فصلنامه علمی-پژوهشی
فناوری اطلاعات و ارتباطات ایران
سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷

انجمن فناوری اطلاعات و ارتباطات ایران

صاحب امتیاز: انجمن فناوری اطلاعات و ارتباطات ایران
مدیر مسؤول: دکتر مسعود شفیعی
سردبیر: دکتر محمدشهرام معین
شمارگان: ۱۰۰۰
حروفچینی و صفحه آرایی: سمیه آهاری

هیئت تحریریه:

فرخ حجت کاشانی (استاد)، دانشگاه علم وصنعت ایران
سید احمدرضا شرافت (استاد)، دانشگاه تربیت مدرّس
مسعود شفیعی (استاد)، دانشگاه صنعتی امیرکبیر
محمدرضا عارف (استاد)، دانشگاه صنعتی شریف
کریم فائز (استاد)، دانشگاه صنعتی امیرکبیر
رضا فرجی دانا (استاد)، دانشگاه تهران
کمال محامدپور (استاد)، دانشگاه خواجه نصیرالدین طوسی
فرخ مروستی (استاد)، دانشگاه صنعتی شریف
سیداحمد معتمدی (استاد)، دانشگاه صنعتی امیرکبیر
روزبه معینی (استاد)، دانشگاه صنعتی امیرکبیر
کاظم اکبری (دانشیار)، دانشگاه صنعتی امیرکبیر
کامبیز بدیع (دانشیار)، مرکز تحقیقات مخابرات ایران
علی اکبر جلالی (دانشیار)، دانشگاه علم و صنعت ایران
حمیدرضا ربیعی (دانشیار)، دانشگاه صنعتی شریف
امیر البدوی (دانشیار)، دانشگاه تربیت مدرس
محمد تشنه لب (دانشیار)، دانشگاه خواجه نصیر طوسی

اعضای هیئت مدیره:

- **اعضای اصلی:**

دکتر مسعود شفیعی (رئیس انجمن)
دانشگاه صنعتی امیرکبیر
دکتر عباس محمدی
دانشگاه صنعتی امیرکبیر
مهندس ناصر یوسف‌پور
شرکت مخابرات ایران
مهندس فرهاد رضایی
شرکت کوشا تجارت نوپدید
دکتر علیرضا اهری لاحق
شرکت داده‌پردازی ایران
دکتر محمد جوادیان
دانشگاه صنعتی کرمانشاه
مهندس سیدعلی صموتی
دانشگاه فنی و حرفه‌ای شهید شمسی‌پور
دکترعلیرضا یاری
پژوهشگاه فناوری اطلاعات و ارتباطات
دکتر سیدعلیرضا هاشمی گلپایگانی
دانشگاه صنعتی امیرکبیر
دکتر اکبر کاری دولت آبادی
دانشگاه شهید ستاری
دکتر جمشید شنبه‌زاده
دانشگاه خوارزمی

- **اعضای علی‌البدل:**

دکتر مهدی شجری (دانشگاه صنعتی امیرکبیر)
مهندس امیرارسلان داداشی‌جوردهی (شرکت ارتباطات سیار ایران)

بازرسان:

بازرس اصلی:
دکتر محمود کمره (دانشگاه تهران)
بازرس علی‌البدل:
مهندس امیر شهاب شاهمیری (دانشگاه جامع علمی کاربردی)

براساس نامهٔ شمارهٔ ۳/۴۸۱۷ مورّخ ۱۳۸۶/۶/۱۱ دفتر امور پژوهشی وزارت علوم، تحقیقات و فناوری و بیانیهٔ رأی کمیسیون بررسی نشریات علمی‌کشور درتاریخ ۱۳۸۶/۴/۲۳، درجهٔ علمی- پژوهشی به این فصلنامه اعطا شده است

نشانی: تهران- خیابان حافظ- دانشگاه صنعتی امیرکبیر- ساختمان ابوریحان- طبقه ششم- اطاق ۶۱۲
تلفن: ۶۶۴۸۵۸۵۶ ، تلفکس: ۶۶۴۹۵۴۳۳

