیق، زوال یا غیرقابل دسترس شدن باشد، می‌تواند تأثیرات قابل‌توجهی بر زندگی اجتماعی مردم داشته باشد [۹].	2009	Japan The Information Security Policy Council	۶
زیرساخت حیاتی به عنوان سیستم و دارایی‌ها، اعم از فیزیکی یا مجازی که ناتوانی یا نابودی چنین سیستم‌ها و دارایی‌ها می‌تواند تأثیر تضعیف‌کننده‌ای بر امنیت، امنیت اقتصادی ملی، سلامت یا ایمنی عمومی ملی یا هر ترکیبی از آنها داشته باشد [۱۰].	2022	Kingdom of Saudi Arabia NISS	۷
زیرساخت‌ها حیاتی هستند زیرا خدماتی را ارائه می‌دهند که برای یک یا چند عملکرد یا ویژگی‌های عمومی دولتی یا اجتماعی حیاتی هستند. این امر می‌تواند با بقای شهروندان تا آنجایی که به امنیت جان آنها مربوط می‌شود یا به کیفیت زندگی آنها مرتبط باشد [۱۱].	2006	Gheorghe et al	۸
زیرساخت حیاتی: سیستم‌ها، خدمات و عملکردهای کلیدی که اختلال یا تخریب آنها می‌تواند تأثیر تضعیف‌کننده‌ای بر سلامت و ایمنی عمومی، تجارت و امنیت ملی یا هر ترکیبی از اینها داشته باشد [۱۲].	2008	ITU	۹
زیرساخت‌های حیاتی آن دسته از امکانات فیزیکی، زنجیره‌های تامین، فناوری‌های اطلاعات و شبکه‌های ارتباطی هستند که در صورت تخریب، تخریب یا غیرقابل دسترس بودن برای مدت طولانی، به طور قابل‌توجهی بر رفاه اجتماعی یا اقتصادی کشور تأثیر می‌گذارند یا بر توانایی استرالیا برای انجام دفاع ملی تأثیر می‌گذارند. امنیت ملی را تضمین کند [۱۳].	2010	Australian Critical Infrastructure Resilience Strategy	۱۰
دارایی‌های فیزیکی، سیستم‌ها یا تأسیسات، که در صورت مختل شدن، به خطر افتادن یا تخریب، تأثیر جدی بر سلامت، ایمنی، امنیت یا رفاه اقتصادی قطر یا عملکرد مؤثر دولت قطر خواهد داشت [۱۴].	2014	Qatar National Cyber Security Strategy	۱۱
سیستم و دارایی‌ها، چه فیزیکی و چه مجازی، برای ایالات متحده آنقدر حیاتی هستند که ناتوانی یا نابودی چنین سیستم‌ها و دارایی‌هایی می‌تواند تأثیر تضعیف‌کننده‌ای بر امنیت، امنیت اقتصادی ملی، سلامت یا ایمنی عمومی ملی یا هر ترکیبی از این موارد داشته باشد [۱۵].	2015	NIST	۱۲
سازمان‌ها و امکاناتی که برای عملکرد جامعه و اقتصاد در کل ضروری است [۱۶].	2013	ISO/IEC TR 27019:	۱۳
زیرساخت‌های اطلاعاتی حیاتی: آن دسته از سیستم‌هایی که برای یک کشور آنقدر حیاتی هستند که ناتوانی یا تخریب آنها می‌تواند بر امنیت ملی، اقتصاد، یا سلامت و ایمنی عمومی تأثیر منفی بگذارد [۱۷].		IETF RFC 449 Internet Security Glossary 2	۱۴
آن دسته از امکانات، سیستم‌ها، سایت‌ها و شبکه‌های ضروری برای عملکرد کشور و ارائه خدمات ضروری که زندگی روزمره در بریتانیا به آن بستگی دارد [۱۸].	2016	UK Centre for the Protection of National Infrastructure	۱۵
زیرساخت حیاتی: سیستم‌ها و دارایی‌های فیزیکی یا مجازی تحت صلاحیت یک دولت که به قدری حیاتی هستند که ناتوانی یا تخریب آنها ممکن است امنیت، اقتصاد، سلامت یا ایمنی عمومی یا محیط‌زیست یک دولت را تضعیف کند [۱۹].	2013	NATO	۱۶
زیرساخت‌های حیاتی به فرایندها، سیستم‌ها، امکانات، فناوری‌ها، شبکه‌ها، دارایی‌ها و خدمات ضروری برای سلامت، ایمنی، امنیت یا رفاه اقتصادی کانادایی‌ها و عملکرد مؤثر دولت اشاره دارد [۲۰].	2022	Canada An Emergency Management Framework for Canada	۱۷
زیرساخت‌های بحرانی: زیرساخت‌هایی که عملکرد آنها ضروری است و راه‌حل‌های جایگزین را نمی‌پذیرد، به همین دلیل قطع یا تخریب آن تأثیر جدی بر خدمات عمومی اساسی یا ساختارهای دولتی خواهد داشت [۲۱].	2018	Cuba Glossary of Cyber terms	۱۸
ساختارها و تأسیسات سازمانی و فیزیکی از چنان اهمیت حیاتی برای جامعه و اقتصاد یک ملت برخوردار است که شکست یا تخریب آنها منجر به کمبود مداوم عرضه، اختلال قابل توجه در امنیت و ایمنی عمومی یا سایر پیامدهای شگرف می‌شود [۲۲].	2009	Germany Germany Federal Ministry of the Interior, FRG	۱۹

مضامین و دیدگاه زیرساخت حیاتی	سال	مرجع	
یک زیرساخت حیاتی یک تاسیسات، سیستم یا بخشی از آن با منافع فدرال است که برای حفظ عملکردهای حیاتی اجتماعی، سلامت، ایمنی، امنیت، رفاه اقتصادی یا اجتماعی مردم ضروری است و در صورت مختل شدن یا تخریب، تاثیر قابل توجهی خواهد داشت [۲۳].	2011	Belgium Federale Overheidsdienst Binnenlandse Zaken	۲۰
زیرساخت‌های حیاتی: شامل بخش‌هایی می‌شود که دارایی‌ها، سیستم‌ها و شبکه‌های آنها، اعم از فیزیکی یا مجازی، آن‌قدر حیاتی تلقی می‌شوند که از کار افتادن یا تخریب آنها تأثیر مخربی بر امنیت، امنیت اقتصادی ملی، سلامت یا ایمنی عمومی ملی یا هر ترکیبی از آنها خواهد داشت [۲۴].	2020- 2023	Brazil National Cybersecurity Strategy	۲۱
زیرساخت حیاتی فدراسیون روسیه موضوعی است که نقض (یا خاتمه) عملیات آن منجر به از دست دادن کنترل، تخریب زیرساخت‌ها، تغییرات منفی غیرقابل برگشت (یا شکست) اقتصاد، موضوع فدراسیون روسیه یا اداری-سرزمینی می‌شود و تاثیر قابل توجهی در سلامت و ایمنی افراد ساکن در این مناطق برای دراز مدت دارد [۲۴].	(2012)	RUSSIA NATIONAL SECURITY OF RUSSIA - Information security	۲۲
زیرساخت‌های حیاتی: زیرساخت‌ها، شامل امکانات، سیستم‌ها، فرآیندها، شبکه‌ها، فناوری‌ها، دارایی‌ها و خدمات - لازم برای حفظ یا بازیابی عملکردهای حیاتی اجتماعی [۲۵].	(2022- 2024)	Denmark Danish Cyber and Information Security Strategy	۲۳
زیرساخت حیاتی به معنای دارایی، تسهیلات، تجهیزات، شبکه یا سیستم، یا بخشی از دارایی، تسهیلات، تجهیزات، شبکه یا سیستم است که برای ارائه یک سرویس ضروری است [۲۶].	(2022)	European Parliament and of the CounCISI	۲۴

می‌شوند. آن شاخص‌ها عبارتند از:

- اهمیت کارکرد هر حوزه در تأمین نیازهای حیاتی مردم در شرایط اضطراری
 - شدت اثرگذاری بر اقتصاد، امنیت ملی و سلامت مردم
 - وابستگی زیرساخت‌های سایر حوزه‌ها به عملکرد آنها
 - شدت پیامد وقوع تهدید مبتنی بر جغرافیا و جمعیت
- همچنین زیرساخت حیاتی کشور را از منظر امنیت سایبری می‌توان به پنج بخش زیرساخت انرژی، زیرساخت حمل و نقل، زیرساخت ارتباطات و فناوری اطلاعات، زیرساخت مالی و زیرساخت سلامت تقسیم نمود. زیرساخت انرژی خود به ۷ بخش گاز، پتروشیمی، نفت، برق، آب، کنترل صنعتی و انرژی اتمی مطابق شکل ۱ تقسیم می‌شود.



شکل ۱. زیرساخت‌های حیاتی جمهوری اسلامی ایران از منظر سایبری

تعریف زیرساخت در این مقاله: به مجموعه‌ای از مراکز و تأسیسات زیربنایی و شریان‌های عمده که خدمات و نیازهای ضروری و اساسی کشور را به مردم و جامعه ارائه می‌کند، اطلاق می‌گردد؛ زیرساخت مشتمل بر فرابخش (حوزه)، بخش، زیربخش، دارایی و اجزاء آن می‌باشد.

ترکیب فهرست بخش زیرساخت حیاتی مبتنی بر موقعیت ملی، از کشوری به کشور دیگر متفاوت است. نمونه‌ای از آنها در جدول ۳ مشاهده می‌شوند. با توجه به اطلاعات قابل مشاهده در این جدول، شش بخش نخست زیرساخت حیاتی، در تمامی کشورها به عنوان زیرساخت حیاتی شناخته می‌شوند. همچنین با توجه به این جدول تعداد زیرساخت‌های موجود در هر کشور نیز مشخص شده است.

با توجه به اطلاعات موجود، و با عنایت به طرح راهبردی حفاظت از زیرساخت‌های کشور ایران مصوب شصت و ششمین جلسه کمیته دائمی (شورای عالی) پدافند غیرعامل در تاریخ ۱۴۰۱/۰۶/۲۹ که به تأیید مقام معظم رهبری و فرماندهی کل قوا (مدظله‌العالی) نیز رسیده است، و بر اساس ماده ۱ بند ۴ آن، حوزه‌های با اهمیت بالا: به هریک از حوزه‌های ۱- انرژی، ۲- آب، ۳- غذا و کشاورزی، ۴- حمل و نقل، ۵- بهداشت و سلامت، ۶- دفاعی و امنیتی، ۷- صنعت، ۸- رسانه، ۹- هسته‌ای، ۱۰- فضا، ۱۱- جمعیت، ۱۲- حاکمیتی، ۱۳- خدمات ضروری و فوریتی، ۱۴- پولی و مالی، ۱۵- ارتباطات و فناوری اطلاعات، حوزه‌های با اهمیت بالا گفته می‌شود که با توجه به شاخص‌های زیر، حوزه‌های ۱- انرژی، ۲- آب، ۳- ارتباطات و فناوری اطلاعات، ۴- حمل و نقل، ۵- بهداشت و سلامت، ۶- غذا و کشاورزی، ۷- دفاعی و امنیتی و ۸- حاکمیتی بعنوان حوزه‌های کلیدی از منظر پدافند غیرعامل دسته‌بندی

جدول ۳. مقایسه برخی زیرساخت‌های حیاتی در چند کشور بطور نمونه

ایران	کره	اسپانیا	انگلستان	فیلادلفیا	آمریکا	اتحادیه اروپا	چک	کرواسی	شیلی	کانادا	بلژیک	بنگلادش	اتریش	استرالیا	زیرساخت	
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	ارتباطات و فناوری اطلاعات	۱
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	مالی	۴
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	حمل و نقل	۵
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	بهداشت و سلامت	۶
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	انرژی	۷
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	هسته‌ای	۸
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	آب	۹
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	پتروشیمی	۱۰
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	نفط	۱۱
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	گاز	۱۲
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	برق	۱۳
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	غذا (کشاورزی)	۱۴
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	پژوهش (تحقیقات)	۱۵
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	شبکه اجتماعی	۱۶
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	رسانه	۱۷
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	سیستم توزیع	۱۸
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	امداد و نجات (خدمات اضطراری)	۱۹
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	حاکمیتی (نهادهای قانون‌گذار)	۲۰
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	آثار تاریخی	۲۱
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	مراکز تولید	۲۲
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	امکانات دولتی	۲۳
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	دفاعی (صنایع دفاعی) (امنیتی)	۲۴
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	خدمات اجتماعی	۲۵
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	علم و آموزش	۲۶
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	شیمیایی	۲۷
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	سد سازی	۲۸
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	امکانات تجاری	۲۹
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	محیط زیست	۳۰
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	نگهداری زیرساخت	۳۱

جدول ۴. مشخصات اعضای شرکت‌کننده در پنل خبرگی

جنسیت	تحصیلات	تعداد شرکت‌کنندگان داخل پنل	حوزه و تخصص
۸ نفر زن	۱۲ نفر دکتری	۳۶	امنیت سایبری، مدیریت ارتباطات و فناوری اطلاعات، ارتباطات و فناوری اطلاعات، کامپیوتر، برق و زیرساخت‌های حیاتی و پدافند غیر عامل
۲۸ نفر مرد	۲۴ نفر کارشناسی ارشد		

در این مرحله پنل خبرگی متشکل از متخصصان و مدیران صاحب‌نظر در حوزه‌های امنیت سایبری و زیرساخت‌های حیاتی با ترکیب جدول ۴ تشکیل گردید و پرسشنامه‌ای متشکل از عناوین زیرساخت‌های حیاتی به منظور تعیین زیرساخت‌های حیاتی کشور از منظر امنیت سایبری در اختیار اعضای پنل قرار گرفت.

در خصوص زیرساخت رسانه رهبر معظم انقلاب با نگاه کلان و راهبردی خود، در بیانات گوناگونی از سال‌های گذشته به مسئله «رسانه» و «جنگ رسانه‌ای» پرداخته‌اند و نکات ارزنده و مهمی در ابعاد گوناگون این مفهوم بیان داشته‌اند. از این‌رو با بهره‌گیری از بیانات ایشان فقط در بازه زمانی یک ماه اول سال ۱۴۰۲ به نقش و اهمیت رسانه در جنگ ترکیبی بی‌سابقه دشمن علیه امنیت ملی کشور در جدول ۵ اشاره شده است که به لزوم در نظر گرفتن زیرساخت رسانه به عنوان یک زیرساخت حیاتی و مهم تاکید دارد.

همچنین با نگاهی به گزارش مرکز ملی مدیریت امداد و هماهنگی رخدادهای رایانه‌ای در پاییز ۱۳۹۹ الی تابستان ۱۴۰۰ در جدول ۶ تعداد حملات سایبری‌ای که در سال‌های اخیر متوجه سرویس‌ها و زیرساخت‌های ایرانی شده به میزان قابل توجهی رسیده است. اما نکته قابل توجه تعداد حملاتی است که در حوزه زیرساخت رسانه صورت پذیرفته و متوجه سازمان‌های مرتبط با حوزه رسانه شده است.

پس از بررسی و تحلیل نتایج ۱۰۰ درصد شرکت کنندگان در پنل زیرساخت ارتباطات و فناوری اطلاعات، زیرساخت مالی (پولی و بانکی)، زیرساخت حمل و نقل، زیرساخت رسانه و زیرساخت برق را از جز زیرساخت‌های حیاتی از منظر امنیت سایبری (تاثیرگذاری و تاثیرپذیری) انتخاب نمودند همچنین زیرساخت پتروشیمی و زیرساخت گاز با کسب ۸۸/۸۸ درصد در رتبه بعدی قرار داشتند ۸۳/۳۳ و انرژی هسته‌ای و زیرساخت آب نیز امتیاز ۸۰/۵۵ درصد را به خود اختصاص دادند اما دیگر زیرساخت‌ها عددی بیش از ۵۰ درصد را کسب نمودند.

با تطبیق نتایج حاصله از نظر خبرگان این پژوهش با اسناد بالادست و دسته‌بندی ارائه شده توسط مرکز ملی فضای مجازی و طرح راهبردی حفاظت از زیرساخت‌های کشور ایران توسط پدافند غیر عامل، نظر خبرگان این پژوهش در انتخاب زیرساخت‌ها به جز زیرساخت رسانه در مستندات مذکور تایید شد.

جدول ۵. بیانات رهبر معظم انقلاب در بازه یک ماه (فروردین ماه ۱۴۰۲) در خصوص مبحث رسانه و جنگ رسانه‌ای

ردیف	تاریخ	محل ایراد بیانات	بیانات معظم له
۱	۱۴۰۲/۰۱/۰۱	در اجتماع زائران و مجاوران حرم مطهر رضوی	در مقابل کودتا، در مقابل تحریم، در مقابل فشارهای سیاسی، در مقابل تهاجم رسانه‌ای؛ این تهاجم رسانه‌ای که برای ایران‌هراسی و انقلاب‌هراسی در دنیا راه انداخته‌اند، بی‌سابقه است؛ چنین چیزی هرگز وجود نداشته. در مقابل توطئه‌های امنیتی کدام ملت بجز ملت ایران میتواند یا توانسته در مقابل اینها ایستادگی کند؟
۲	۱۴۰۲/۰۱/۰۱	در اجتماع زائران و مجاوران حرم مطهر رضوی	در جنگ ترکیبی، دشمن از رسانه استفاده میکند، از عامل فرهنگی استفاده میکند، از عامل امنیتی استفاده میکند، از نفوذ استفاده میکند، از عامل اقتصادی استفاده میکند؛ از همه این عوامل استفاده میکنند برای اینکه ملت را در محاصره قرار بدهد، برای اینکه ملت را دچار یأس کند، ملت را از نیروی خودش غافل کند.
۳	۱۴۰۲/۰۱/۰۱	در اجتماع زائران و مجاوران حرم مطهر رضوی	یک توصیه مهم من به همه کسانی که توانایی سخن گفتن با مردم را دارند و رسانه در اختیارشان است — چه در فضای مجازی، چه در مطبوعات، چه در صدا و سیما — امیدآفرینی است. دشمن سعی میکند جوانهای ما را ناامید کند؛ ما باید متقابلاً امیدآفرینی کنیم.
۴	۱۴۰۲/۰۱/۱۵	در دیدار مسئولان نظام	انصافاً گاهی انسان فکرهای بسیار خوب، بسیار نو مشاهده میکند؛ خوب اینها امکانات کشور است؛ اینها همه باید در خدمت مهار تورم و رشد تولید قرار بگیرد. امکانات اجرایی داریم، امکانات تقنینی داریم، امکانات قضائی داریم، امکانات رسانه‌ای داریم، تجربه‌ها و عناصر مجرب داریم؛ آمریکا با فشار تحریم میخواست مسئله‌ی هسته‌ای را طبق برنامه‌ی خودش به پایان برساند، نتوانست؛ این ضعف آمریکا است. خیلی تلاش کرد هیاهو کرد [از طریق] رسانه، غیر رسانه؛ سیاست، تحریم، غیره [ولی] قادر نشد، نتوانست مسئله‌ی هسته‌ای را طبق برنامه‌ی خودش حل کند.
۵	۱۴۰۲/۰۱/۱۵	در دیدار مسئولان نظام	آخرین مسئله هم مسئله‌ی رسانه است که مهم است. من بارها درباره‌ی رسانه صحبت کرده‌ام، باز هم عرض میکنم. باید سیاه‌نمایی‌های دشمن، تخریب‌های دشمن، توطئه‌ها علیه اقتدار کشور که به وسیله‌ی دشمن در فضای مجازی و مانند اینها صورت میگیرد، باید افشا بشود؛ این به عهده‌ی رسانه است. خوب، رسانه‌ی ملی بحمدالله دست افراد مؤمن و پُرانگیزه است؛ تلاش کنند که تلاش دشمن در این باره را خنثی کنند.
۶	۱۴۰۲/۰۱/۱۶	دیدار شاعران و اساتید ادبیات فارسی	حضرت آیت‌الله خامنه‌ای، بخش دیگری از هجوم متنوع بدخواهان به ایران را هجوم رسانه‌ای و استفاده آنها از هزاران رسانه برای ترویج دروغ، شایعه و انحراف دانستند و افزودند: هدف دشمن از این تهاجم، سلب نقاط قوت فکری و معارفی و تضعیف روحیه استقلال و استقامت ملی و وحدت و عمل اسلامی است.
۷	۱۴۰۲/۱/۲۹	در دیدار رضائی دانشجویان	یک مثال دیگر برای بدبین کردن ما به خود [این است که] رسانه‌های جورواجور بدخواه، اصرارشان این است که ثابت کنند ملت ایران از اعتقادات دینی روگردان شده، از احساسات انقلابی روگردان شده. اصرار دارند؛ این را بارها و بارها آنها میگویند، یک دنباله‌هایی هم اینجا دارند که اینها هم میگویند؛ در فضای مجازی میگویند.

جدول ۶. گزارش مرکز ملی مدیریت امداد و هماهنگی رخدادهای رایانه‌ای

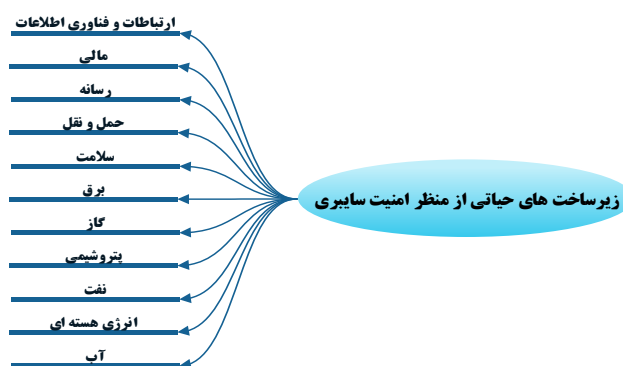
ردیف	تاریخ	سازمان مورد هدف
۱	۲۲ مهر ۱۳۹۹	سازمان بنادر و کشتیرانی
۲	۲۸ آبان ۱۳۹۹	شرکت ملی گاز ایران
۳	۲۲ فروردین ۱۴۰۰	تجهیزات هسته‌ای نطنز
۴	۱۸ تیر ۱۴۰۰	شرکت راه‌آهن
۵	۱۹ تیر ۱۴۰۰	پورتال وزارت راه
۶	۳۱ مرداد ۱۴۰۰	زندان اوین
۷	۴ آبان ۱۴۰۰	سامانه هوشمند توزیع سوخت
۸	۳۰ آبان ۱۴۰۰	شرکت هواپیمایی ماهان
۹	۷ بهمن ۱۴۰۰	صدا و سیما
۱۰	۱۲ بهمن ۱۴۰۰	سامانه تلویزیون
۱۱	۱۸ بهمن ۱۴۰۰	زندان قزلحصار
۱۲	۲۳ اسفند ۱۴۰۰	وزارت ارشاد
۱۳	۲۰ خرداد ۱۴۰۱	شهرداری تهران
۱۴	۶ تیر ۱۴۰۱	صنایع فولاد
۱۵	۱۲ تیر ۱۴۰۱	سازمان فرهنگ و ارتباطات اسلامی

معرض واکنش‌های زنجیره‌ای ناکارآمدی هستند [۲۷]. همچنین از آنجایی که عملیات تجاری به طور فزاینده‌ای بر فناوری اطلاعات تکیه می‌کنند، زیرساخت‌های مدرن به طور فزاینده‌ای به هم مرتبط شده‌اند. در نتیجه، خطر این که حتی اختلالات جزئی در یک زیرساخت می‌تواند منجر به یک آبخار فاجعه بار از خرابی در شبکه‌های زیرساخت‌ها شود در حال افزایش است. همچنین توجه دولت‌ها به زیرساخت‌ها و وابستگی‌های متقابل آنها در حال افزایش است. این امر بسیاری از محققان را تحریک می‌کند تا رویکردهای نوآورانه‌ای را برای شناسایی، توصیف و مدل سازی چنین وابستگی‌های متقابل بین زیرساخت‌ها توسعه دهند [۳]. آنها همچنین استدلال می‌کنند که یک زیرساخت را نمی‌توان به عنوان سیستم جدا شده از سایر زیرساخت‌ها در نظر گرفت. ابتدا تعاریف وابستگی و وابستگی متقابل بین زیرساخت‌ها به شرح زیر ارائه می‌شود.

وابستگی یک پیوند یا ارتباط بین دو زیرساخت است که به وسیله آن وضعیت یک زیرساخت بر وضعیت زیرساخت تأثیر می‌گذارد یا به آن متکی است. وابستگی متقابل یک رابطه دوسویه بین دو زیرساخت است که در آن وضعیت هر زیرساخت بر وضعیت دیگری تأثیر می‌گذارد یا به آن وابسته است.

به طور کلی، محققان از وابستگی و وابستگی متقابل برای توصیف مفهوم پیوند مستقیم از یک زیرساخت به زیرساخت بر اساس معیارهای خاص استفاده می‌کنند. رینالدی ایده‌ای از توصیف کیفی وابستگی‌های متقابل مستقیم بر اساس محصولات و خدماتی را ارائه می‌دهد که زیرساخت‌های حیاتی به یکدیگر ارائه می‌کنند [۳].

با توجه به موارد ذکر شده زیرساخت رسانه از اهمیت قابل توجهی در حوزه امنیت سایبری برخوردار بوده و در این پژوهش به زیرساخت‌های تعیین شده توسط اسناد بالادست اضافه می‌گردد. لذا زیرساخت‌های حیاتی کشور از منظر امنیت سایبری مطابق شکل ۲ تعیین می‌گردد.



شکل ۲. زیرساخت‌های حیاتی جمهوری اسلامی ایران از منظر امنیت سایبری بر اساس نظر خبرگان در این پژوهش

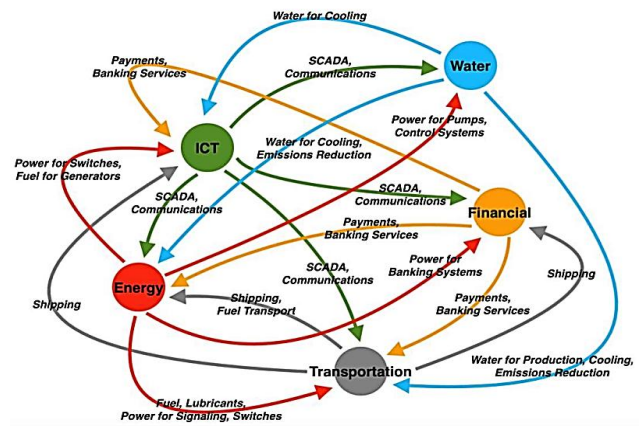
۲- وابستگی زیرساخت‌های حیاتی (پیشینه پژوهش)

زیرساخت‌های حیاتی ستون فقرات جوامع مدرن هستند زیرا به طور منظم برای اطمینان از وجود و عملکرد فعالیت‌های روزانه مورد استفاده قرار می‌گیرند. با این حال، با توجه به تحولات مداوم در فناوری و جهانی شدن جوامع، شبکه‌های زیرساختی از نظر قابلیت عملیاتی به یکدیگر وابسته‌تر می‌شوند. از این رو، این شبکه‌ها در برابر عوامل استرس‌زای خارجی آسیب پذیرتر می‌شوند زیرا در



شکل ۴. انواع رویکردهای مدل‌سازی وابستگی زیرساخت‌های حیاتی [36]

رویکردهای تجربی وابستگی‌های متقابل زیرساخت‌های حیاتی را با توجه به داده‌های سوانح یا فاجعه تاریخی و تجربه کارشناسان تحلیل می‌کنند. مطالعات با این نوع رویکردها می‌توانند الگوهای شکست مکرر و قابل توجه را شناسایی کنند. در رویکردهای عامل محور به دلیل پیچیدگی ذاتی زیرساخت‌های حیاتی و فرآیندهای تصمیم‌گیری مرتبط، زیرساخت‌های حیاتی‌ها معمولاً به عنوان سیستم‌های تطبیقی پیچیده در نظر گرفته می‌شوند لذا برای تجزیه و تحلیل این سیستم‌ها، یک راه موثر، رویکردهای مبتنی بر عامل است، که روشی از پایین به بالا را اتخاذ می‌کند و فرض می‌کند رفتار یا پدیده پیچیده از بسیاری از تعاملات فردی و نسبتاً ساده عوامل مستقل ظاهر می‌شود. رویکردهای مبتنی بر پویایی سیستم از روشی از بالا به پایین برای مدیریت و تجزیه و تحلیل سیستم‌های انطباقی پیچیده که شامل وابستگی‌های متقابل است استفاده می‌کنند. حلقه‌های بازخورد اتصال و جهت اثرات بین اجزای زیرساخت‌های حیاتی را نشان می‌دهد. وابستگی‌های متقابل کشورهای مستقل مشترک المنافع را می‌توان از طریق مدل‌های وابستگی متقابل اقتصادی تحلیل کرد، دو نوع نظری اقتصادی برای مدل‌سازی وابستگی‌های متقابل زیرساخت‌های حیاتی استفاده می‌شود: ورودی-خروجی و تعادل عمومی قابل محاسبه. رویکردهای مبتنی بر شبکه زیرساخت‌های حیاتی جایی که گره‌ها اجزای مختلف زیرساخت‌های حیاتی را نشان می‌دهند و پیوندها از ارتباطات فیزیکی و رابطه‌ای بین آنها تقلید می‌کنند. بسته به مدل‌سازی جریان ذرات در زیرساخت‌های حیاتی، این بخش را بر توپولوژی و



شکل ۳. نمونه‌ای از توصیف کیفی وابستگی‌های متقابل [3]

مطالعات دیگر که عمدتاً به مطالعات مفهومی اشاره دارد که پیچیدگی مدل‌سازی وابستگی‌های متقابل زیرساخت‌های حیاتی را نشان می‌دهد، تلاش‌های تحقیقاتی بعدی بر تعیین معیارها و چارچوب‌های ریاضی در میان زیرساخت‌های حیاتی برای تعیین کمیت تأثیرات آبخاری تمرکز می‌کنند [۲۸]. اوپانگ بررسی جامع مدل‌سازی وابستگی متقابل زیرساخت‌های حیاتی را به عنوان یک موضوع نابالغ، اما به سرعت در حال رشد تعریف می‌کند. یک مشکل عمده فقدان داده‌های در دسترس عموم در مورد زیرساخت‌های حیاتی است که محققان را وادار می‌کند تا تحلیل‌های کیفی را برخلاف تحلیل‌های کمی انجام دهند به عنوان مثال اخیراً، درک زیرساخت‌های حیاتی به عنوان «سیستم سیستم‌ها» [۲۹] و «شبکه شبکه‌ها» [30] منجر به بررسی‌های کمی عمیق‌تر وابستگی‌ها در زیرساخت‌های حیاتی و وابستگی‌های متقابل بین آن‌ها می‌شود. با این حال، پیچیدگی مدل‌سازی وابستگی متقابل اغلب بسیاری از مطالعات را محدود می‌کند که تنها یک زیرساخت یا تعداد کمی از آنها را در نظر بگیرند [۳۱]. نان لی در سال ۲۰۲۲ با ارائه چارچوبی از یک رویکرد شبیه‌سازی مبتنی بر معماری سطح بالا برای مدل‌سازی امنیت زیرساخت‌های وابسته به هم با یکپارچه‌سازی دانش موجود، داده‌ها و مدل‌های خاص برای هر حوزه زیرساخت، و شبیه‌سازی انتشار شکست در سیستم‌ها استفاده کرد [۳۲]. همچنین مین اوپانگ در سال ۲۰۱۳، رویکردهای مدل‌سازی و شبیه‌سازی در تحقیق و عمل وابستگی‌های متقابل زیرساخت را گروه‌بندی و بررسی می‌کند و آنها را در رویکردهای تجربی، رویکردهای مبتنی بر عامل، رویکردهای مبتنی بر پویایی سیستم، رویکردهای مبتنی بر نظریه اقتصادی، رویکردهای مبتنی بر شبکه، و موارد دیگر تقسیم‌بندی می‌نماید. خلاصه آن در شکل ۴ مشاهده می‌شود.

روش‌های مبتنی بر جریان گروه‌بندی می‌کنند. علاوه بر رویکردهای مذکور، رویکردهای دیگری وجود دارد، مانند روش مدل‌سازی هولوگرافیک سلسله مراتبی، روش مبتنی بر معماری سطح بالا و غیره. او همچنین برای مقایسه رویکردهای مختلف، معیارهای

جدول ۷. مقایسه رویکردهای مدل‌سازی وابستگی زیرساخت‌های حیاتی

مقایسه رویکردها از چند معیار						
نوع رویکرد	رویکرد فرعی	کمیت داده های ورودی	قابلیت دسترسی به داده های ورودی	انواع وابستگی های متقابل	هزینه های محاسباتی	بلوغ
تجربی		متوسط، زیاد	متوسط	فیزیکی، سایبری، جغرافیایی و منطقی	کم	متوسط
مبتنی بر عامل		زیاد	کم	فیزیکی، سایبری، جغرافیایی و منطقی	زیاد	زیاد
مبتنی بر پویایی سیستم		زیاد	کم	فیزیکی، سایبری و منطقی	کم	زیاد
مبتنی بر نظریه اقتصادی	ورودی خروجی تعادل عمومی قابل محاسبه	متوسط	زیاد	فیزیکی، سایبری	کم	زیاد
مبتنی بر شبکه توپولوژی	روش مبتنی بر توپولوژی	کم، متوسط	متوسط	فیزیکی، سایبری، جغرافیایی و منطقی	کم، متوسط	زیاد
روش مبتنی بر جریان	روش مبتنی بر جریان	زیاد	کم	فیزیکی، سایبری، جغرافیایی و منطقی	زیاد	زیاد

(۴) انحصاری/یا وابستگی: زمانی که یک زیرساخت قادر به کار با یک زیرساخت دیگر نباشد. و (۵) وابستگی همزمان: زمانی که اجزای دو یا چند زیرساخت در یک مکان قرار دارند.

می‌توان بین چهار و پنج نوع وابستگی متقابل که در بالا مورد بحث قرار گرفت، روابط برقرار کرد. وابستگی‌های فیزیکی و سایبری که وابستگی سیستم اجزای مادی یا غیر مادی را عنوان می‌کند، می‌تواند ارتباط نزدیکی با وابستگی ورودی و متقابل داشته باشد، در حالی که وابستگی جغرافیایی و همزمان هر دو به نزدیکی مکان‌های اجزای سیستم مرتبط هستند. در نهایت، وابستگی مشترک و انحصاری انواع منطقی وابستگی سیستم هستند [۳۴].

۳- اهمیت و لزوم پژوهش

ادبیات، مطالعات گذشته و مدل‌های ارائه شده در غالب رویکردهای گوناگون اطلاعات و دانش خوبی جهت تحلیل، ایجاد و توسعه یک مدل بومی در اختیار می‌گذارند و از آنجایی که انواع زیرساخت‌های حیاتی مبتنی بر مقتضیات ملی، از کشوری به کشور دیگر متفاوت است، مثلاً ایالات متحده بیشترین تعداد زیرساخت حیاتی را دارد، از طرف دیگر میزان و نوع وابستگی هر زیرساخت به زیرساخت‌های دیگر نیز بر مبنای موقعیت هر کشور متفاوت است به عنوان مثال: در مقایسه نوع و میزان وابستگی در زیرساخت برق، بر اساس گزارش

همانگونه که در جدول ۷ مشخص است چهار نوع وابستگی متقابل، در میان سیستم‌های زیرساختی، مشخص شد [۳]

(۱) فیزیکی: زمانی که وابستگی متقابل به دلیل تکیه بر جریان مواد بین دو یا چند سیستم باشد.

(۲) سایبر: زمانی که وابستگی متقابل به دلیل اتکا به انتقال اطلاعات بین دو یا چند سیستم باشد.

(۳) جغرافیایی: زمانی که وابستگی متقابل به دلیل نزدیکی دو یا چند سیستم باشد. و

(۴) منطقی: زمانی که وابستگی متقابل ناشی از عوامل دیگری باشد که در سه دسته فوق قرار نمی‌گیرند.

به طور متناوب، پنج نوع دیگر از وابستگی متقابل زیرساخت ارائه شد [۳۳].

(۱) وابستگی ورودی: زمانی که یک زیرساخت به ورودی از زیرساخت دیگری نیاز دارد.

(۲) وابستگی متقابل: زمانی که حداقل یک فعالیت در یک زیرساخت به فعالیت دیگری از زیرساخت دیگر وابسته است، در حالی که حداقل یک فعالیت در زیرساخت بعدی به فعالیت دیگری از زیرساخت قبلی وابسته است.

(۳) وابستگی مشترک: زمانی که اجزا یا فعالیت‌های فیزیکی بین زیرساخت‌ها به اشتراک گذاشته می‌شود.

۴- سوالات پژوهش

سوالات این پژوهش به شرح ذیل است:

- ۱- کدام زیرساخت‌ها از منظر امنیت سایبری تاثیرگذار و کدام یک تاثیرپذیرتر هستند؟
- ۲- ارتباط زیرساخت‌های حیاتی در کشور جمهوری اسلامی ایران چگونه است؟
- ۳- اهمیت و اولویت زیرساخت‌های حیاتی با رویکرد سایبری کدام است؟
- ساختار کلی پژوهش جهت پاسخ به سوالات آن در شکل ۵ نشان داده شده است.



شکل ۵. نمودار ساختار کلی پژوهش

به جهت شناسایی ارتباط و وابستگی زیرساخت‌های حیاتی در این پژوهش از رویکرد تجربی استفاده شد، به همین منظور به جهت جامعیت پژوهش از هر زیرساخت حیاتی مدیران و متخصصین مربوط به همان زیرساخت انتخاب و در گروه کانونی دسته‌بندی شدند. گروه کانونی به بحث اکتشافی گروهی اطلاق میشود که به منظور به دست آوردن ادراک در خصوص موضوع‌هایی خاص در فضایی تعریف شده صورت می‌گیرد. این روش به طور فزاینده‌ای به عنوان یک ابزار پژوهش در علوم اجتماعی و در ابتدا در جامعه‌شناسی مورد استفاده قرار گرفت. گروه‌های کانونی ممکن

وضعیت صنعت هسته‌ای جهان در سال ۲۰۲۲ (WNISR2022) کشور فرانسه تقریباً دوسوم برق خود را از منابع هسته‌ای تولید می‌کند که این مقدار بیشتر از هر کشور دیگری است، این در حالی است که با بررسی نمودار انرژی در ایران در سال ۲۰۲۲، انرژی هسته‌ای تنها حدود ۱ درصد از تولید برق کشور را به خود اختصاص داده است^۱ و گاز طبیعی بیشترین منبع تولید برق در ایران است. این تفاوت باعث تغییر در نوع و میزان وابستگی زیرساخت برق به زیرساخت هسته‌ای یا زیرساخت گاز است. از منظر دیگرناتجی و خروجی‌های این پژوهش می‌تواند به مدیران ارشد حوزه‌های زیرساختی کشور در اتخاذ تصمیمات راهبردی کلان کمک کند، به عنوان مثال مهمترین دلایل خاموشی‌های گسترده در سال ۱۴۰۰ (زیرساخت برق) افت تولید برقایی‌ها در حوزه (زیرساخت آب) به دلیل خالی شدن مخازن سدها عنوان شد، از طرفی در دی ماه سال ۱۴۰۱ شاهد مشکلات حوزه گاز (زیرساخت گاز) در کشور بودیم که منجر به ایجاد نارضایتی و قطع گاز صنایع و مشترکین در برخی استان‌ها شد، لذا با بررسی و تحلیل دقیق جریان انرژی در کشور و تجربیات سال‌های گذشته و بررسی میزان تاثیر پذیری و تاثیرگذاری هر زیرساخت بر زیرساخت دیگر میتوان راهبردهای مناسبی را اتخاذ نمود به عنوان مثال سهم گاز را در تولید برق کاهش داد و سهم تولید برق را از انرژی‌های تجدید پذیر یا انرژی هسته‌ای (زیرساخت هسته-ای) افزایش داد یا نگاه ویژه‌ای به موضوع تولید برقایی با توجه به وضعیت آب در کشور در سال‌های اخیر نمود که به طبع آن اهمیت، نوع و میزان اثر گذاری و اثر پذیری هر زیرساخت تغییر می‌نماید. این موضوع زمانی اهمیت بیشتری دارد که به این مهم توجه نماییم که تعداد حملات سایبری‌ای که در سال‌های اخیر متوجه سرویس‌ها و زیرساخت‌های ایرانی شده به میزان قابل توجهی رسیده است. اما نکته‌ای که در این میان قابل توجه است این است که حملاتی که از دو سال قبل تا به امروز انجام شده فقط یک حمله سایبری معمولی نیستند بلکه می‌توان گفت ایران در حوزه امنیت سایبری از همراه سال ۹۹ وارد یک جنگ سایبری شده است. به عنوان نمونه در ۲۸ آبان ۱۳۹۹ حمله به شرکت ملی گاز ایران و ۲۲ فروردین ۱۴۰۰ حمله به تجهیزات هسته‌ای نظنز را میتوان نام برد^۲.

همچنین بند ۴ طرح راهبردی حفاظت از زیرساخت‌های حیاتی کشور مصوب ۱۴۰۱/۶/۲۹ در (شورای عالی) پدافند غیرعامل کشور، موضوع وابستگی زیرساخت‌های حیاتی به یکدیگر مورد توجه قرار گیرد.

³ <https://cert.ir>

¹ World Nuclear Industry Status Report 2022 (WNISR2022)

^۲ <https://pep.moe.gov.ir>

است شامل دو نفر، سه نفر، چهار تا شش نفر (گروه بسیار کوچک)، هفت تا ده نفر (گروه کوچک) یا یازده تا بیست نفر (گروه بزرگ) شرکت‌کننده باشد در گروه کانونی، افراد شرکت‌کننده برداشت‌ها، احساسات و تجربه‌های خود را به اشتراک می‌گذارند؛ بنابراین محدوده دیدگاه‌ها درباره موضوعات خاص گسترش می‌یابد و از مشکلات ناشی از تمایلات یکطرفه جلوگیری می‌شود [۳۵].

جدول ۸. توصیف خبرگان ۱۱ گروه کانونی پژوهش

تعداد جلسات	تحصیلات		تخصص	شرکت کنندگان	زیرساخت	
۳	دکتری	۴ نفر	مهندسی برق مخابرات، مهندسی فناوری اطلاعات مهندسی کامپیوتر، مدیریت فناوری اطلاعات	۶	ارتباطات و فناوری اطلاعات	۱
	کارشناسی ارشد	۲ نفر				
۲	دکتری	۱ نفر	اقتصاد، مهندسی سیستم‌های اقتصادی، مدیریت مالی	۴	مالی	۲
	کارشناسی ارشد	۳ نفر				
۲	کارشناسی ارشد	۳ نفر	رسانه، مدیریت صنعتی، تولید	۳	رسانه	۳
۳	دکتری	۲ نفر	چشم پزشکی فلوشیپ شبکه، ارتوپدی فوق تخصص جراحی زانو، امنیت شبکه	۳	سلامت	۴
	کارشناسی ارشد	۱ نفر				
۱	دکتری	۱ نفر	شهرسازی گرایش مدیریت شهری، عمران، برنامه ریزی حمل و نقل	۵	حمل و نقل	۵
	کارشناسی ارشد	۴ نفر				
۲	دکتری	۲ نفر	مندسی برق قدرت، مدیریت دولتی	۳	برق	۶
	کارشناسی ارشد	۱ نفر				
۱	دکتری	-	مهندسی صنایع، مهندسی ایمنی	۳	گاز	۷
	کارشناسی ارشد	۳ نفر				
۱	دکتری	-	مهندسی نفت استخراج و مخازن، مهندسی صنایع، مهندسی فناوری اطلاعات	۶	نفت	۸
	ارشد	۶ نفر				
۲	دکتری	۱ نفر	مهندسی شیمی، مهندسی صنایع	۲	پتروشیمی	۹
	کارشناسی ارشد	۱ نفر				
۱	دکتری	-	مهندسی منابع آب مکانیک هیدرولیک مجاری روباز	۳	آب	۱۰
	کارشناسی ارشد	۳ نفر				
۱	دکتری	۱ نفر	مکانیک، صنایع	۲	اتمی (هسته‌ای)	۱۱
	کارشناسی ارشد	۱ نفر				
۴۰ نفر			مجموع شرکت کنندگان			

۵- روش شناسی

۵-۱- تکنیک دیمتل

روش ارزیابی تصمیم‌گیری DEMATEL برای اولین بار توسط Fontela و Gabus در سال ۱۹۷۶ مورد استفاده قرار گرفت و توانسته است با در نظر گرفتن قضاوت متخصصان بسیاری از مشکلات پیچیده جهانی را در حوزه‌های علمی، سیاسی و اقتصادی حل کند. موسسه BMI از روش DEMATEL برای اجرای پروژه‌های بزرگ و پیچیده استفاده کرد. روش DEMATEL در ژاپن محبوب‌تر شده است، زیرا یک تکنیک گسترده است که قادر است تمام روابط علت و معلولی در هم تنیده را در هر مدل ساختاری ارزیابی و فرموله کند. DEMATEL یک روش جامع برای ساخت و تجزیه و تحلیل یک مدل ساختاری است که شامل روابط علی بین عوامل پیچیده است. مراحل اساسی DEMATEL شامل ساخت سلسله مراتب ارزیابی،

در یک سیستم دارای وابستگی داخلی، تمام معیارهای سیستم‌ها دو به دو مشابه، مستقیم یا غیرمستقیم هستند. بنابراین، هر رابطه داخلی با یکی از معیارها روی سایر معیارها نیز تأثیر می‌گذارد. از این رو، پیدا کردن اولویت در عمل کار بسیار دشواری است. روش DEMATEL بر مبنای تئوری گراف، ما را قادر می‌سازد تا مسائل را بهتر برنامه ریزی و حل کنیم؛ به نحوی که ممکن است چندین معیار را در گروه علت معلول برای درک بهتر روابط علی، در جهت ترسیم نقشه روابط شبکه‌ای تقسیم کنیم. این روش شناسی ممکن است تأییدکننده روابط متقابل میان متغیر معیارها و محدودکننده روابطی باشد که در یک روند توسعه‌ای و سیستماتیک ضروری هستند محصول نهائی فرآیند DEMATEL ارائه تصویری است که پاسخگو بر اساس آن فعالیت‌های خود را سازمان می‌دهد.

$$a_{ij} = \frac{1}{M} \sum_{k=1}^M x_{ij}^k$$

قبل از آغاز مرحله بعدیه منظور بررسی اعتبار و پایایی ماتریس میانگین از فرمول ذیل استفاده می‌نماییم. که باید عددی کمتر از ۰/۰۵ باشد.

مرحله ۳: ماتریس مستقیم اولیه نرمال شده D ساخته می‌شود.

$$D = A \times S$$

$$T = (I - D)^{-1} \quad \text{where } S = \min\left(\frac{1}{\max \sum_{j=1}^n a_{ij}}, \frac{1}{\max \sum_{i=1}^n a_{ij}}\right)$$

هر عنصر در ماتریس D بین صفر و ۱ قرار دارد.

$$R_i = \sum_{j=1}^n t_{ij} \quad (2)$$

$$C_j = \sum_{i=1}^n t_{ij}$$

$R_i + C_j$ is called the "Prominence" and $R_i - C_j$ is called the "Relation".

مرحله ۴: ماتریس رابطه کل T ایجاد می‌شود. I ماتریس هویت است.

مرحله ۵: امتیاز (برتری) و رابطه برای هر معیار محاسبه می‌شود.

مرحله ۶: نمودار علت و معلول با نگاهت مجموعه داده ایجاد می‌شود.

$$(R_i + C_j : R_i - C_j) \quad (3)$$

۶- تحلیل داده‌ها و یافته‌ها

پس از تشکیل جلسه‌های گروه کانونی خبرگان حوزه زیرساخت‌های حیاتی و با بحث و بررسی پیرامون مبانی نظری و تجربه‌های کاربردی و بر اساس روش DEMATEL با کمک ۱۱ نفر از اعضای کارگروه تاثیر میزان هر زیرساخت بر زیرساخت دیگر بر اساس ماتریس (۱۱*۱۱) جمع آوری شده و میانگین ماتریس Z به شرح جدول ۹ احصاء شد.

در این مرحله جهت بررسی اعتبار و پایایی ماتریس میانگین، ادغام نظرات ۱۰ نفر و ماتریس ادغام نظرات ۹ نفر را بر اساس فرمول ذیل

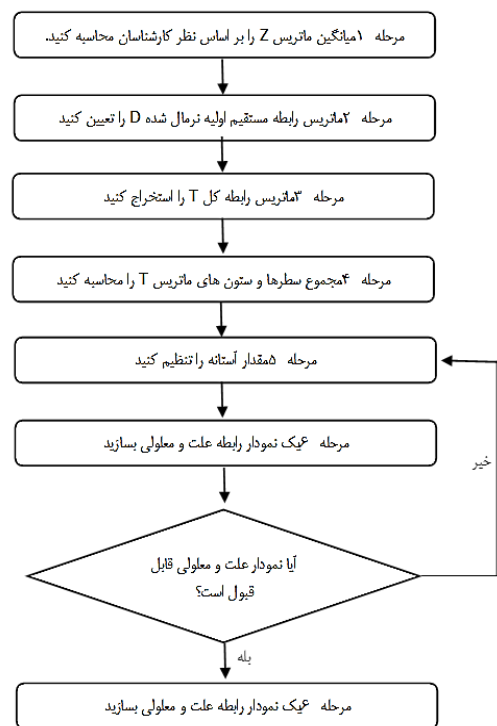
$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{|g_c^{ij\rho} - g_c^{ij(\rho-1)}|}{g_c^{ij\rho}} \times 100\% \quad (4)$$

عمل شد عدد حاصل برابر با ۰/۴۹۴ شد که کمتر از ۰/۰۵ می‌باشد و پایایی ماتریس Z را تایید می‌نماید. در ادامه با محاسبه ماتریس نرمال و کسر ماتریس I (ماتریس واحد) از آن، ماتریس حاصله را معکوس نموده و ماتریس T را از ضرب ماتریس نرمال در ماتریس معکوس بدست می‌آوریم جدول (۱۲) تا (۱۵).

انتخاب تیم خبره، محاسبه ماتریس کل رابطه، تعیین درجه تاثیر است و $R + C$ و $R - C$ برای ترسیم نقشه تاثیر رابطه کل محاسبه می‌شود.

مرحله ۱: در پانل گروه‌های کانونی و ابعاد / معیارها تعیین می‌شود. در این مرحله، گروهی از کارشناسان برای جمع آوری نظرات ذهنی انتخاب می‌شوند. بر اساس ادبیات و نظر کارشناسان، چالش‌ها تعیین و بحث می‌شود.

مرحله ۲: ماتریس رابطه مستقیم اولیه ساخته می‌شود. این ماتریس با انجام مقایسه‌های زوجی بین چالش‌ها/معیارها ایجاد می‌شود. برای ارزیابی رابطه بین معیارها از مقیاس زیر استفاده می‌شود.



مراحل DEMATEL: 2021, j. Thakkar

شکل ۶. مراحل گام به گام DEMATEL

$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{|g_c^{ij\rho} - g_c^{ij(\rho-1)}|}{g_c^{ij\rho}} \times 100\% \quad (1)$$

صفر: بدون تاثیر. ۱: تاثیر بسیار کم. ۲: تاثیر کم. ۳: تاثیر بالا. ۴: تاثیر بسیار بالا.

$i = j$, the x_{ij}^k نشان دهنده میزان تأثیر معیار i بر معیار j است که توسط متخصص k مشاهده می‌شود. اجازه دهید، M متخصص و n معیار وجود داشته باشد. ماتریس متوسط A با فرمول زیر محاسبه می‌شود:

جدول ۹. میانگین ماتریس Z

ماتریس میانگین Z	فازا	مالی	سلامت	حمل و نقل	برق	آب	نفت	گاز	پتروشیمی	هسته ای	رسانه
فازا	۰	۳/۸	۳/۱	۳/۴	۳/۱	۲/۷	۲/۶	۲/۸	۳/۵	۱/۶	۳/۸
مالی	۲/۳	۰	۱/۸	۲/۲	۱/۸	۰/۹	۰/۹	۱/۱	۱/۷	۰/۱	۲/۸
سلامت	۰/۴	۱/۱	۰	۲	۰/۴	۰/۶	۰/۳	۰/۳	۰/۳	۰	۰/۹
حمل و نقل	۱/۷	۲/۷	۲	۰	۱/۳	۰/۴	۲/۱	۲/۱	۳	۰/۱	۲
برق	۲/۶	۳/۱	۳	۳/۲	۰	۱/۶	۱/۴	۱/۲	۲/۹	۱/۴	۳/۱
آب	۰/۱	۰/۴	۲/۷	۰/۵	۱/۶	۰	۰/۳	۱/۴	۱/۷	۱/۱	۰/۳
نفت	۰/۳	۱/۴	۰/۳	۲/۹	۲/۱	۰/۲	۰	۱/۵	۳/۷	۰	۰/۸
گاز	۰/۳	۱/۱	۱	۳/۱	۴	۰/۹	۱/۲	۰	۲/۳	۰	۰/۵
پتروشیمی	۱/۹	۲/۷	۱/۴	۲/۶	۱/۴	۰/۳	۱/۶	۰/۹	۰	۰/۷	۰/۸
هسته ای	۰	۰/۱	۱/۱	۰	۱/۱	۰	۰	۰/۲	۰	۰	۰
رسانه	۲/۶	۳	۲/۴	۲/۴	۲/۱	۱/۸	۱/۶	۱/۹	۱/۹	۱/۶	۰

جدول ۱۰. ماتریس نرمال

ماتریس نرمال	ذخایا	مالی	سلامت	حمل و نقل	برق	آب	نفت	گاز	پتروشیمی	هسته ای	رسانه
ذخایا	۰/۰۰۰۰	-/۰۱۲۵۰	-/۰۱۰۲۰	-/۰۱۱۱۸	-/۰۱۰۲۰	-/۰۸۸۸	-/۰۸۵۵	-/۰۹۲۱	-/۰۱۱۵۱	-/۰۵۲۶	-/۰۱۲۵۰
مالی	-/۰۷۵۷	۱/۰۰۰۰	-/۰۵۹۲	-/۰۷۲۴	-/۰۵۹۲	-/۰۲۹۶	-/۰۲۹۶	-/۰۳۶۲	-/۰۵۵۹	-/۰۰۲۳	-/۰۹۲۱
سلامت	-/۰۱۳۲	-/۰۳۶۲	۱/۰۰۰۰	-/۰۶۵۸	-/۰۱۳۲	-/۰۱۹۷	-/۰۰۹۹	-/۰۰۹۹	-/۰۰۹۹	۱/۰۰۰۰	-/۰۳۶۲
حمل و نقل	-/۰۵۵۹	-/۰۸۸۸	-/۰۶۵۸	۱/۰۰۰۰	-/۰۴۲۸	-/۰۱۳۲	-/۰۶۹۱	-/۰۶۹۱	-/۰۹۸۷	-/۰۰۲۳	-/۰۶۵۸
برق	-/۰۸۵۵	-/۰۱۰۲۰	-/۰۹۸۷	-/۰۱۰۵۳	۱/۰۰۰۰	-/۰۵۲۶	-/۰۴۶۱	-/۰۳۶۵	-/۰۵۴	-/۰۴۶۱	-/۰۱۰۲۰
آب	-/۰۰۲۳	-/۰۱۳۲	-/۰۸۸۸	-/۰۱۶۴	-/۰۵۲۶	۱/۰۰۰۰	-/۰۰۹۹	-/۰۴۶۱	-/۰۵۵۹	-/۰۳۶۲	-/۰۰۹۹
نفت	-/۰۰۹۹	-/۰۴۶۱	-/۰۰۹۹	-/۰۹۵۴	-/۰۶۹۱	-/۰۰۶۶	۱/۰۰۰۰	-/۰۴۹۳	-/۰۱۲۱۷	۱/۰۰۰۰	-/۰۲۶۲
گاز	-/۰۰۹۹	-/۰۳۶۲	-/۰۳۶۲	-/۰۱۰۲۰	-/۰۱۳۱۶	-/۰۲۹۶	-/۰۳۶۵	۱/۰۰۰۰	-/۰۷۵۷	۱/۰۰۰۰	-/۰۱۶۴
پتروشیمی	-/۰۶۲۵	-/۰۸۸۸	-/۰۴۶۱	-/۰۸۵۵	-/۰۴۶۱	-/۰۰۹۹	-/۰۵۲۶	-/۰۲۹۶	۱/۰۰۰۰	-/۰۲۳۰	-/۰۲۶۲
هسته ای	۱/۰۰۰۰	-/۰۰۲۳	-/۰۳۶۲	۱/۰۰۰۰	-/۰۳۶۲	۱/۰۰۰۰	۱/۰۰۰۰	-/۰۰۶۶	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
رسانه	-/۰۸۵۵	-/۰۹۸۷	-/۰۷۸۹	-/۰۷۸۹	-/۰۶۹۱	-/۰۵۹۲	-/۰۵۲۶	-/۰۶۲۵	-/۰۶۲۵	-/۰۵۲۶	۱/۰۰۰۰

جدول ۱۱. ماتریس I-N

ماتریس I-N	ذخایا	مالی	سلامت	حمل و نقل	برق	آب	نفت	گاز	پتروشیمی	هسته ای	رسانه
ذخایا	۱/۰۰۰۰	-/۰۱۲۵۰	-/۰۱۰۲۰	-/۰۱۱۱۸	-/۰۱۰۲۰	-/۰۸۸۸	-/۰۸۵۵	-/۰۹۲۱	-/۰۱۱۵۱	-/۰۵۲۶	-/۰۱۲۵۰
مالی	-/۰۷۵۷	۱/۰۰۰۰	-/۰۵۹۲	-/۰۷۲۴	-/۰۵۹۲	-/۰۲۹۶	-/۰۲۹۶	-/۰۳۶۲	-/۰۵۵۹	-/۰۰۲۳	-/۰۹۲۱
سلامت	-/۰۱۳۲	-/۰۳۶۲	۱/۰۰۰۰	-/۰۶۵۸	-/۰۱۳۲	-/۰۱۹۷	-/۰۰۹۹	-/۰۰۹۹	-/۰۰۹۹	۱/۰۰۰۰	-/۰۳۶۲
حمل و نقل	-/۰۵۵۹	-/۰۸۸۸	-/۰۶۵۸	۱/۰۰۰۰	-/۰۴۲۸	-/۰۱۳۲	-/۰۶۹۱	-/۰۶۹۱	-/۰۹۸۷	-/۰۰۲۳	-/۰۶۵۸
برق	-/۰۸۵۵	-/۰۱۰۲۰	-/۰۹۸۷	-/۰۱۰۵۳	۱/۰۰۰۰	-/۰۵۲۶	-/۰۴۶۱	-/۰۳۶۵	-/۰۵۴	-/۰۴۶۱	-/۰۱۰۲۰
آب	-/۰۰۲۳	-/۰۱۳۲	-/۰۸۸۸	-/۰۱۶۴	-/۰۵۲۶	۱/۰۰۰۰	-/۰۰۹۹	-/۰۴۶۱	-/۰۵۵۹	-/۰۳۶۲	-/۰۰۹۹
نفت	-/۰۰۹۹	-/۰۴۶۱	-/۰۰۹۹	-/۰۹۵۴	-/۰۶۹۱	-/۰۰۶۶	۱/۰۰۰۰	-/۰۴۹۳	-/۰۱۲۱۷	۱/۰۰۰۰	-/۰۲۶۲
گاز	-/۰۰۹۹	-/۰۳۶۲	-/۰۳۶۲	-/۰۱۰۲۰	-/۰۱۳۱۶	-/۰۲۹۶	-/۰۳۶۵	۱/۰۰۰۰	-/۰۷۵۷	۱/۰۰۰۰	-/۰۱۶۴
پتروشیمی	-/۰۶۲۵	-/۰۸۸۸	-/۰۴۶۱	-/۰۸۵۵	-/۰۴۶۱	-/۰۰۹۹	-/۰۵۲۶	-/۰۲۹۶	۱/۰۰۰۰	-/۰۲۳۰	-/۰۲۶۲
هسته ای	۱/۰۰۰۰	-/۰۰۲۳	-/۰۳۶۲	۱/۰۰۰۰	-/۰۳۶۲	۱/۰۰۰۰	۱/۰۰۰۰	-/۰۰۶۶	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
رسانه	-/۰۸۵۵	-/۰۹۸۷	-/۰۷۸۹	-/۰۷۸۹	-/۰۶۹۱	-/۰۵۹۲	-/۰۵۲۶	-/۰۶۲۵	-/۰۶۲۵	-/۰۵۲۶	۱/۰۰۰۰

جدول ۱۲. معکوس ماتریس I-N

معکوس I-N	ذغال	مالی	سلامت	حمل و نقل	برق	آب	نفت	گاز	پتروشیمی	هسته‌ای	رسانه
ذغال	۱/۰۹۰۹	-/۰۲۴۶۶	-/۰۲۱۴۷	-/۰۲۴۹۴	-/۰۲۱۰۳	-/۰۱۴۴۳	-/۰۱۶۲۰	-/۰۱۷۳۰	-/۰۲۴۰۷	-/۰۰۹۱۰	-/۰۲۱۸۲
مالی	-/۰۱۲۵۶	۱/۰۸۰۹	-/۰۱۲۷۸	-/۰۱۵۴۴	-/۰۱۲۳۲	-/۰۰۶۷۵	-/۰۰۷۹۳	-/۰۰۸۸۰	-/۰۱۲۲۶	-/۰۰۲۹۷	-/۰۱۴۹۵
سلامت	-/۰۰۳۵۲	-/۰۰۶۶۱	۱/۰۲۸۵	-/۰۰۹۵۰	-/۰۰۳۹۱	-/۰۰۳۳۴	-/۰۰۳۰۱	-/۰۰۳۱۶	-/۰۰۴۲۳	-/۰۰۰۹۲	-/۰۰۵۳۹
حمل و نقل	-/۰۱۰۹۶	-/۰۱۶۵۹	-/۰۱۳۳۰	۱/۰۹۴۳	-/۰۱۱۳۹	-/۰۰۵۱۲	-/۰۱۱۷۶	-/۰۱۱۹۰	-/۰۱۷۶۲	-/۰۰۲۷۷	-/۰۱۲۶۷
برق	-/۰۱۵۲۹	-/۰۲۰۱۶	-/۰۱۸۸۶	-/۰۲۱۲۹	۱/۰۹۰۴	-/۰۰۹۹۹	-/۰۱۱۰۸	-/۰۱۰۸۵	-/۰۱۹۴۲	-/۰۰۷۷۱	-/۰۱۷۹۳
آب	-/۰۰۲۹۳	-/۰۰۵۱۲	-/۰۱۲۰۰	-/۰۰۶۱۱	-/۰۰۸۳۳	۱/۰۱۷۱	-/۰۰۳۲۷	-/۰۰۶۷۱	-/۰۰۸۹۷	-/۰۰۴۶۷	-/۰۰۳۸۸
نفت	-/۰۰۵۷۹	-/۰۱۱۱۸	-/۰۰۶۶۴	-/۰۱۶۲۳	-/۰۱۲۰۳	-/۰۰۳۴۰	۱/۰۴۳۱	-/۰۰۸۹۳	-/۰۱۸۱۴	-/۰۰۱۸۹	-/۰۰۷۶۵
گاز	-/۰۰۶۱۳	-/۰۱۰۷۷	-/۰۰۹۵۷	-/۰۱۷۳۸	-/۰۱۸۰۴	-/۰۰۵۹۶	-/۰۰۸۳۱	۱/۰۴۵۵	-/۰۱۴۵۶	-/۰۰۲۱۹	-/۰۰۷۴۰
پتروشیمی	-/۰۱۰۶۵	-/۰۱۵۲۷	-/۰۱۰۵۰	-/۰۱۵۶۴	-/۰۱۰۳۲	-/۰۰۴۱۹	-/۰۰۹۴۴	-/۰۰۷۵۱	۱/۰۷۱۹	-/۰۰۴۲۰	-/۰۰۸۳۶
هسته‌ای	-/۰۰۰۷۶	-/۰۰۱۴۰	-/۰۰۴۵۱	-/۰۰۱۲۸	-/۰۰۴۳۵	-/۰۰۰۵۴	-/۰۰۰۵۹	-/۰۰۱۲۲	-/۰۰۱۰۰	۱/۰۰۳۴	-/۰۰۰۹۴
رسانه	-/۰۱۴۳۳	-/۰۱۸۵۹	-/۰۱۶۱۲	-/۰۱۷۸۶	-/۰۱۴۸۸	-/۰۱۰۱۲	-/۰۱۰۹۲	-/۰۱۲۲۲	-/۰۱۵۵۸	-/۰۰۷۹۵	۱/۰۰۷۶۷

جدول ۱۳. ماتریس T

ارتباطات کل T	فاوا	مالی	سلامت	حمل و نقل	برق	آب	نفت	گاز	پتروشیمی	هسته‌ای	رسانه
فاوا	-/۰۰۹۰۹	-/۰۲۴۶۶	-/۰۲۱۴۷	-/۰۲۴۹۴	-/۰۲۱۰۳	-/۰۱۴۴۳	-/۰۱۶۲۰	-/۰۱۷۳۰	-/۰۲۴۰۷	-/۰۰۹۱۰	-/۰۲۱۸۲
مالی	-/۰۱۲۵۶	-/۰۰۸۰۹	-/۰۱۲۷۸	-/۰۱۵۴۴	-/۰۱۲۳۲	-/۰۰۶۷۵	-/۰۰۷۹۳	-/۰۰۸۸۰	-/۰۱۲۲۶	-/۰۰۲۹۷	-/۰۱۴۹۵
سلامت	-/۰۰۳۵۲	-/۰۰۶۶۱	-/۰۰۲۸۵	-/۰۰۹۵۰	-/۰۰۳۹۱	-/۰۰۳۳۴	-/۰۰۳۰۱	-/۰۰۳۱۶	-/۰۰۴۲۳	-/۰۰۰۹۲	-/۰۰۵۳۹
حمل و نقل	-/۰۱۰۹۶	-/۰۱۶۵۹	-/۰۱۳۳۰	-/۰۰۹۴۳	-/۰۱۱۳۹	-/۰۰۵۱۲	-/۰۱۱۷۶	-/۰۱۱۹۰	-/۰۱۷۶۲	-/۰۰۲۷۷	-/۰۱۲۶۷
برق	-/۰۱۵۲۹	-/۰۲۰۱۶	-/۰۱۸۸۶	-/۰۲۱۲۹	-/۰۰۹۰۴	-/۰۰۹۹۹	-/۰۱۱۰۸	-/۰۱۰۸۵	-/۰۱۹۴۲	-/۰۰۷۷۱	-/۰۱۷۹۳
آب	-/۰۰۲۹۳	-/۰۰۵۱۲	-/۰۱۲۰۰	-/۰۰۶۱۱	-/۰۰۸۳۳	-/۰۰۱۷۱	-/۰۰۳۲۷	-/۰۰۶۷۱	-/۰۰۸۹۷	-/۰۰۴۶۷	-/۰۰۳۸۸
نفت	-/۰۰۵۷۹	-/۰۱۱۱۸	-/۰۰۶۶۴	-/۰۱۶۲۳	-/۰۱۲۰۳	-/۰۰۳۴۰	-/۰۰۴۳۱	-/۰۰۸۹۳	-/۰۱۸۱۴	-/۰۰۱۸۹	-/۰۰۷۶۵
گاز	-/۰۰۶۱۳	-/۰۱۰۷۷	-/۰۰۹۵۷	-/۰۱۷۳۸	-/۰۱۸۰۴	-/۰۰۵۹۶	-/۰۰۸۳۱	-/۰۰۴۵۵	-/۰۱۴۵۶	-/۰۰۲۱۹	-/۰۰۷۴۰
پتروشیمی	-/۰۱۰۶۵	-/۰۱۵۲۷	-/۰۱۰۵۰	-/۰۱۵۶۴	-/۰۱۰۳۲	-/۰۰۴۱۹	-/۰۰۹۴۴	-/۰۰۷۵۱	-/۰۰۷۱۹	-/۰۰۴۲۰	-/۰۰۸۳۶
هسته‌ای	-/۰۰۰۷۶	-/۰۰۱۴۰	-/۰۰۴۵۱	-/۰۰۱۲۸	-/۰۰۴۳۵	-/۰۰۰۵۴	-/۰۰۰۵۹	-/۰۰۱۲۲	-/۰۰۱۰۰	۱/۰۰۳۴	-/۰۰۰۹۴
رسانه	-/۰۱۴۳۳	-/۰۱۸۵۹	-/۰۱۶۱۲	-/۰۱۷۸۶	-/۰۱۴۸۸	-/۰۱۰۱۲	-/۰۱۰۹۲	-/۰۱۲۲۲	-/۰۱۵۵۸	-/۰۰۷۹۵	-/۰۰۷۶۷

حیاتی را دارد و همچنین روابط علت و معلولی احتمالی بین آن‌ها با کمک از ماتریس T موارد مذکور را در جدول ۱۴ استخراج می‌نماییم.

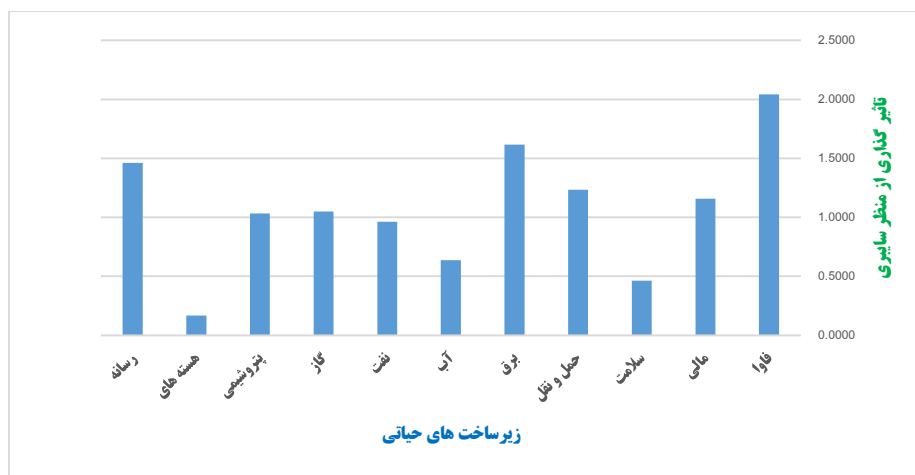
جهت مشخص شدن تاثیرگذارترین زیرساخت، تاثیرپذیرترین زیرساخت، زیرساختی که بیشترین ارتباط با دیگر زیرساخت‌های

جدول ۱۴. تحلیل ماتریس ارتباطات کل

زیرساخت	D	R	D+R	D-R
فاوا	۲/۰۴۱	۰/۹۲۰	۲/۹۶۱	۱/۱۲۱
مالی	۱/۱۵۸	۱/۳۸۴	۲/۵۴۲	-/۰۲۲۶
سلامت	-/۰۴۶۴	۱/۲۸۶	۱/۷۵۰	-/۰۸۲۱
حمل و نقل	۱/۲۳۵	۱/۵۵۱	۲/۷۸۶	-/۰۳۱۶
برق	۱/۶۱۶	۱/۲۵۵	۲/۸۷۲	-/۰۳۶۱
آب	-/۰۶۳۷	-/۰۶۵۶	-/۰۱۲۹۳	-/۰۰۱۹
نفت	-/۰۹۶۲	-/۰۸۹۹	-/۰۱۸۶۱	-/۰۰۹۴
گاز	۱/۰۴۹	۰/۹۳۱	۱/۹۸۰	-/۰۱۱۷
پتروشیمی	۱/۰۳۲	۱/۴۴۰	۲/۴۷۲	-/۰۴۰۸
هسته‌ای(اتمی)	-/۰۱۶۸	-/۰۴۴۷	-/۰۶۱۵	-/۰۲۲۷۹
رسانه	۱/۴۶۳	۱/۰۸۷	۲/۵۴۹	-/۰۳۷۶

تاثیرگذارترین زیرساخت حیاتی: زیرساخت فناوری اطلاعات و ارتباطات (فاوا) است بعد از آن زیرساخت برق در رتبه دوم زیرساخت رسانه در رتبه سوم زیرساخت حمل و نقل در رتبه چهارم زیرساخت مالی در رتبه پنجم زیرساخت گاز و پتروشیمی به ترتیب رتبه ششم و هفتم و زیرساخت نفت در رتبه هشتم زیرساخت آب در رتبه نهم و زیرساخت سلامت و هسته‌ای در رتبه‌های دهم و یازدهم قرار گرفتند. این رتبه‌بندی در نمودار شکل ۷ به نمایش در آمده است.

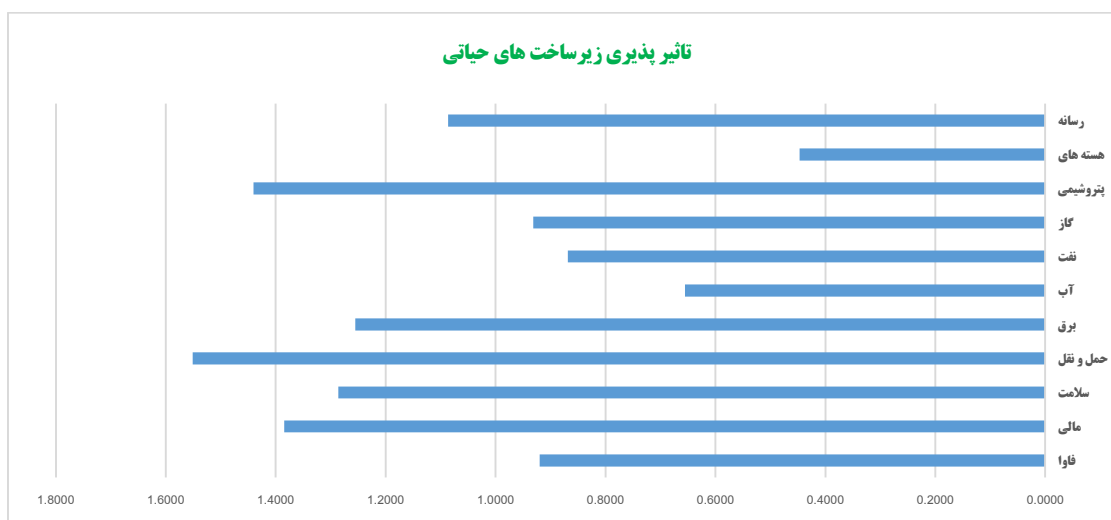
D: مجموع سطرها (بزرگترین عدد بدست آمده نشان‌دهنده تاثیرگذارترین زیرساخت حیاتی).
R: مجموع ستون‌ها (بزرگترین عدد نشان دهنده تاثیرپذیرترین زیرساخت حیاتی).
D+R: بزرگترین عدد نشان دهنده زیرساخت حیاتی که بیشترین ارتباط را با دیگر زیرساخت‌های حیاتی دارد.
D-R: اعداد مثبت به عنوان معیارهای (زیرساخت‌های) علت و اعداد منفی به عنوان معیارهای (زیرساخت‌های) معلول. بر اساس نتایج حاصله



شکل ۷. نمودار رتبه‌بندی زیرساخت‌های حیاتی از نظر تاثیرگذاری

رسانه و گاز نیز به ترتیب در رتبه‌های پنجم تا هفتم زیرساخت فاوا در رتبه هشتم و زیرساخت نفت و آب و هسته‌ای نیز در رتبه‌های نهم تا یازدهم قرار گرفتند این رتبه‌بندی نیز در نمودار شکل ۸ به نمایش در آمده است.

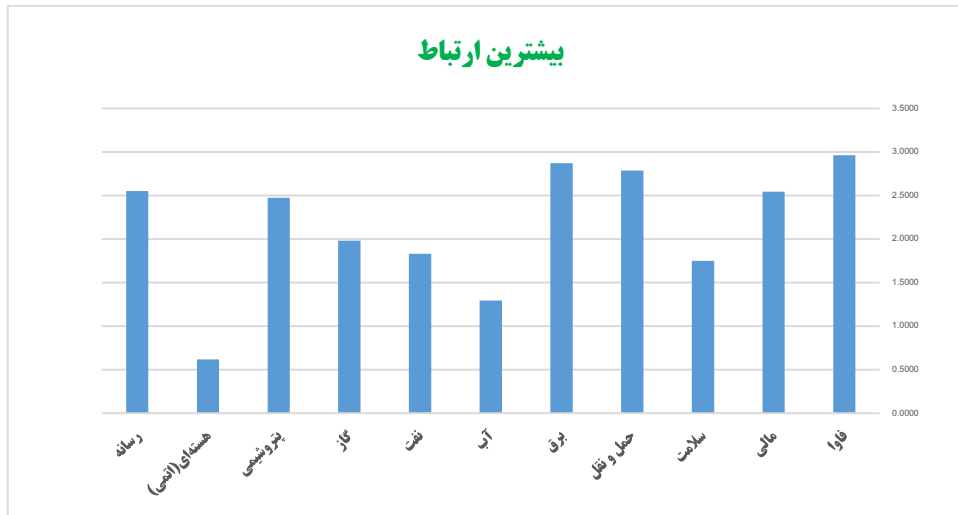
تاثیرپذیرترین زیرساخت حیاتی: زیرساخت حمل و نقل است و بعد از آن زیرساخت پتروشیمی در رتبه دوم و زیرساخت مالی و سلامت به ترتیب در رتبه‌های سوم و چهارم و زیرساخت‌های برق و



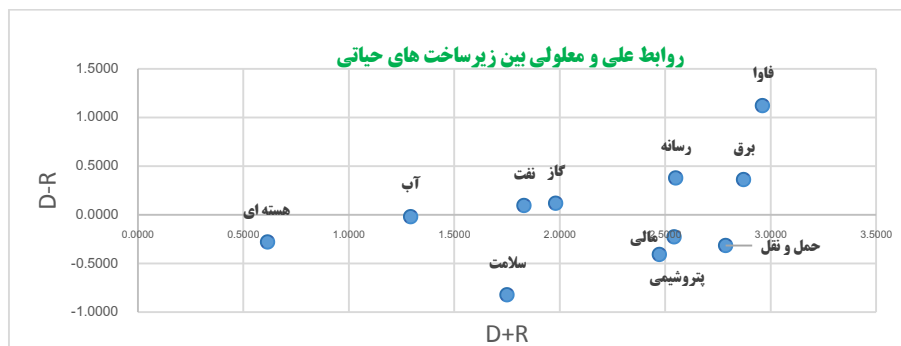
شکل ۸. رتبه‌بندی زیرساخت‌های حیاتی از نظر تاثیرپذیری

ارتباط را دارا بودند این رتبه‌بندی از ستون سوم جدول ۱۴ احصا شده است و در نمودار شکل ۹ این رتبه‌بندی به نمایش درآمده است همچنین همانگونه که در نمودار اسکرین ذیل مشخص است زیرساخت فاوا، برق رسانه گاز، نفت و آب علت و زیرساخت‌های حمل و نقل، مالی، پتروشیمی، هسته‌ای و سلامت معلول هستند

نتیجه دیگری که از جدول ۱۴ قابل استنباط است شناخت زیرساخت‌های حیاتی است که دارای بیشترین ارتباط با دیگر زیرساخت‌ها هستند در این تحلیل نیز زیرساخت فاوا از منظر سایبری بیشترین ارتباط را با دیگر زیرساخت‌های حیاتی دارا بود همچنین زیرساخت برق و حمل و نقل بعد از زیرساخت فاوا بیشترین



شکل ۹. نمودار رتبه‌بندی زیرساخت‌های حیاتی از نظر دارا بودن بیشترین ارتباط با دیگر زیرساخت‌ها



شکل ۱۰. نمودار اسکرین

عدد آستانه کمتر بود ۰ (صفر) منظور نموده به معنی عدم ارتباط است و هر عددی که بیش از عدد آستانه بود نشان دهنده وجود ارتباط بین معیارها (زیرساخت‌های حیاتی) است و خود عدد را به شرح ماتریس جدول ۱۵ منظور می‌نماییم.

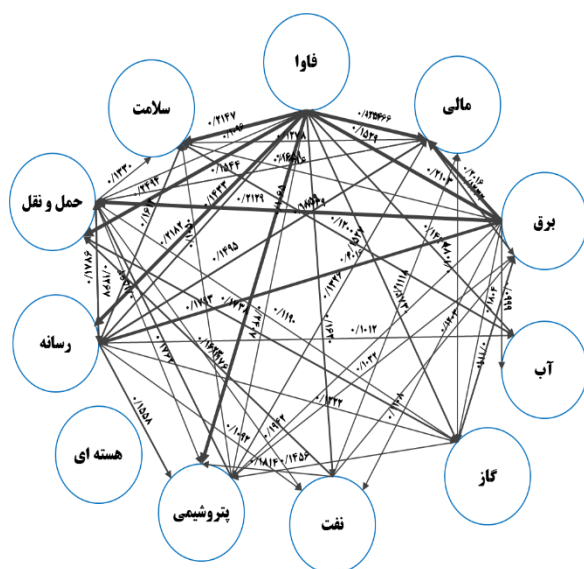
در ادامه تحلیل داده‌ها، جهت مشخص نمودن روابط بین زیرساخت‌های حیاتی ابتدا با محاسبه میانگین داده‌های ماتریس T عدد آستانه را بدست آورده و ماتریس ارتباط را ایجاد می‌نماییم بدین صورت که در ماتریس ارتباط هر عددی در ماتریس T را که از

جدول ۱۵. ماتریس ارتباط

ماتریس ارتباط	فاوا	مالی	سلامت	حمل و نقل	برق	آب	نفت	گاز	پتروشیمی	هسته ای	رسانه
فاوا	۰	۰/۲۴۶۶	۰/۲۱۴۷	۰/۲۴۹۴	۰/۲۱۰۳	۰/۱۴۴۳	۰/۱۶۲۰	۰/۱۷۳۰	۰/۲۴۰۷	۰	۰/۲۱۸۲
مالی	۰/۱۲۵۶	۰	۰/۱۲۷۸	۰/۱۵۴۴	۰/۱۲۳۲	۰	۰	۰	۰/۱۲۳۶	۰	۰/۱۴۹۵
سلامت	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
حمل و نقل	۰/۱۰۹۶	۰/۱۶۵۹	۰/۱۲۳۰	۰	۰/۱۱۳۹	۰	۰/۱۱۷۶	۰/۱۱۹۰	۰/۱۷۶۲	۰	۰/۱۲۶۷
برق	۰/۱۵۲۹	۰/۲۰۱۶	۰/۱۸۸۶	۰/۲۱۲۹	۰	۰/۰۹۹۹	۰/۱۱۰۸	۰/۱۰۸۵	۰/۱۹۴۲	۰	۰/۱۷۹۳
آب	۰	۰	۰/۱۲۰۰	۰	۰	۰	۰	۰	۰	۰	۰
نفت	۰	۰/۱۱۱۸	۰	۰/۱۶۲۳	۰/۱۲۰۳	۰	۰	۰	۰/۱۸۱۴	۰	۰
گاز	۰	۰/۱۰۷۷	۰	۰/۱۷۳۸	۰/۱۸۰۴	۰	۰	۰	۰/۱۴۵۶	۰	۰
پتروشیمی	۰/۱۰۶۵	۰/۱۵۲۷	۰/۱۰۵۰	۰/۱۵۶۴	۰/۱۰۳۲	۰	۰	۰	۰	۰	۰
هسته ای	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
رسانه	۰/۱۴۳۳	۰/۱۸۵۹	۰/۱۶۱۲	۰/۱۷۸۶	۰/۱۴۸۸	۰/۱۰۱۲	۰/۱۰۹۲	۰/۱۲۲۲	۰/۱۵۵۸	۰	۰

در شکل ۱۸ ارتباطات زیرساخت گاز نمایش داده شده است. زیرساخت برق و حمل و نقل به ترتیب بیشترین وابستگی را به زیرساخت گاز دارند. در شکل ۱۹ نیز ارتباطات زیرساخت نفت و وابستگی دیگر زیرساخت‌ها به آن به نمایش در آمده است، زیرساخت پتروشیمی بیشترین وابستگی را به آن دارد.

در شکل ۲۰ ارتباطات زیرساخت آب نمایش داده شده است. زیرساخت سلامت تنها زیرساختی است که به آن وابستگی را نشان می‌دهد.

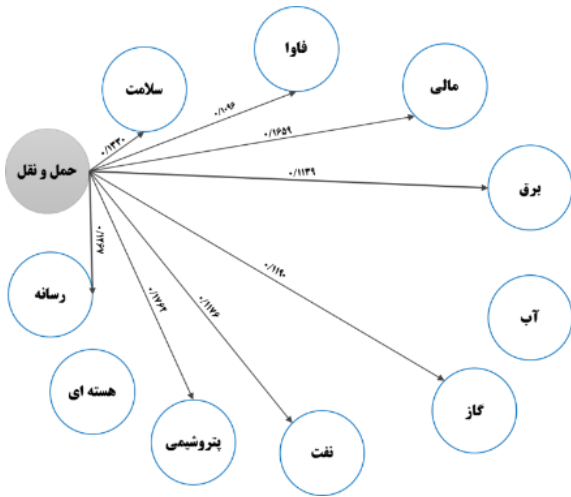


شکل ۱۱. ارتباط بین زیرساخت‌های حیاتی

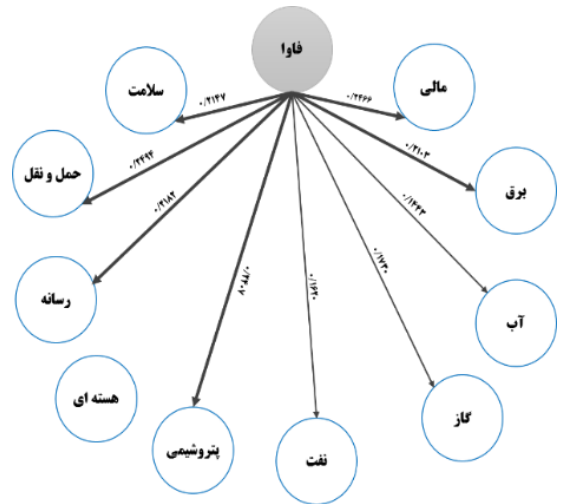
ارتباط بین زیرساخت‌های حیاتی و تحلیل نتایج ماتریس ارتباط برای هر یک از زیرساخت‌های حیاتی و مجموع ارتباط آن‌ها در شکل ۱۱ ارائه شده است. به منظور نمایش بهتر و تحلیل این ارتباطات و وابستگی بین زیرساخت‌ها ارتباط هر زیرساخت به طور مجزا در اشکال ۱۲ الی ۲۰ ارائه شده است.

همانگونه که در شکل ۱۲ ارتباطات زیرساخت فاوا نمایش داده شده است و وابستگی زیرساخت‌های سلامت، حمل و نقل، پتروشیمی، رسانه، برق، مالی به زیرساخت فاوا بیشتر است و زیرساخت نفت، گاز و آب وابستگی کمتری دارند به همین ترتیب در شکل ۱۳ ارتباطات زیرساخت رسانه و وابستگی دیگر زیرساخت‌ها به آن به نمایش در آمده است. در شکل ۱۴ ارتباطات زیرساخت برق نمایش داده شده است و وابستگی زیرساخت‌های دیگر به آن به روشنی نمایش داده شده است. ضمناً این زیرساخت با زیرساخت هسته‌ای ارتباطی ندارد. در شکل ۱۵ نیز ارتباطات زیرساخت حمل و نقل و وابستگی دیگر زیرساخت‌ها به آن به نمایش در آمده است، این زیرساخت ارتباطی با زیرساخت آب و هسته‌ای ندارد و به مفهوم دیگر زیرساخت آب و هسته‌ای به آن وابسته نیستند.

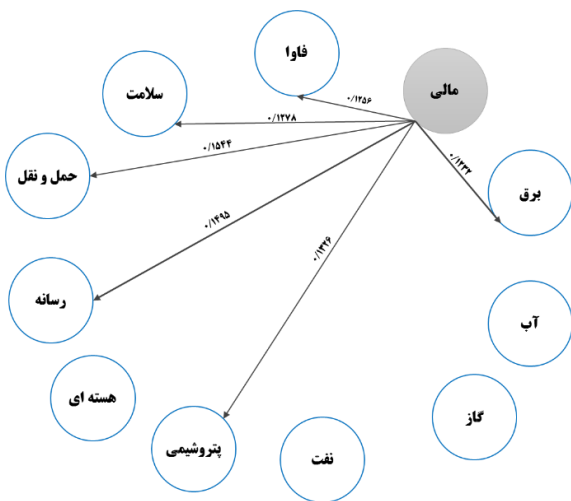
در شکل ۱۶ ارتباطات زیرساخت مالی نمایش داده شده است و وابستگی زیرساخت‌های دیگر به آن به روشنی نمایش داده شده است. ضمناً این زیرساخت با زیرساخت‌های آب، گاز، نفت و هسته‌ای ارتباطی ندارد. در شکل ۱۷ نیز ارتباطات زیرساخت پتروشیمی و وابستگی دیگر زیرساخت‌ها به آن به نمایش در آمده است، این زیرساخت بیشترین ارتباط را با زیرساخت حمل و نقل داشته و ارتباطی با زیرساخت آب، گاز نفت و هسته‌ای ندارد و به مفهوم دیگر زیرساخت‌های اشاره شده وابسته به آن نیستند.



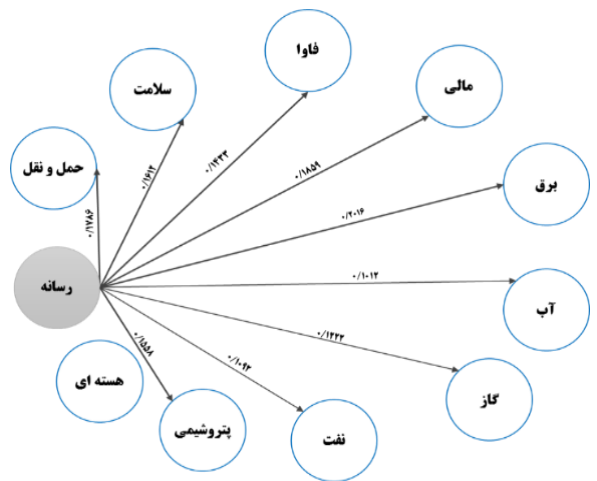
شکل ۱۵. ارتباطات و وابستگی زیرساخت حمل و نقل



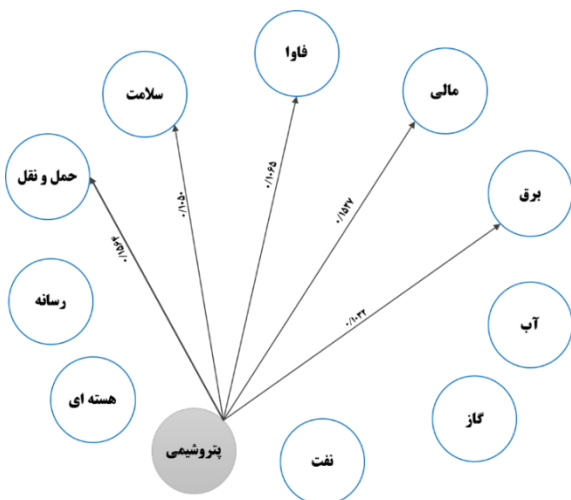
شکل ۱۲. ارتباطات و وابستگی زیر ساخت فایا



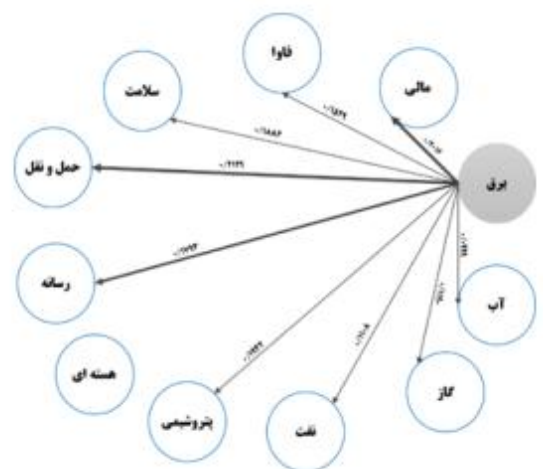
شکل ۱۶. ارتباطات و وابستگی زیرساخت مالی



شکل ۱۳. ارتباطات و وابستگی زیرساخت رسانه



شکل ۱۷. ارتباطات و وابستگی زیرساخت پتروشیمی



شکل ۱۴. ارتباطات و وابستگی زیرساخت برق

۷- نتیجه‌گیری و پیشنهاد

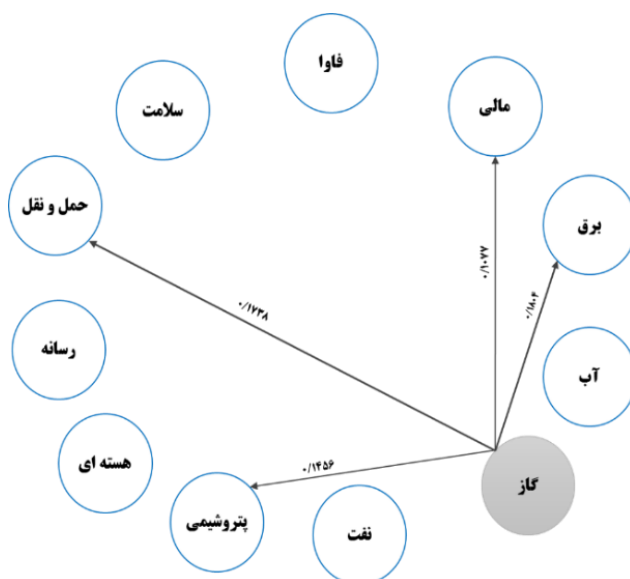
در خصوص شناسایی زیرساخت‌های حیاتی بعد از مطالعات تطبیقی با دیگر کشورها زیرساخت‌ها از منظر امنیت سایبری با تشکیل پنل خبرگی احصاء شد با تطابق خروجی‌های پنل با اسناد بالادست و سیاست‌های کلی نظام مشخص شد زیرساخت رسانه می‌بایست به عنوان یک زیرساخت حیاتی از منظر امنیت سایبری مورد توجه ویژه قرار گیرد همچنین خروجی جلسات کانونی و تحلیل نتایج پژوهش نشان می‌دهد که زیرساخت رسانه به عنوان سومین زیرساخت بعد از زیرساخت‌های فاوا و برق جزو تاثیرگذارترین زیرساخت‌ها به شمار می‌آید و همچنین در تاثیرپذیری رتبه ششم را در بین یازده زیرساخت معرفی شده داراست و از نظر ارتباط با دیگر زیرساخت‌های حیاتی در رتبه چهارم قرار دارد این موارد خود تایید و تبیین‌کننده دغدغه مقام معظم رهبری مدظله‌العالی در خصوص رسانه و جنگ رسانه‌ای در فضای مجازی است.

همچنین بر مبنای یافته‌ها زیرساخت فناوری اطلاعات و ارتباطات (فاوا) به عنوان تاثیرگذارترین زیرساخت حیاتی از منظر سایبری بر دیگر زیرساخت‌های حیاتی شناخته شد، زیرساخت برق در حوزه انرژی بعد از آن جزو تاثیرگذارترین زیرساخت‌ها قرار دارد همچنین تاثیرپذیرترین زیرساخت حیاتی حمل و نقل معرفی شد و بعد از آن زیرساخت پتروشیمی در رتبه دوم قرار گرفت.

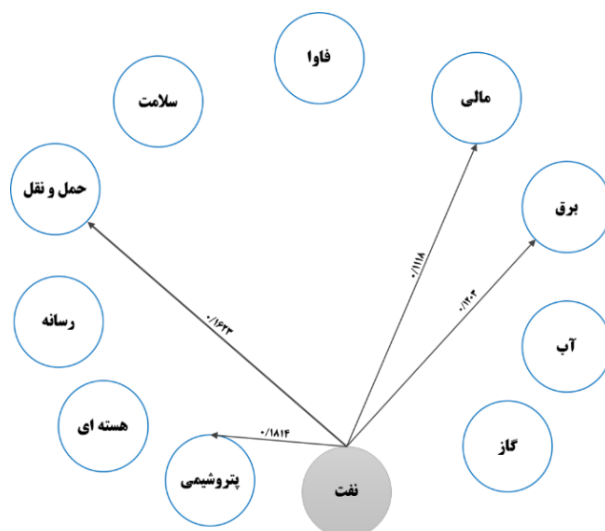
در این پژوهش کلیه ارتباط و وابستگی‌های بین زیرساخت‌های حیاتی احصاء شد و زیرساخت هسته‌ای و سلامت هیچ ارتباطی با دیگر زیرساخت‌های حیاتی از منظر امنیت سایبری نداشتند.

نتایج و خروجی‌های این پژوهش می‌تواند مدیران ارشد حوزه‌های زیرساختی کشور را در اتخاذ تصمیمات راهبردی کلان یاری نماید.

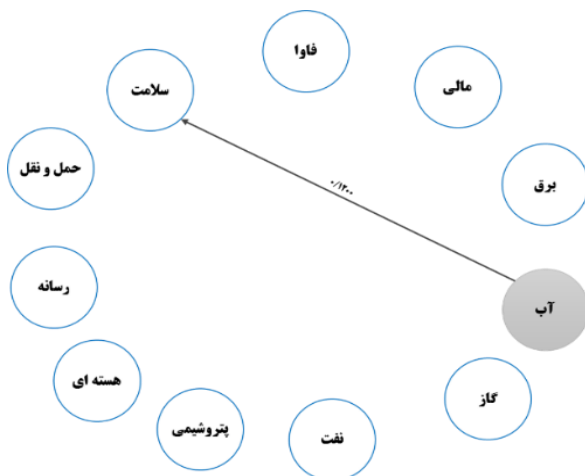
پیشنهاد می‌گردد زیرساخت رسانه جزو زیرساخت‌های حیاتی از منظر امنیت سایبری لحاظ گردد همچنین خروجی و نتایج این پژوهش می‌تواند در طراحی، ایجاد و راه‌اندازی سامانه ملی اشتراک‌گذاری هشدار به منظور احصاء آگاهی وضعیت ملی در حوزه سایبری و دیگر تحقیقات متکی بر وابستگی زیرساخت‌های حیاتی که در طرح مصوب راهبردی حفاظت از زیرساخت‌های کشور ایران نیز به صراحت به آن اشاره شده است، مفید واقع شود. اما از آنجایی که تهدیدات و حملات سایبری بر روی دارایی‌های زیرساخت‌های حیاتی صورت می‌گیرد، شناخت، دسته‌بندی و اولویت‌بندی دارایی‌های زیرساخت‌های حیاتی، نگاشت این دارایی‌ها بر خدمات علاوه بر شناخت خدمات و مأموریت‌های زیرساخت‌های حیاتی ضرورت پیدا می‌کند این مهم می‌تواند منجر به شناسایی بهتر



شکل ۱۸. ارتباطات و وابستگی زیرساخت گاز



شکل ۱۹. ارتباطات و وابستگی زیرساخت نفت



شکل ۲۰. ارتباطات و وابستگی زیرساخت آب

- ISO/IEC 27002 for process control systems specific to the energy utility industry. 2013.
- [17] IETF RFC449 Internet Security Glossary 2: <https://tools.ietf.org/html/rfc449>.
- [18] UK.: Centre for the Protection of National Infrastructure (CPNI)
- [19] NATO: Tallinn Manual on the International Law Applicable to Cyber Warfare. 2013.
- [20] Canada: An Emergency Management Framework for Canada (Second Edition).
- [21] Cuba : Glossary of Cyber terms/Glosario de términos, Centro de Seguridad del CISberespaCISO.
- [22] FRG, G., National strategy for critical infrastructure protection. 2009: p. 1-18.
- [23] Brussels, B., Commission of the European Communities. Retrieved from <http://eur-lex>.
- [24] National Cybersecurity Strategy - Towards A Secure Cyberspace 2020-2023. 2020.
- [25] Danish Cyber and Information Security Strategy (2022-2024).
- [26] European Parliament and of the Council 2022.
- [27] Karakoc, D.B., K. Barker, and A.D. González, Analyzing the tradeoff between vulnerability and recoverability investments for interdependent infrastructure networks. *Socio-Economic Planning Sciences*, 2023. **87**: p. 101508.
- [28] Zimmerman, R. Decision-making and the vulnerability of interdependent critical infrastructure. in 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583). 2004. IEEE.
- [29] Eusgeld, I., C. Nan, and S. Dietz, "System-of-systems" approach for interdependent critical infrastructures. *Reliability Engineering & System Safety*, 2011. **96**(6): p. 679-686.
- [30] GAO, U., Critical infrastructure protection: Challenges and efforts to secure control systems. US GAO, 2004.
- [31] Eusgeld, I. and C. Nan. Creating a simulation environment for critical infrastructure interdependencies study. in 2009 IEEE International Conference on Industrial Engineering and Engineering Management. 2009. IEEE.
- [32] Wang, F., J.J. Magoua, and N. Li, Modeling cascading failure of interdependent critical infrastructure systems using HLA-based co-simulation. *Automation in Construction*, 2022. **133**: p. 104008.
- [33] Lee II, E.E., J.E. Mitchell, and W.A. Wallace, Restoration of services in interdependent infrastructure systems: A network flows approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2007. **37**(6): p. 1303-1317.
- [34] Haggag, M., et al., Resilient cities critical infrastructure interdependence: a meta-research. *Sustainable and resilient infrastructure*, 2022. **7**(4): p. 291-312.
- [35] Fisher, E., What practitioners consider to be the skills and behaviours of an effective people project manager. *International journal of project management*, 2011. **29**(8): p. 994-1002.
- [36] Min Ouyang Review on modeling and simulation of interdependent critical infrastructure systems, *Reliability Engineering & System Safety* Volume 121, 2014, Pages 43-60, ISSN 0951-8320, <https://doi.org/10.1016/j.res.2013.06.040>.
- و دقیق‌تر ارتباط بین زیرساخت‌های حیاتی گردد لذا پیشنهاد می‌گردد الگویی تدوین شود تا بوسیله آن زیرساخت‌های حیاتی بتوانند دارائی‌های خود را شناسائی نموده و نگرانی بین دارائی‌ها و خدماتی که ارائه می‌نمایند حاصل گردد این نکته نیز حائز اهمیت است که اطلاعات دارائی‌های زیرساخت‌های حیاتی جزء اسناد طبقه‌بندی شده کشورها می‌باشد و محدودیتی به جهت پژوهش از این روش فراهم می‌نماید.
- ## مراجع
- [1] Ziring, N. NATIONAL CYBER RESILIENCE AND ROLES FOR PUBLIC AND PRIVATE SECTOR STAKEHOLDERS. in International Conference on Critical Infrastructure Protection. 2022. Springer.
- [2] Cyber Security and Infrastructure Security Agency, Critical Infrastructure Sectors, Arlington, Virginia (www.dhs.gov/CISsa/critical-infrastructure-sectors). 2020.
- [3] Rinaldi, S.M., J.P. Peerenboom, and T.K. Kelly, Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE control systems magazine*, 2001. **21**(6): p. 11-25.
- [4] Thissen, W. and P. Herder, Synthesis of Approaches and Insights: Conclusions and Research Agenda. *Critical Infrastructures State of the Art in Research and Application*, 2003: p. 283-300.
- [5] Clinton, W.J., Executive order 13010-critical infrastructure protection. *Federal register*, 1996. **61**(138): p. 37347-37350.
- [6] Council, E., Communication from the commission to the councils and the European Parliament: Critical infrastructure protection in the fight against terrorism. 2004: p. 1-11.
- [7] India : workshop presentation by the NATIONAL CRITICAL INFORMATION INFRASTRUCTURE PROTECTION CENTRE (NCISIPC), . 2015.
- [8] Israel, [https://ironscales.com/blog-how-machine-learning-can-stop-phishing-attacks-critical-infrastructure/\(2023\)](https://ironscales.com/blog-how-machine-learning-can-stop-phishing-attacks-critical-infrastructure/(2023)). 2023.
- [9] Japan, The Information Security Policy Council, The Second Action Plan on Information Security Measures for Critical Infrastructures, Japan (2009). 2009.
- [10] Arabia, K.o.S., Developing National Information Security Strategy for the Kingdom of Saudi Arabia NISS draft 7. 2022.
- [11] Gheorghie, A., et al., Critical infrastructures at risk. *Securing the European electric power system*, 2006.
- [12] ITU Study Group Q.22/1 Report on Best Practices for a National Approach to Cybersecurity: A Management Framework for Organizing National Cybersecurity Efforts, ITU-D Secretariat, Geneva. 2008.
- [13] Australian :Critical Infrastructure Resilience Strategy. 2010.
- [14] QATAR National Cyber Security Strategy. May 2014.
- [15] NIST Glossary/ NIST SP 800-30 / CNSSI 4009-2015. 2015.
- [16] ISO/IEC TR 27019:2013 Information technology -- Security techniques -- Information security management guidelines based on

Providing A Lightweight Encryption Solution to Secure Data in Internet Of Things

Wahab Aminiazar^{1*}, Rasoul Farahi² and Fatemeh Dashti³

¹ Department of Computer Engineering and Information Technology, Faculty of Electrical and Computer Engineering, Islamic Azad University, Mahabad Branch, Mahabad, Iran

² Department of Computer Engineering and Information Technology, Faculty of Electrical and Computer Engineering, Islamic Azad University, Mahabad Branch, Mahabad, Iran

³ PhD Student in Computer Science, Tabriz Electricity Distribution Company, Tabriz, Iran

Received: 04 May 2024, Revised: 16 August 2024, Accepted: 31 August 2024

Paper type: Research

Abstract

In order to use the Internet of Things as a secure infrastructure, there are various challenges and problems, of which security is one of the most important. Establishing security in such networks has many complications due to the limitation of various resources, including processing resources and low energy, and there is a need to establish a kind of compromise between security and available resources. These conditions have caused security to become an important challenge in these networks, and various methods have been presented to improve and optimize this challenge. Accordingly, in this article, a lightweight encryption solution based on symmetric and asymmetric encryption is presented to ensure data security on the Internet of Things. In the proposed method, first, the main data is encrypted by the symmetric Bluefish algorithm, and then its key is secured with the help of the elliptic curve encryption algorithm, so that as a result, data security can be ensured in a short time and with high security in infrastructures based on the Internet of Things provided. In the end, the proposed solution has been evaluated through the Eclipse simulator and by testing on the data volume of 20 to 1000 kilobytes. The simulation results show that the proposed method performs more optimally compared to other encryption algorithms in terms of evaluation criteria such as execution time and encryption and decryption throughput. These results indicate that the proposed solution, while establishing security, has had the least negative impact on the processing resources of IoT nodes.

Keywords: Internet of things, security, Lightweight Encryption Algorithms, Elliptic Curve Encryption Algorithms.

* Corresponding Author's email: Aminiazar@iau-mahabad.ac.ir

ارائه یک راهکار رمزنگاری سبک وزن به منظور امنیت داده در اینترنت اشیا

وهاب امینی‌آذر^{۱*}، رسول فرحی^۲، فاطمه دشتی^۳

^۱ دانشکده مهندسی برق و کامپیوتر، دانشگاه آزاد اسلامی واحد مهاباد، مهاباد، ایران

^۲ دانشکده مهندسی برق و کامپیوتر، دانشگاه آزاد اسلامی واحد مهاباد، مهاباد، ایران

^۳ دانشجوی دکترای کامپیوتر، شرکت توزیع نیروی برق تبریز، تبریز، ایران

تاریخ دریافت: ۱۴۰۳/۰۲/۱۵ تاریخ بازبینی: ۱۴۰۳/۰۵/۲۶ تاریخ پذیرش: ۱۴۰۳/۰۶/۱۰

نوع مقاله: پژوهشی

چکیده

جهت بکارگیری اینترنت اشیا به عنوان یک زیرساخت مطمئن، چالش‌های و مشکلات مختلفی وجود دارد که امنیت یکی از مهم‌ترین آنها می‌باشد. برقراری امنیت در چنین شبکه‌هایی با توجه به محدودیت منابع مختلف از جمله منابع پردازشی و انرژی پایین، دارای پیچیدگی‌های زیادی است و نیاز به برقراری یک نوع مصالحه بین امنیت و منابع در دسترس است. این شرایط موجب شده است تا امنیت به یک چالش مهم در این شبکه‌ها تبدیل شود و روش‌های مختلفی نیز برای بهبود و بهینگی این چالش ارائه شده است. بر همین اساس در این مقاله یک راهکار رمزنگاری سبک‌وزن مبتنی بر رمزنگاری متقارن و نامتقارن جهت تأمین امنیت داده در اینترنت اشیا ارائه شده است. در روش پیشنهادی در ابتدا داده اصلی توسط الگوریتم متقارن بلوفیش رمزنگاری می‌شود و سپس کلید آن به کمک الگوریتم رمزنگاری خم‌های بیضوی ایمن‌سازی می‌شود تا در نتیجه بتوان در زمان کم و با امنیت بالا امنیت داده را در زیرساخت‌های مبتنی بر اینترنت اشیا تأمین کرد. در انتها راهکار پیشنهادی، از طریق شبیه‌ساز Eclipse و با آزمایش بر روی حجم داده ۲۰ تا ۱۰۰۰ کیلوبایت مورد ارزیابی قرار داده شده است. نتایج حاصل از شبیه‌سازی نشان می‌دهد که روش پیشنهادی در مقایسه با سایر الگوریتم‌های رمزنگاری از نظر معیارهای ارزیابی هم چون زمان اجرا و توان عملیاتی رمزنگاری و رمزگشایی بهینه‌تر عمل می‌نماید. این نتایج؛ بیانگر آن است که راهکار پیشنهادی ضمن برقراری امنیت، کمترین تأثیر منفی را بر روی منابع پردازشی گره‌های IoT داشته است.

کلیدواژه‌گان: اینترنت اشیا، امنیت، الگوریتم‌های رمزنگاری سبک وزن، الگوریتم‌های رمزنگاری خم بیضوی.

* رایانامه نویسنده مسؤول: Aminiazar@iau-mahabad.ac.ir

۱- مقدمه

راهکارهای سنتی به سادگی قابل پیش‌بینی و حل نیستند. علاوه بر این، باید تأکید شود که بیشتر رویکردهای امنیتی بر پایه معماری‌های متمرکز تکیه می‌کنند، که در اینترنت اشیا به دلیل تعداد زیاد اشیا، استفاده از آنها به طرز چشمگیری پیچیده‌تر می‌شود [۵]. بر همین اساس در این مقاله به عنوان نوآوری یک راهکار رمزنگاری ترکیبی و سبک‌وزن، مبتنی بر الگوریتم بلوفیش و خم‌های بیضوی ارائه می‌شود بطوریکه از طریق آن علاوه بر بهبود امنیت در اینترنت اشیا بتوان زمان اجرا و توان عملیاتی را نیز بهینه نمود. بر همین اساس در راهکار پیشنهادی ابتدا داده اصلی توسط الگوریتم رمزنگاری متقارن بلوفیش که زمان اجرای پایینی دارد رمزنگاری می‌شود، سپس به منظور ایمن‌سازی کلید آن در فرایند ارسال و دریافت، از الگوریتم رمزنگاری خم‌های بیضوی استفاده می‌شود. از این طریق می‌توان یک مصالحه بین امنیت و کاهش مصرف منابع در فرایند رمزنگاری بوجود آورد.

این مقاله از پنج بخش تشکیل شده است. در بخش بعدی، خلاصه‌ای از روش‌های قبلی رمزنگاری به منظور تأمین امنیت داده در اینترنت اشیا بررسی می‌شود. بخش سوم، جزئیات الگوریتم پیشنهادی را ارائه می‌کند. نتایج شبیه‌سازی و ارزیابی روش پیشنهادی در بخش چهارم شرح داده می‌شود و در نهایت بخش پنجم به نتیجه‌گیری و کارهای آتی می‌پردازد.

۲- کارهای مرتبط

فراگا و همکاران [۶] یک چارچوب رمزنگاری با در نظر گرفتن پیچیدگی عملیات XOR و چکیده‌ساز برای حفاظت از داده‌ها در فناوری اینترنت اشیا در رایانش مه ارائه داده‌اند. در این پروتکل پیشنهادی، از رمزنگاری‌های پیچیده استفاده نشده است و از فرایند رمزنگاری فقط برای انتقال داده‌های محرمانه یا به منظور اهداف تأیید کاربر و داده‌ها استفاده می‌شود. برای اطمینان از اینکه پروتکل پیشنهادی امن است، توابع چکیده‌ساز استفاده می‌شوند. در نهایت، پس از اعمال معروف‌ترین حملاتی مانند BFU^1 بر روی الگوریتم پیشنهادی، نتایج نشان دادند که عملکرد و امنیت آن بالا است.

موسوی و همکاران [۷] یک رویکرد ترکیبی رمزنگاری مبتنی بر رمزنگاری منحنی بیضوی (ECC) برای سیستم‌های آبیاری هوشمند مبتنی بر اینترنت اشیا را ارائه داده‌اند. نتایج ارزیابی نشان می‌دهد که این رویکرد ترکیبی در مقابل حملات MiM^2 ایمن بوده و عملکرد بهتری نسبت به سایر الگوریتم‌های رمزنگاری دارد.

اینترنت اشیا یک ادغام از حسگرها و اشیاء مختلف است که بدون دخالت انسان می‌توانند مستقیماً با یکدیگر ارتباط برقرار کنند. «اشیا» در «اینترنت اشیا» شامل دستگاه‌های فیزیکی مانند حسگرها هستند که انواع داده‌ها را درباره ماشین‌ها و زندگی انسان‌ها جمع‌آوری و نظارت می‌کنند [۱]. ظهور اینترنت اشیا منجر به اتصال همیشگی و جهانی افراد، اشیا، حسگرها و خدمات شده است. هدف اصلی اینترنت اشیا ارائه زیرساخت شبکه‌ای با پروتکل‌ها و نرم‌افزارهای ارتباطی قابل تعامل برای اتصال و یکپارچه‌سازی حسگرهای فیزیکی/مجازی، رایانه‌های شخصی، دستگاه‌های هوشمند، خودروها و اشیاء مانند یخچال، ماشین ظرفشویی، فر و غذا و دارو را در هر زمان و هر شبکه امکان‌پذیر کند [۱] و [۲]. توسعه فناوری تلفن همراه اجازه می‌دهد تا اشیاء بی‌شماری از طریق حسگرهای مختلف تلفن همراه به بخشی از اینترنت اشیا تبدیل شوند. با این حال، نیازهای موردنیاز برای استقرار گسترده اینترنت اشیا به سرعت در حال افزایش است که منجر به نگرانی جدی امنیتی می‌شود. مسائل امنیتی مانند حریم خصوصی، اعتبارسنجی، تأیید هویت، کنترل دسترسی، پیکربندی سیستم، ذخیره و مدیریت اطلاعات، چالش‌های اصلی در محیط اینترنت اشیا هستند [۲]. به‌عنوان مثال، برنامه‌های اینترنت اشیا مانند تلفن همراه و دستگاه‌های جاسازی‌شده، به فراهم کردن یک محیط دیجیتال برای اتصال جهانی که با حساسیت، سازگاری و واکنش‌گری نسبت به نیازهای انسانی عمل می‌کند، کمک می‌کنند. با این حال، امنیت تضمین نمی‌شود و حریم خصوصی کاربران ممکن است به خطر بیافتد و اطلاعات آنها هنگامی که ارتباط کاربر متوقف یا دچار اختلال شود، قابل نفوذ گردد. برای بهره‌گیری حداکثری از اینترنت اشیا، این مسئله باید موردتوجه قرار گیرد تا اعتماد کاربران در امر حریم خصوصی و کنترل اطلاعات شخصی فراهم شود [۳]. در واقع توسعه اینترنت اشیا به‌طور قابل توجهی به حل مسائل امنیتی وابسته است. در سال‌های اخیر، بسیاری از تحقیقات به منظور مقابله با چالش‌های امنیتی مرتبط با اینترنت اشیا، مانند مسائل مدیریت کلید [۴]، حریم خصوصی، صحت داده، حفظ محرمانگی و امنیت صورت گرفته است. اکثر پژوهش‌های پیشین تلاش کرده‌اند تا راهکارهای امنیتی ارائه شده برای شبکه‌های بی‌سیم حسگری و اینترنت را در زمینه اینترنت اشیا انطباق دهند. با این حال، باید به این نکته اشاره کرد که چالش‌های اینترنت اشیا ابعاد جدیدی دارند و با استفاده از

¹ Brute Force Attack

² Man-in-the-Middle Attack

زو و همکاران [۱۵] پروتکل احراز هویت سه عاملی سبک و ناشناس مورد بررسی قرار دادند. نتایج آزمایشات نشان‌دهنده این می‌باشد که رویکرد پیشنهادی دارای محرمانگی است اما می‌تواند در برابر حملات آفلاین حدس زدن رمز عبور و همگام سازی آسیب‌پذیر باشد. یک پروتکل ساده احراز هویت متقابل و توافق کلید توسط سونی و سینگ در [۱۶] ارائه شده است. که این پروتکل بین بیماران و سرورهای مراکز درمانی اجرا می‌شود تا داده‌های بیمار به طور امن روی سرور ذخیره نماید. این پروتکل از تابع زنجیره‌ای هش و XOR استفاده می‌کند، اما در برابر حملات آفلاین حدس زدن رمز عبور آسیب‌پذیر است و به احراز هویت پزشکان به سرورها یا نحوه اجرای سیاست‌های کنترل دسترسی اشاره ای نمی‌کند.

اثر شریا و همکاران [۱۷] یک پروتکل احراز هویت متقابل و توافق کلید بین کاربر (پزشک)، سرور ابری پزشکی و دروازه اینترنت اشیا را ارائه می‌دهد. این دروازه داده‌ها را از حسگرهای پزشکی جمع آوری می‌کند تا توسط کادر درمان مورد پردازش قرار گیرد. این پروتکل از رمزگذاری/رمزگشایی متقارن، هش و عملیات XOR استفاده می‌کند. اما این رویکرد پیشنهادی در برابر حدس زدن رمز عبور آفلاین آسیب‌پذیر است.

پاتل و همکاران [۱۸] یک رویکرد ترکیبی رمزنگاری مبتنی بر رمزنگاری منحنی بیضوی را به منظور افزایش امنیت داده‌های کاربران در فضای ابری موبایل ارائه کردند. روش پیشنهادی از یک فرایند تصادفی‌سازی برای رمزگذاری و رمزگشایی داده‌ها استفاده می‌کند که این عمل منجر به تولید کلید بهینه می‌شود، در نتیجه با اعمال این فرایند مهاجم قادر نخواهد بود داده‌ها را با موفقیت رمزگشایی کند حتی اگر کلید رمزگذاری را در اختیار داشته باشد. نتایج شبیه‌سازی نشان می‌دهد که روش پیشنهادی زمان فرایندهای رمزنگاری و رمزگشایی را در مقایسه با الگوریتم اصلی بلوفیش کاهش می‌دهد.

گانگیدی و همکاران [۱۹] یک رویکرد جدید را برای افزایش امنیت سایبری در فضای ابری مبتنی بر الگوریتم بلوفیش بهبود یافته به منظور انتخاب کلید بهینه معرفی کرده‌اند. در این رویکرد اطلاعات محرمانه با استفاده از الگوریتم خوشه‌بندی k-medoid خوشه‌بندی می‌شود و سپس داده‌ها با استفاده از رمزگذاری بلوفیش در فضای ابری رمزگذاری و ذخیره می‌شوند. نتایج شبیه‌سازی نشان می‌دهد که روش پیشنهادی دقت امنیت سایبری را برای تمام اطلاعات مخفی بهبود می‌بخشد و در مقایسه با الگوریتم‌های استاندارد هم چون بلوفیش، RSA و AES، زمان اجرای کمتری را در هنگام فرایندهای رمزگذاری و رمزگشایی به دست می‌آورد.

دوی پریا و همکاران [۸] یک طرح احراز هویت چندعاملی مبتنی بر زنجیره هش را برای اینترنت اشیا ارائه کرده‌اند، آنها ثابت کرده‌اند که پروتکل احراز هویت آنها در مقابل حملات مختلف از امنیت کافی برخوردار است، ولی باید به این نکته توجه کرد که طرح آنها به‌تازگی گواهینامه‌ها بستگی دارد و مشکلاتی که طرح‌های مبتنی بر گواهینامه‌ها دارند را طرح آنها نیز داراست.

چن و همکارانش [۹] یک طرح احراز هویت دوعاملی سبک را ارائه کرده‌اند. اما این رویکرد پیشنهادی در برابر حدس زدن رمز عبور آفلاین آسیب‌پذیر است.

ژیا و همکاران [۱۰] پروتکل احراز هویت گروهی چاین [۹] را مورد بررسی قرار دادند و نشان دادند که مهاجم می‌تواند در یک مدل ارتباطی ناهمزمان، بدون شناسایی شدن، به عنوان یک گره مشروع خود را جا بزند.

معمد و همکاران [۱۱] یک پروتکل احراز هویت گروهی مقیاس‌پذیر بر مبنای طرح‌های ترکیباتی با قابلیت تحمل‌پذیری خطا برای شبکه‌های IoT ارائه دادند. نظریه طراحی بیضوی، تحمل خطا بر مبنای تعداد اعضای گروه یک آستانه را مشخص می‌کند.

یائو و همکاران [۱۲] یک رویکرد احراز هویت و ایجاد کلید نشست سبک به نام LBAKA برای شبکه‌های UND کاربرمحور ارائه دادند که رویکرد احراز هویت دسته‌ای سبک وزن و توافق کلید یک به یک را با یکدیگر مطابقت می‌دهد. ارتباطات D2D به‌عنوان تکنولوژی ارتباط مستقیم کاربردهای گسترده‌ای دارد و نقش مهمی را در 5G بازی می‌کند.

سان و همکاران [۱۳] یک مکانیزم کشف دستگاه نزدیک با قابلیت گمنامی و مکانیزم احراز هویت دستی برای D2Dهای ناهمگن بر مبنای یک امضای دسته‌ای بدون گواهینامه بی‌نیاز از جفت شدن و کارا ارائه دادند که احراز هویت متقابل و تائید دسته‌ای را بدون نیاز به گواهینامه فراهم می‌آورد.

پارک و همکاران [۱۴] یک پروتکل طرح احراز هویت مبتنی بر زنجیره هش را برای شناسایی حملات در برابر گره‌ها حسگر ارائه نمودند. در این پروتکل پیشنهادی یک نفوذگر ممکن است یک گره حسگر مورد تعرض قرار دهد و به محتویات درون آن دسترسی پیدا نماید. یکی ضعف‌های این پروتکل پیشنهادی در دستیابی به محرمانه بودن مقادیر داخل حسگرها می‌باشد به این دلیل که کلیدهای جلسه Ks را به عنوان یک مقدار هش محاسبه می‌کند و اگر کلید سرور kser شناخته شده باشد، می‌توان همه ورودی‌های تابع هش را به راحتی محاسبه کرد.

۳- روش پیشنهادی

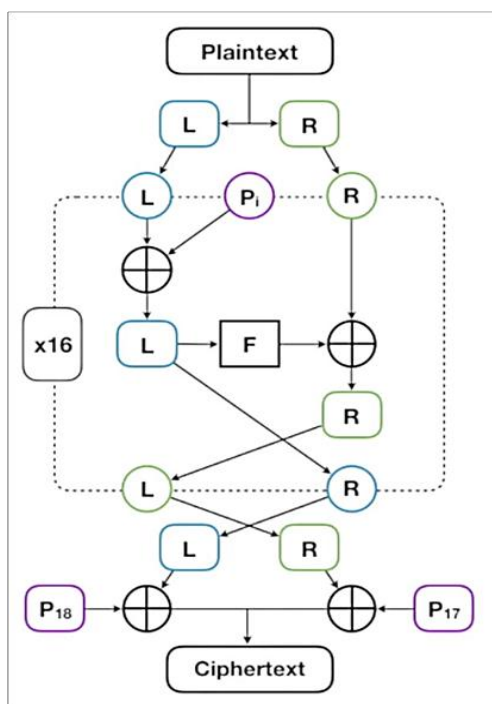
رشد پیوسته اینترنت اشیا و استفاده گسترده آن در هر مکان، باعث می شود که امنیت به یکی از مسائل اساسی در این زیرساخت تبدیل شود. از سوی دیگر، منابع محدود گره های اینترنت اشیا باعث می شود که استفاده از رمزنگاری های مختلف در این بستر به چالشی تبدیل شود. به همین دلیل، تمرکز اصلی بر استفاده از روش های رمزنگاری سبک وزن به منظور ارائه امنیت با حداقل مصرف منابع است. در این پژوهش، قصد داریم از یک راهکار ترکیبی برای تضمین امنیت در اینترنت اشیا استفاده کنیم. در این روش، داده اصلی با استفاده از الگوریتم متقارن بهبود یافته بلوفیش رمزنگاری می شود و سپس کلید رمزگذاری با استفاده از خم های بیضوی ایمن سازی می شود. به این ترتیب، با کمترین زمان و با سطح بالای امنیت، فرایند تبادل داده در زیرساخت های مبتنی بر اینترنت اشیا انجام می شود. این بهبود در رمزنگاری می تواند مزایای زیادی در محیط اینترنت اشیا داشته باشد. در ادامه این بخش، الگوریتم بلوفیش مورد بررسی قرار می گیرد و سپس نحوه استفاده از راهکار ترکیبی شرح داده می شود.

۳-۱- الگوریتم بلوفیش

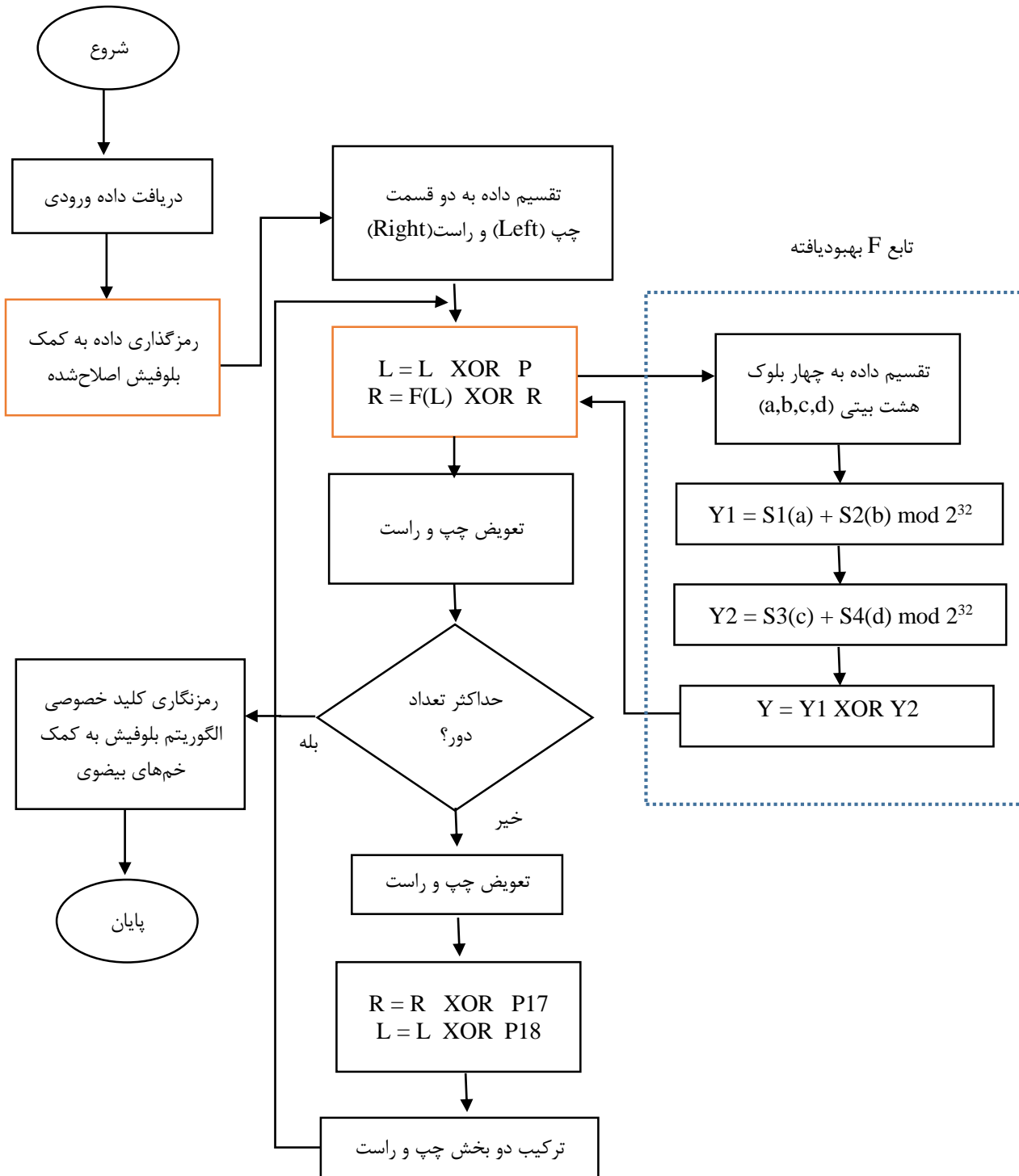
الگوریتم بلوفیش یک راهکار رمزنگاری متقارن است که از یک کلید برای عملیات رمزنگاری و رمزگشایی داده ها بهره می برد. این الگوریتم یک رمزگذار بلوکی است که پیام را در طول فرآیند رمزنگاری و رمزگشایی به بلوک هایی با طول ثابت تقسیم می کند. طول بلوک در بلوفیش ۶۴ بیت است و برای پیام هایی که طولشان مضرری از هشت بایت نیست، باید با فاصله گذاری تصحیح شوند. این الگوریتم شامل عملیات جمع، جستجوی جدول و XOR است. همچنین دارای یک جدول شامل چهار S-box و یک آرایه است. مهم ترین نکته این است که بلوفیش یک روش رمزنگاری مبتنی بر فایستل است و از تابع F برای ساده سازی اصول استفاده شده در DES استفاده می کند تا با سرعت و کارایی بالاتری امنیت را فراهم آورد. این الگوریتم بسیار سریع است و می تواند با حافظه کمتر از ۵ کیلوبایت اجرا شود. بلوفیش دارای دو بخش است: بسط کلید و رمزنگاری داده ها. در مرحله بسط، کلید مربوطه به چندین آرایه از کلید و یک مجموعه ۴۱۶۸ بیتی تبدیل می شود. همچنین آرایه P وجود دارد که شامل هجده P-box ۳۲ بیتی است و همچنین چهار آرایه S-box ۳۲ بیتی با ۲۵۶ درایه وجود دارد [۲۰]. در مرحله مقاداردهی، ۳۲ بیت اول کلید با استفاده از P1 (اولین بسته ۳۲ بیتی در آرایه p) XOR می شوند. سپس ۳۲ بیت دوم کلید با استفاده از

P2 دوباره XOR می شود و این فرآیند تا زمانی که تمام ۴۴۸ بیت کلید XOR شوند، ادامه می یابد. این چرخه روی بیت های کلید از طریق بازگشت به ابتدای کلید تکرار می شود تا در نتیجه کل آرایه P با کلید XOR شود. سپس تمام رشته ها با استفاده از الگوریتم و تابع F رمزنگاری می شوند تا به یک بلوک ۶۴ بیتی برسیم. سپس P1 با ۳۲ بیت اول خروجی و P2 با ۳۲ بیت دوم خروجی (از بلوک ۶۴ بیتی) جایگزین می شوند. خروجی ۶۴ بیتی به عنوان ورودی برای بلوفیش استفاده می شود تا یک بلوک ۶۴ بیتی جدید تولید شود. این فرآیند برای تمامی مقادیر در آرایه P و تمامی S-box ها تکرار می شود. در شکل ۱ نمای کلی رمزنگاری مبتنی بر الگوریتم بلوفیش نشان داده شده است.

با توجه به آنکه تابع فایستل F یکی از عوامل اساسی برای پیچیدگی زمانی الگوریتم است، در این مقاله، هدف اصلی اصلاح تابع F است تا زمان اجرا را کاهش داده و کارایی الگوریتم را افزایش دهیم. سپس با استفاده از الگوریتم بلوفیش، داده اصلی بهینه شده و در نهایت با استفاده از الگوریتم خم های بیضوی، کلید رمزنگاری می شود بر این اساس، در ادامه این بخش، ابتدا نحوه اصلاح تابع F به منظور بهبود کارایی الگوریتم بررسی شده و سپس در قدم بعدی نحوه ایمن سازی کلید آن با استفاده از رمزنگاری نامتقارن تشریح می شود. در شکل ۲ فلوجارت و راهکار پیشنهادی براساس الگوریتم بلوفیش بهبود یافته پیشنهادی نشان داده شده است.



شکل ۱. فرایند رمزنگاری مبتنی بر الگوریتم بلوفیش [۱۷].



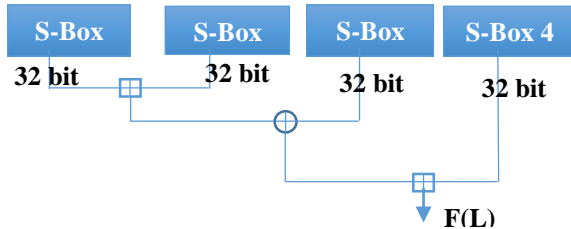
شکل ۲. فلوجارت الگوریتم پیشنهادی

اصلاح این تابع به منظور کاهش زمان اجرا گرفته شده است. در حالت کلی، تابع F به شکلی که در شکل ۳ نمایش داده شده است عمل می‌کند.

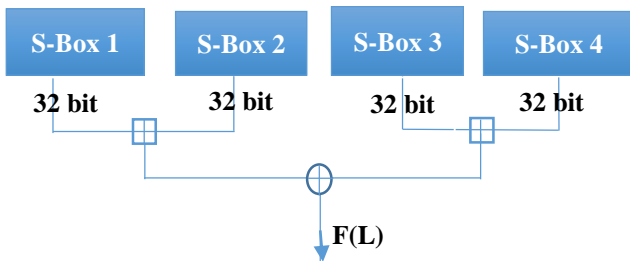
۳-۱-۱- تابع F بهبودیافته داده شده در روش پیشنهادی

تابع F در الگوریتم رمزنگاری بلوفیش نقش مهمی ایفا می‌کند و بر روی زمان اجرا تأثیر قابل توجهی دارد. در این مقاله، تصمیم به

همانند در منحنی در نظر گرفته می‌شود. ترتیب منحنی برابر با تعداد نقاط مشخص در منحنی است و شامل نقطه صفر نیز می‌شود [۲۱].



شکل ۳. نحوه عملکرد تابع F



شکل ۴. نحوه عملکرد تابع F بهبود یافته

Algorithm 1: The Bluefish algorithm has been improved along with the F function

Input: data d
Output: encrypted d
 Split d into two 32-bit: dLeft, dRight
For j = 1 to 16:
 dL = dL XOR Pj
 dR = F(xL) XOR dR
 Swap dL & dR
Next j
Swap dL & dR
 dR = dR XOR P17
 dL = dL XOR P18
 Merge xL & xR
 Func F
 Split dL in four 8-bit: a, b, c, d
 $F(dL) = (S_{1,a} + S_{2,b} \bmod 2^{32}) \text{ XOR } (S_{3,c} + S_{4,d} \bmod 2^{32})$

شکل ۵. شبه کد تابع F بهبود یافته داده شده روش پیشنهادی

$$F(DL) = ((S_{1,a} + S_{2,b} \bmod 2^{32}) \text{ XOR } S_{3,c}) + S_{4,d} \bmod 2^{32} \quad (1)$$

براساس این تغییرات، می‌توان تابع F را به منظور موازی‌سازی و اجرای سریع‌تر الگوریتم به شکل رابطه (۲) تغییر داد. تابع F بهبود داده شده به شکلی که در شکل ۴ نمایش داده شده است تغییر پیدا می‌کند.

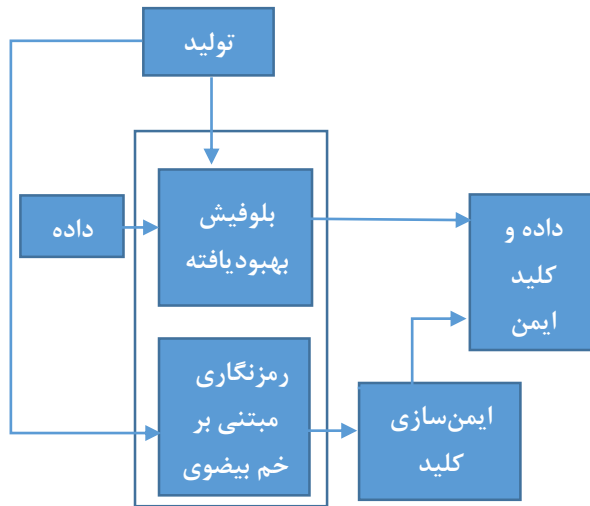
$$F'(XL) = (S_{1,a} + S_{2,b} \bmod 2^{32})(S_{3,c} + S_{4,d} \bmod 2^{32}) \quad (2)$$

تغییرات اعمال شده در تابع F منجر به اجرای همزمان دو عملیات XOR می‌شود. در تابع F اصلی، این عملیات به صورت سریالی اجرا می‌شود و نیازمند ۳۲ عمل جمع و ۱۶ عمل XOR بود. اما با تغییرات اعمال شده در تابع F اصلاح شده، نیاز به انجام ۴۸ عمل منطقی (۳۲ عمل XOR و ۱۶ عمل جمع) است. با این حال، زمان صرف شده برای اجرای این ۴۸ عمل به دلیل مکانیزم چندبخشی کاهش می‌یابد. در ادامه شبه کد تابع F بهبود یافته داده شده روش پیشنهادی در شکل ۵ نشان داده شده است.

به منظور حل مشکلات مربوط به انتقال کلید خصوصی در الگوریتم بلوفیش، در این تحقیق از رمزنگاری نامتقارن مبتنی بر خم‌های بیضوی (EC) استفاده می‌شود. رمزنگاری خم‌های بیضوی یک روش برای رمزنگاری مبتنی بر کلید عمومی است که بر اساس منحنی‌های بیضوی در میدان‌های متناهی عمل می‌کند [۲۱]. این الگوریتم قابلیت ارائه امنیت با استفاده از کلید ۱۶۴ بیتی را دارا بوده و به دلیل مصرف انرژی کمتر، عملکرد بهتری نسبت به روش‌های قبلی ارائه می‌دهد. در واقع، نسبت به سایر روش‌های رمزنگاری نامتقارن، این روش دارای بالاترین سطح محرمانگی، مصرف کمتر انرژی و نیاز کمتر به حافظه در سیستم است [۲۲]. یکی از ویژگی‌های مهم این رمزنگاری، انجام عملیات بر روی میدان‌های متناهی است. رمزنگاری مبتنی بر خم‌های بیضوی بر اساس مسئله لگاریتم منحنی گسسته بیضوی تعریف می‌شود که یک مسئله NP-Hard است. یک منحنی بیضوی توسط رابطه (۳) تعریف می‌شود:

$$y^2 + xy = x^3 + ax + b \quad (3)$$

یکی از ویژگی‌های اساسی در منحنی بیضوی این است که برای یافتن نقطه سوم روی منحنی، می‌توان یک قانون جهت اضافه کردن دو نقطه به منظور به دست آوردن آن نقطه سوم روی منحنی، مشخص کرد. این نقاط و قانون اضافه کردن، یک دسته آبدی متناهی را شکل می‌دهد. برای تعریف مناسب عمل اضافه کردن دو نقطه، لازم است یک نقطه صفر اضافی را مشخص کنیم که معادلات منحنی بیضوی را برآورده نمی‌کند. این نقطه صفر به عنوان نقطه



شکل ۶. روش ترکیبی پیشنهادی بمنظور ایمن‌سازی داده در کنار کلید خصوصی

این روش امنیتی بر پایه محاسبات ریاضی بر روی خم‌های بیضوی در میدان متناهی عمل می‌کند. در این راهکار، ابتدا یک منحنی بیضوی غیرمحرمانه و یک نقطه ثابت نیز غیرمحرمانه تعیین می‌شوند. سپس کاربران (مانند آلیس، باب، کتی و دیوید) اعداد صحیح محرمانه خود را بر روی منحنی بیضوی انتخاب می‌کنند و از نقطه‌ای بر روی منحنی به عنوان کلید عمومی خود استفاده می‌کنند. برای ایمن‌سازی داده، کاربران با استفاده از کلیدهای عمومی یکدیگر، ابتدا نقاطی را روی منحنی بیضوی محاسبه می‌کنند و سپس از این نقاط برای رمزنگاری و امضای داده‌ها استفاده می‌کنند. برای محاسبه این نقاط، اعمالی مشخص روی منحنی بیضوی (مانند جمع نقاط) انجام می‌شود. امنیت این روش بر پایه پیچیدگی محاسباتی محاسبه کلیدها و عملیات روی خم‌های بیضوی استوار است. با توجه به خواص ریاضی خم‌های بیضوی، محاسبه کلیدها و عملیات روی آنها با توان مصرفی کمتر و در زمان کمتری نسبت به روش‌های دیگر صورت می‌گیرد. این باعث می‌شود که کلیدهای عمومی و امضای مبتنی بر ECC به نسبت الگوریتم‌های نامتقارن دیگر کوچکتر و موثرتر باشند، در حالی که همان سطح امنیت را فراهم می‌کنند. به این ترتیب، راهکار ترکیبی مبتنی بر ECC امنیت بالا و کارایی برتر را با هم ترکیب می‌کند، که آن را به یک گزینه محبوب برای رمزنگاری و امضای داده‌ها در سیستم‌های امنیتی مختلف می‌کند.

۴- شبیه‌سازی روش پیشنهادی

راهکار پیشنهادی و روش‌های مورد مقایسه با استفاده از محیط Eclipse و با استفاده از JDK نسخه ۷ کدنویسی شده‌اند و در سیستم عامل ویندوز ۶۴ بیتی و با حافظه 16G اجرا گردیده است.

۳-۱-۲- انتخاب منحنی ثابت

برای انتخاب منحنی ثابت، فرض می‌گردد عدد p یک عدد اول بزرگ است و میدان متناهی $GF(p)$ شامل مجموعه‌ای از اعداد کمتر از p می‌باشند. در نتیجه، از طریق حذف عبارت xy از رابطه (۳) و با توجه به رابطه $4a^3 + 27b^2 \neq 0$ می‌توان خم بیضوی را به صورت رابطه (۴) تعریف کرد:

$$y_2 = x_3 + ax^2 + b \quad (4)$$

برای رمزنگاری مبتنی بر الگوریتم EC، ابتدا یک عدد تصادفی k از مجموعه محدود $\{1, \dots, n-1\}$ که در میدان قرار دارد، انتخاب می‌شود.

این عدد تصادفی در واقع به عنوان کلید خصوصی در نظر گرفته می‌شود. سپس کلید عمومی R با استفاده از رابطه $R = Fk$ محاسبه می‌شود، که در آن F یک نقطه بر روی خم بیضوی می‌باشد. در این صورت، محاسبه k با استفاده از نقاط Q و F دارای پیچیدگی زمانی نمایی است. برای افزایش امنیت، نقطه‌ی ثابت و منحنی به گونه‌ای انتخاب می‌شوند که مرتبه نقطه‌ی F یک عدد اول بسیار بزرگ باشند. مرتبه منحنی به وسیله‌ی الگوریتم اسچوف مشخص می‌گردد. اگر مرتبه نقطه‌ی ثابت F یک عدد اول n -بیتی باشد، محاسبه‌ی k با استفاده از kF و F به‌طور تقریبی به $2^{\frac{n}{2}}$ عملیات نیاز دارد. این امر نشان می‌دهد که استفاده از منحنی‌های بیضوی توجیه‌پذیر است، به این معنی که کلیدهای عمومی نسبت به الگوریتم نامتقارن از جمله RSA بسیار کوچکتر خواهد بود، در حالی که همچنان امنیت لازم را فراهم می‌کنند.

۳-۱-۳- نحوه رمزنگاری EC

با توجه به فرضیات مطرح شده، در این سناریو باب، آلیس، دیوید و کتی بر روی یک منحنی بیضوی غیرمحرمانه و یک نقطه منحنی ثابت غیرمحرمانه f توافق کرده‌اند. آلیس عدد صحیح محرمانه مانند Key_k را تعیین می‌کند که به عنوان کلید رمز او عمل می‌کند و نقطه منحنی $Key_p = Key_k F$ را به عنوان کلید عمومی خود منتشر می‌کند. باب، کتی و دیوید همچنین همین عمل را انجام می‌دهند. حال آلیس می‌خواهد پیغامی را به باب ارسال کند. برای این منظور، آلیس ابتدا key_{kBp} را محاسبه می‌کند و از حاصل آن به عنوان کلید برای رمزنگاری بهره می‌برد. همچنین، باب هم می‌تواند این عدد را از طریق محاسبه $BkKey_p$ تعیین کند. زیرا داریم:

$$B_k A_p = B_k \cdot (A_k F) = A_k \cdot (B_k F) = A_k B_{kp} \quad (5)$$

امنیت راهکار ترکیبی برای ایمن‌سازی داده و کلید در شکل ۶ معمولاً بر اساس رمزنگاری مبتنی بر خم‌های بیضوی استوار است.

۴-۲-۱- بررسی زمان اجرا

این معیار میزان زمان مورد نیاز برای انجام عملیات رمزنگاری و رمزگشایی را نشان می‌دهد. کاهش زمان اجرا نشانگر کارایی بالاتر راهکار است. زمان اجرا از طریق رابطه (۶) محاسبه می‌شود:

$$\text{Execution Time(ms)} = \text{EncryotStartTime} - \text{End Time} \quad (۶)$$

در ادامه این بخش و به ترتیب در شکل‌های ۷ و ۸ مقایسه زمان اجرای عملیات رمزنگاری و رمزگشایی روش پیشنهادی با سایر روش‌های موجود در [۷] آمده است. همان‌طور که در شکل ۹ مشاهده می‌شود، میانگین زمان اجرای رمزنگاری و رمزگشایی روش پیشنهادی به طرز چشمگیری کمتر از سایر روش‌های مورد ارزیابی است. در روش پیشنهادی، به ترتیب زمان اجرای رمزنگاری ۱۸،۴۷ میلی‌ثانیه و زمان اجرای رمزگشایی برابر با ۲۱،۴۵ میلی‌ثانیه می‌باشد. دلیل این کارایی بالا روش پیشنهادی در مقایسه با الگوریتم‌های مورد ارزیابی استفاده از رویکرد رمزنگاری پیشنهادی است که در رویکرد پیشنهادی عملیات رمزنگاری داده‌ها با استفاده از الگوریتم متقارن بلوفیش صورت می‌گیرد و همچنین با استفاده از روش رمزنگاری نامتقارن مبتنی بر خم‌های بیضوی ایمن‌سازی کلید انجام می‌شود، بر همین اساس این موضوع باعث کمترین تأثیر روی زمان اجرا می‌شود.

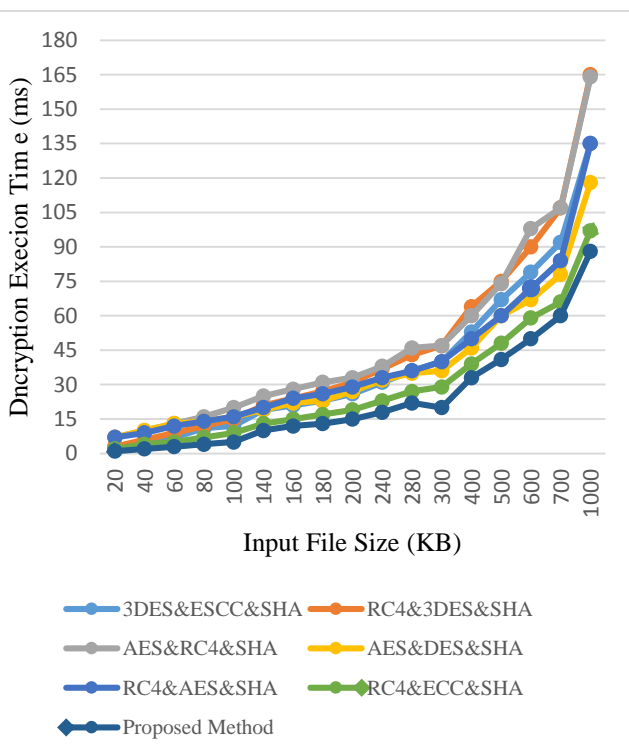
۴-۱- ارزیابی روش پیشنهادی

در JDK، توابع رمزنگاری با استفاده از دو کتابخانه اصلی JCA و JCE قابل استفاده هستند. کتابخانه JCA به طور کامل با هسته API‌های جاوا ترکیب شده است و قادر است بیشتر قابلیت‌های پایه‌ای رمزنگاری را فراهم کند و همچنین، برای عملیات‌های پیشرفته رمزنگاری از کتابخانه JCE استفاده شده است.

برای ارزیابی راهکار ارائه شده، از معیارهای مختلفی مانند زمان اجرا، توان عملیاتی و میانگین میزان محرمانگی استفاده شده است. به دلیل تأثیر حجم داده‌ها بر زمان اجرا و توان عملیاتی الگوریتم‌های رمزنگاری، در ادامه ارزیابی‌ها با استفاده از حجم داده‌ها متنی (۲۰ تا ۱۰۰۰ کیلوبایت) اجرا و کارایی آنها بررسی شده است. بر همین اساس در ابتدا، ارزیابی‌ها با حجم کوچک داده‌ها متنی انجام می‌شوند و سپس به حجم داده‌های متنی افزوده می‌شود تا بتوان به میزان کارایی راهکارها در هر دو حالت پرداخت. در این حالت، تأثیر افزایش حجم داده‌ها بر عملکرد راهکارها سنجیده می‌شود زیرا فرایند رمزنگاری تا حد زیادی تحت تأثیر اندازه داده‌ها قرار می‌گیرد و الگوریتمی که در زمان کمتری عملیات رمزنگاری را انجام دهد، مناسب‌تر است.

۴-۲- معیارهای ارزیابی

همان‌طور که گفته شد معیارها و پارامترهای ارزیابی که در روش پیشنهادی در نظر گرفته شده است زمان اجرا، توان عملیاتی و میانگین میزان محرمانگی می‌باشند. کارهایی که جهت مقایسه با روش پیشنهادی در نظر گرفته شده است الگوریتم‌های رمزنگاری در مقاله [۷] می‌باشند. علت انتخاب این روش‌ها سازگاری بیشتر روش‌های مورد مقایسه با محیط شبیه‌سازی و استفاده از الگوریتم‌های رمزنگاری متقارن و غیر متقارن می‌باشد. ایده روش پیشنهاد شده در این مقاله، ارائه یک راهکار رمزنگاری ترکیبی و سبک‌وزن، مبتنی بر الگوریتم بلوفیش و خم‌های بیضوی برای حل مساله بهبود امنیت در اینترنت اشیا می‌باشد که در رویکرد ارائه شده سعی می‌شود براساس راهکار رمزنگاری ترکیبی و سبک‌وزن بتوان در بهینه نمودن زمان اجرا، توان عملیاتی و میانگین میزان محرمانگی تأثیرگذار باشیم، برای بررسی روش پیشنهادی و روش‌های مورد مقایسه از شبیه‌ساز Eclipse استفاده شده است. در ادامه نمودارهای حاصل از شبیه‌سازی بر روی حجم داده‌ها (۲۰ تا ۱۰۰۰ کیلوبایت) در شکل‌های ۷ تا ۱۳ برای معیارهای رویکرد پیشنهادی مورد نظر آورده شده است.



شکل ۷. مقایسه زمان اجرای رمزنگاری روش پیشنهادی با سایر روش‌های پیشین

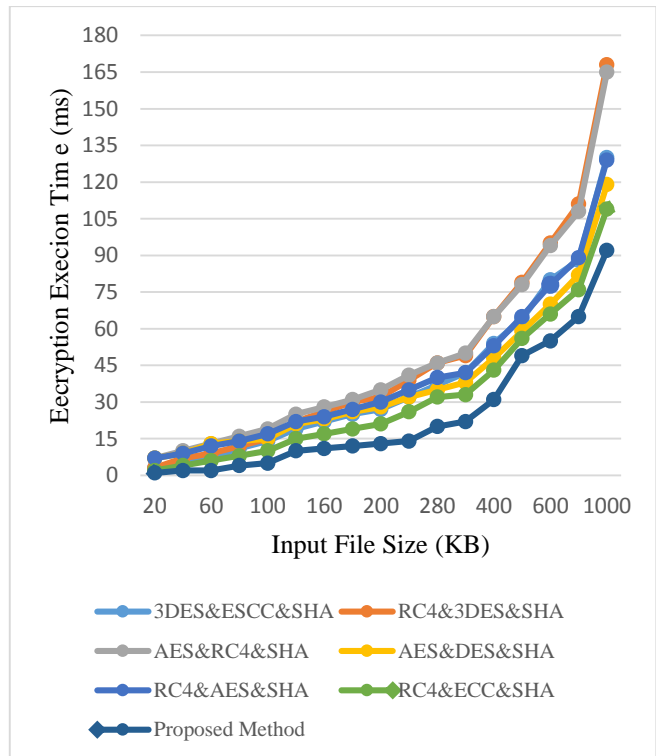
۴-۲-۲- بررسی توان عملیاتی

این معیار میزان تعداد عملیات رمزنگاری و رمزگشایی انجام شده در واحد زمان را نشان می‌دهد. برای ارزیابی کارایی راهکار، باید تعداد عملیات موفقیت‌آمیز واحد زمان محاسبه شود. همچنین به کمک رابطه (۷) زیر می‌توان به ترتیب میزان توان عملیاتی عملیات رمزنگاری و رمزگشایی در طی فرایند ارزیابی را بدست آورد:

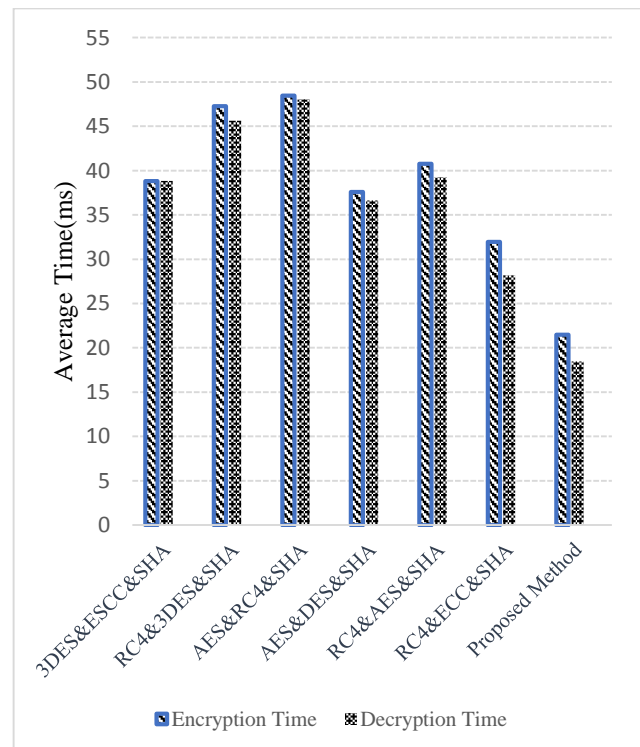
$$\text{Throughput} \left(\frac{\text{KB}}{\text{ms}} \right) = \frac{\text{Data Size}}{\text{Execution Time}} \quad (7)$$

در رابطه (۷)، توان عملیاتی به وسیله تقسیم سایز داده‌هایی که در ارزیابی استفاده می‌شوند بر زمان اجرای مورد نیاز محاسبه می‌شود. به عبارتی می‌توان گفت که مقدار توان عملیاتی با تقسیم سایز داده‌های مورد ارزیابی بر زمان اجرای لازم به دست می‌آید. در شکل‌های ۱۰ و ۱۱ مقایسه توان عملیاتی عملیات رمزنگاری و رمزگشایی روش پیشنهادی با سایر روش‌های موجود در مقاله [۷] آمده است.

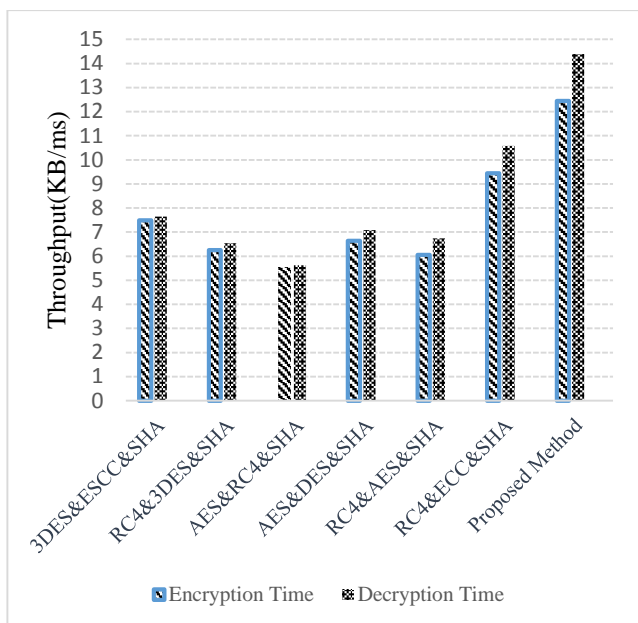
همانطور که در شکل ۱۲ ملاحظه می‌شود، توان عملیاتی راهکار پیشنهادی در حالت رمزنگاری برابر ۱۲,۴۵ کیلوبایت/ثانیه و در حالت رمزگشایی برابر با ۱۴,۴۰ کیلوبایت/ثانیه می‌باشد که بر همین اساس توان عملیاتی روش پیشنهادی در مقایسه با سایر راهکارهای مورد ارزیابی بالاتر است. به عبارت دیگر، در راهکار پیشنهادی، با زمان کمتری، حجم بیشتری از داده‌ها رمزنگاری می‌شود که نشان از بهبود عملکرد و کارایی راهکار می‌دهد. این افزایش کارایی به دلیل استفاده از راهکار رمزنگاری ترکیبی و اصلاح هسته رمزنگاری الگوریتم بلوفیش است. در این حالت، تابع F که عامل اصلی پیچیدگی الگوریتم است، تغییر داده شده است و این تغییر باعث اجرای همزمان دو عملیات XOR در فرایند رمزنگاری می‌شود. این موضوع سبب کاهش پیچیدگی و افزایش زمان اجرا و توان عملیاتی الگوریتم می‌شود. در واقع، در این حالت تغییراتی که در تابع F اعمال شده، باعث گردیده که پیچیدگی الگوریتم کاهش پیدا کرده و در نتیجه زمان اجرا و توان عملیاتی بهبود یابد. در کل، راهکار پیشنهادی با ترکیب این دو رویکرد موفق به ارائه راهکاری با کارایی بالا در زمینه رمزنگاری شده است.



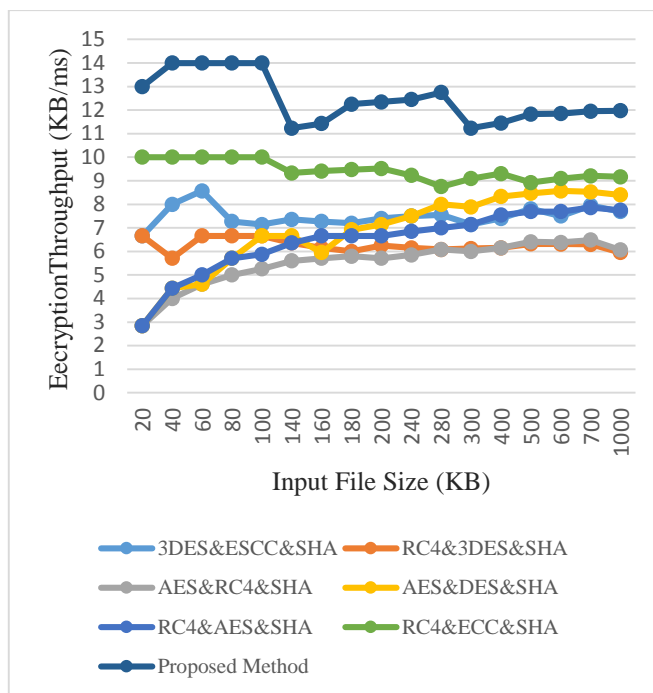
شکل ۸. مقایسه زمان اجرای رمزگشایی روش پیشنهادی با سایر روش‌های پیشین



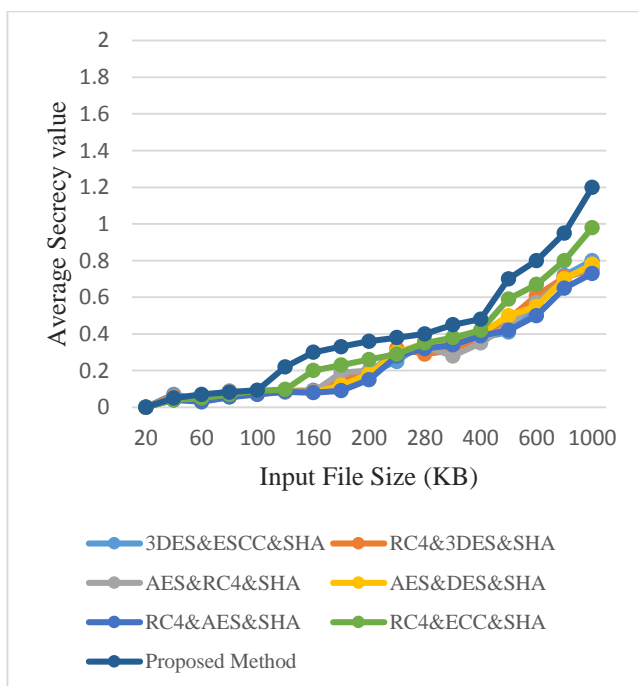
شکل ۹. مقایسه میانگین زمان اجرای رمزنگاری و رمزگشایی روش پیشنهادی با سایر روش‌های پیشین



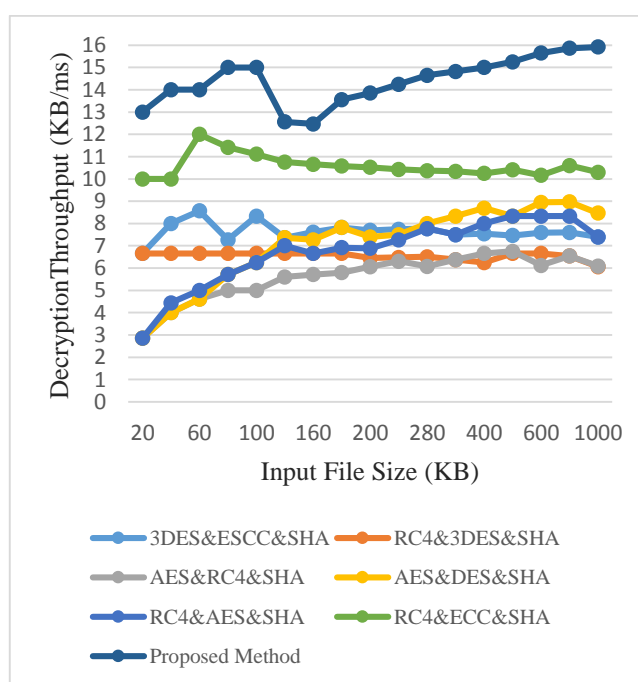
شکل ۱۲. مقایسه میانگین توان عملیاتی رمزنگاری و رمزگشایی روش پیشنهادی با سایر روش‌های پیشین



شکل ۱۰. مقایسه توان عملیاتی رمزنگاری روش پیشنهادی با سایر روش‌های پیشین



شکل ۱۳. مقایسه میانگین میزان محرمانگی روش پیشنهادی با سایر روش‌های پیشین



شکل ۱۱. مقایسه توان عملیاتی رمزگشایی روش پیشنهادی با سایر روش‌های پیشین

در شکل ۱۳ میانگین میزان محرمانگی بدست آمده توسط مدل پیشنهادی، DES&ECC&SHA، RC4&3DES&SHA، RC4&AES&SHA، AES&DES&SHA، ECS&RC4&SHA، DES&ECC&SHA به ترتیب ۰،۹۸۲۸، ۰،۸۴۴۳، ۰،۹۹۸۸، ۰،۸۴۵۵، ۱،۷۱۵، ۰،۹۹۸۸، ۰،۹۷، ۱،۰۹۷

۴-۲-۳- بررسی میزان محرمانگی

در این بخش به مقایسه تحلیل امنیتی مدل پیشنهادی با سایر مدل‌های مقاله [۷] پرداخته شده است. بر همین اساس اصل حریم خصوصی با استفاده از قانون شانون [۲۳] محاسبه می‌شود که هدف از این معیار بررسی محرمانه بودن داده‌ها است.

مدل پیشنهادی را با مدل‌های دیگر بر اساس چندین ویژگی امنیتی مانند احراز هویت، محرمانگی، یکپارچگی و مقاوم بودن در برابر حملات مختلف نشان می‌دهد [۲۴]. همانطور از مقایسه‌هایی که در جدول ۱ انجام گرفته است، واضح است که مدل پیشنهادی با تمام الزامات امنیتی موافق است و احراز هویتی را ارائه می‌کند که تاکنون اکثر مدل‌ها نتوانستند این الزامات امنیتی را کامل کنند. همچنین، به دلیل اینکه مدل پیشنهادی مبتنی بر ECC است، یکپارچگی و محرمانه بودن اصلی را فراهم می‌کند.

۲۰ تا ۱۰۰۰ کیلوبایت نشان داده شده است. همانطور که در شکل ۱۳ ملاحظه می‌شود، روش پیشنهادی از نظر میانگین میزان محرمانگی مزایای آشکاری نسبت به سایر الگوریتم‌های مذکور را دارا می‌باشد.

۴-۲-۴- بررسی ویژگی‌های امنیتی

در این بخش، تجزیه و تحلیل امنیتی براساس ۱۰ پارامتر به منظور مقایسه روش پیشنهادی با سایر الگوریتم‌های رمزنگاری مقاله [۷] انجام شده است. بر همین اساس جدول ۱ تجزیه و تحلیل امنیتی

جدول ۱. مقایسه مدل پیشنهادی براساس ویژگی‌های امنیتی با سایر روش‌ها

Model	3DES&ECC&SHA	RC4&3DES&S&SHA	AES&RC4&SHA	AES&3DES&S&SHA	RC4&AES&SHA	3DES&EC&C&SHA	Proposed Method
Authentication key	✓	x	x	x	x	x	✓
Key agreement	✓	x	x	x	x	x	✓
Integrity	✓	✓	✓	✓	x	✓	✓
Confidentiality	✓	x	x	x	x	x	✓
Secrecy	x	x	x	x	x	x	✓
Resistant to man-in-the-middle attack	✓	x	x	x	x	✓	✓
Resistant to malicious user attack	x	x	x	x	x	x	✓
Resistant to insider attack	x	x	x	x	x	✓	✓
Resistant to brute Force	x	x	x	x	x	x	✓
Attack Key exchange	✓	x	x	x	x	✓	✓

خم‌های بیضوی یک روش رمزنگاری مبتنی بر کلید عمومی است که بر پایه منحنی‌های بیضوی در میدان‌های متناهی عمل می‌کند. این الگوریتم از کلید ۱۶۴ بیتی استفاده می‌کند و در مقایسه با روش‌های قبلی، به دلیل مصرف انرژی کمتر، دارای کارایی بالاتری است. همچنین، این روش نیاز به حافظه کمتری دارد و از نظر محرمانگی و امنیت، محدودیت‌های کمتری نسبت به روش‌های دیگر دارد. به طور کلی، استفاده از رمزنگاری ترکیبی مبتنی بر الگوریتم بلوفیش و خم‌های بیضوی می‌تواند امنیت بالا، کارایی مناسب و مصرف انرژی کمتری را در فرایند رمزنگاری و رمزگشایی داده‌ها فراهم کند. در پایان، پیاده‌سازی راهکار پیشنهادی با استفاده از زبان برنامه‌نویسی جاوا و کیت توسعه جاوا ۷ در محیط Eclipse IDE انجام شده است. این راهکار با استفاده از الگوریتم‌های رمزنگاری موجود در مقاله [۷] برای حجم مختلف داده‌ها مقایسه شده است. نتایج ارزیابی نشان می‌دهد که این راهکار، با توجه به معیارهای زمان اجرا و توان عملیاتی بهینه‌تر عمل کرده است. در مقایسه با سایر الگوریتم‌ها، راهکار پیشنهادی در تمامی ارزیابی‌های انجام شده، کمترین زمان اجرا و بیشترین توان عملیاتی را داشته است. این نتایج نشان می‌دهد که عملکرد این راهکار بهبود یافته است و در مقایسه

۵- نتیجه‌گیری

حفظ امنیت در همه مراحل اینترنت اشیا از اهمیت بالایی برخوردار است و نیازمند رویکردهای جامع و سیستماتیک برای حفاظت از داده‌ها و سیستم‌ها در این زمینه می‌باشد. تلاش‌ها در جهت بهبود استانداردها و فناوری‌های امنیتی در حال ادامه است تا راهکارهایی مؤثر و کارآمد در این زمینه ارائه شوند. راهکار رمزنگاری ترکیبی مبتنی بر الگوریتم بلوفیش و خم‌های بیضوی که در این پژوهش ارائه شده است، بهبودهایی را در کارایی و امنیت فراهم می‌کند. از آنجا که الگوریتم بلوفیش در این روش استفاده می‌شود، تغییراتی در الگوریتم اعمال شده است تا پیچیدگی کلی رمزنگاری کاهش یابد و بلوفیش تغییر یافته، تابع F که به طور عادی به صورت سریالی اجرا می‌شود، به صورت همزمان اجرا می‌شود و نیازمند ۴۸ عمل منطقی است که شامل ۳۲ عمل XOR و ۱۶ عمل جمع است. با این وجود، زمان اجرای این ۴۸ عمل به دلیل مکانیزم چندبخشی کاهش می‌یابد. بعد از انجام رمزنگاری با الگوریتم بلوفیش تغییر یافته، کلید خصوصی ایمن‌سازی می‌شود. این عملیات ایمن‌سازی از طریق رمزنگاری نامتقارن خم‌های بیضوی انجام می‌شود. رمزنگاری

- با روش‌های دیگر، کارایی بالاتری دارد. برای کارهای آتی استفاده از رویکرد پیشنهادی در رایانش لبه می‌تواند به دلیل مصرف پایین حافظه، منجر به بهبود عملکرد و کارایی سامانه‌های موجود در لبه شبکه شود. به عبارتی، با کاهش مصرف حافظه، منابع محدود رایانش لبه بیشتر مورد بهره‌برداری قرار خواهند گرفت و تعداد بیشتری از وظایف رمزنگاری و رمزگشایی قابل انجام خواهند بود. همچنین، با مصرف کمتر حافظه و کاهش زمان اجرا، امکان اجرای همزمان بیشتری از وظایف رمزنگاری و رمزگشایی در دستگاه‌های رایانش لبه وجود دارد که بهبود قابل توجهی در پاسخگویی سیستم به دست می‌آید. بنابراین، از این راهکار پیشنهادی در رایانش لبه می‌توان بهره‌برداری کرده و با بهینه‌سازی مصرف حافظه و افزایش توان عملیاتی، کارایی و کاربردی تر شدن سامانه‌های رایانش لبه را تسهیل کرد.
- ### مراجع
- [1] Xia Z, Liu Y, Hsu CF, Chang CC. Cryptanalysis and improvement of a group authentication scheme with multiple trials and multiple authentications. *Security and Communication Networks*. 2020 Jul 13;2020:1-8.
 - [2] El Mouaatamid O, Lahmer M, Belkasm M. A scalable group authentication scheme based on combinatorial designs with fault tolerance for the Internet of things. *SN Computer Science*. 2020 Jul;1:1-3.
 - [3] Yao Y, Chang X, Mišić J, Mišić VB. Lightweight batch AKA scheme for user-centric ultra-dense networks. *IEEE Transactions on Cognitive Communications and Networking*. 2020 Mar 20;6(2):597-606.
 - [4] Sun Y, Cao J, Ma M, Zhang Y, Li H, Niu B. EAP-DDBA: efficient anonymity proximity device discovery and batch authentication mechanism for massive D2D communication devices in 3GPP 5G HetNet. *IEEE transactions on dependable and secure computing*. 2020 Apr 23;19(1):370-87.
 - [5] Park K, Noh S, Lee H, Das AK, Kim M, Park Y, Wazid M. LAKS-NVT: Provably secure and lightweight authentication and key agreement scheme without verification table in medical internet of things. *IEEE Access*. 2020 Jun 29;8:119387-404.
 - [6] Zhu L, Xiang H, Zhang K. A Light and Anonymous Three-Factor Authentication Protocol for Wireless Sensor Networks. *Symmetry* 2022, 14, 46. Optimization and Applications of Modern Wireless Networks and Symmetry. 2021:3.
 - [7] Wu F, Li X, Sangaiah AK, Xu L, Kumari S, Wu L, Shen J. A lightweight and robust two-factor authentication scheme for personalized healthcare systems using wireless medical sensor networks. *Future Generation Computer Systems*. 2018 May 1;82:727-37.
 - [8] Shreya S, Chatterjee K, Singh A. A smart secure healthcare monitoring system with Internet of Medical Things. *Computers and Electrical Engineering*. 2022 Jul 1;101:107969.
 - [9] Patel P, Patel R, Patel N. Integrated ECC and Blowfish for smartphone security. *Procedia Computer Science*. 2016 Jan 1;78:210-6.
 - [10] Gangireddy VK, Kannan S, Subburathinam K. RETRACTED ARTICLE: Implementation of enhanced blowfish algorithm in cloud environment. *Journal of Ambient Intelligence and Humanized Computing*. 2021 Mar;12(3):3999-4005.
 - [11] Adhikary T, Jana AD, Chakrabarty A, Jana SK. The internet of things (iot) augmentation in healthcare: An application analytics. *ICICCT 2019—System Reliability, Quality Control, Safety, Maintenance and Management: Applications to Electrical, Electronics and Computer Science and Engineering*. 2020:576-83.
 - [12] Dhillon PK, Kalra S. Elliptic curve cryptography for real time embedded systems in IoT networks. In 2016 5th international conference on wireless networks and embedded systems (WECON) 2016 Oct 14 (pp. 1-6). IEEE.
 - [13] Durairaj M, Muthuramalingam K. A new authentication scheme with elliptical curve cryptography for internet of things (IoT) environments. *Int. J. Eng. Technol*. 2018;7(2.26):119-24.
 - [14] Weerasinghe TD. An effective RC4 stream cipher. In 2013 IEEE 8th international conference on industrial and information systems 2013 Dec 17 (pp. 69-74). IEEE.
 - [15] Lohachab A, Karambir B. Critical analysis of DDoS—An emerging security threat over IoT networks. *Journal of Communications and Information Networks*. 2018 Sep;3:57-78.
 - [16] Hassija V, Chamola V, Saxena V, Jain D, Goyal P, Sikdar B. A survey on IoT security: application areas, security threats, and solution architectures. *IEEE Access*. 2019 Jun 20;7:82721-43.
 - [17] Ammar M, Russello G, Crispo B. Internet of Things: A survey on the security of IoT frameworks. *Journal of Information Security and Applications*. 2018 Feb 1;38:8-27.
 - [18] Mrabet H, Belguith S, Alhomoud A, Jemai A. A survey of IoT security based on a layered architecture of sensing and data analysis. *Sensors*. 2020 Jun 28;20(13):3625.
 - [19] HaddadPajouh H, Dehghantanha A, Parizi RM, Aledhari M, Karimipour H. A survey on internet of things security: Requirements, challenges, and solutions. *Internet of Things*. 2021 Jun 1;14:100129.
 - [20] Mousavi SK, Ghaffari A, Besharat S, Afshari H. Security of internet of things based on cryptographic algorithms: a survey. *Wireless Networks*. 2021 Feb;27(2):1515-55.
 - [21] Fraga-Lamas P, Fernández-Caramés TM, Suárez-Albela M, Castedo L, González-López M. A review on internet of things for defense and public safety. *Sensors*. 2016 Oct 5;16(10):1644.
 - [22] Mousavi SK, Ghaffari A, Besharat S, Afshari H. Improving the security of internet of things using cryptographic algorithms: a case of smart irrigation systems. *Journal of Ambient Intelligence and Humanized Computing*. 2021 Feb;12(2):2033-51.
 - [23] Wang, W. Li, and P. Wang, "Measuring TwoFactor Authentication Schemes for Real-Time Data Access in Industrial Wireless Sensor Networks," *IEEE Trans. Ind. Informatics*, vol. 14, no. 9, pp. 4081–4092, 2018.
 - [24] C. M. Chen, S. Liu, X. Li, S. Kumari, and L. Li, "Design and Analysis of a Provable Secure TwoFactor Authentication Protocol for Internet of Things," *Secur. Commun. Networks*, vol. 2022.