فهرست

فصلنامه علمی- پژوهشی فناوری اطلاعات و ارتباطات ایران

سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷

- **تعیین عوامل موثر بر تامین مالی جمعی شرکت‌های دانش بنیان IT** ۱

  علی حاجی غلام سریزدی، علی رجب زاده قطری، علینقی مشایخی، علیرضا حسن زاده

- **ارائه متدولوژی انتخاب استراتژی مدیریت پروژه‌های فناوری اطلاعات** ۱۷

  مونا جامی‌پور، سید محمدباقر جعفری، مرضیه نصراللهی

- **طراحی پردازنده مبتنی بر FPGA برای الگوریتم‌های رمزنگاری سری SHA-2** ۳۵

  ندا صدق اهرابی، محمدعلی جبرئیل‌جمالی

- **یادگیری رتبه‌بندی محتوای فارسی وب بر مبنای برنامه‌نویسی ژنتیک چند لایه** ۴۵

  امیر حسین کیهانی پور

- **زمانبندی کارها در محیط‌های ابری با استفاده از چارچوب نگاشت – کاهش و الگوریتم ژنتیک** ۷۱

  سید نیما خضر، نیما جعفری نویمی‌پور

- **یک روش جدید حریصانه مبتنی بر مدل آبشاری برای محاسبه‌ی حداکثر سازی نفوذ در شبکه‌های اجتماعی** ۸۵

  عسگرعلی بویر، حمید احمدی بنی

سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷
صص: ۱- ۱۶

فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

تعیین عوامل موثر بر تامین مالی جمعی شرکت‌های دانش بنیان IT

[*]علی حاجی غلام سریزدی [**]علی رجب زاده قطری [***]علینقی مشایخی [***]علیرضا حسن زاده

[*]دکتری، مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران

** دانشیار، مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران

*** استاد، مدیریت، دانشکده مدیریت و اقتصاد، دانشگاه صنعتی شریف، تهران

**** دانشیار، مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران

تاریخ دریافت:۱۳۹۸/۰۳/۱۲ تاریخ پذیرش: ۱۳۹۸/۱۲/۲۰

چکیده

روش تامین مالی جمعی در دنیا بدلیل نیاز به تامین مالی در مراحل اولیه تاسیس کسب و کارهای نوپا و همچنین پیشرفت در فناوری اطلاعات بسرعت گسترش یافته است. در ایران نیز تاکنون چندین پلتفرم تامین مالی ایجاد شده است که بعضی موفق و بعضی ناموفق بوده اند. لذا نیاز است با بررسی عوامل موثر بر این روش به توسعه آن کمک کرد. از آنجا که تامین مالی جمعی پدیده‌ای نو و جدید می‌باشد ضرورت دارد تا با روشی مناسب ضمن تعیین عوامل موثر بر این روش، به افزایش آگاهی آن در جامعه پرداخت. روش مدل‌سازی جمعی مبتنی بر شبکه‌های اجتماعی و وب ۲ با هدف شناخت پدیده‌های جدید می‌باشد. لذا در این مقاله با استفاده از مدل‌سازی جمعی به احصا عوامل موثر بر تامین مالی جمعی در ایران در راستای حمایت از شرکت‌های نوپای حوزه IT پرداخته شده است.

واژه‌های کلیدی: تامین مالی جمعی؛ کسب و کارهای نوپا حوزه IT؛ مدل‌سازی جمعی؛ رویکرد پویایی شناسی سیستم‌ها؛ تلگرام.

نویسندهٔ عهده‌دار مکاتبات: علی حاجی غلام سریزدی A.hajigholam@modares.ac.ir

۱-مقدمه

در عصر اقتصاد دیجیتالی، شرکت‌های دانش بنیان IT یکی از محورهای توسعه هر کشوری می‌باشد. شرکت‌های نوپای دانش بنیان IT در مراحل اولیه تاسیس دارای چالش‌های مختلف از جمله تامین مالی می‌باشند [۷] که این چالش از طریق بانک‌ها و سایر روش‌های تامین مالی سنتی قابل رفع نمی‌باشد [۱۸] و و [۱] از طرف دیگر توسعه اینترنت و ظهور وب ۲ و شبکه‌های اجتماعی زمینه‌ساز ارائه روش‌های جدید تامین مالی مانند تامین مالی جمعی شده‌اند [۲۲] و [۱۴].

طبق تعریف بلفلیم و همکاران[۲] (۲۰۱۴) "تامین مالی جمعی یک دعوت اینترنتی برای جذب منابع مالی در قالب کمک مالی (هدیه)، پیش خرید و کسب جایزه می‌باشد". بنابراین روش تامین مالی جمعی به عنوان یک روش مبتنی بر وب و از طریق انبوه مردم به کمک صاحبان ایده و کسب و کارهای نوپا جهت تجاری‌سازی ایده و راه‌اندازی کسب و کار می‌پردازد.

در جهان این رویکرد بشدت گسترش یافته و دارای رشد نسبتا زیاد و پایدار از جنبه‌های مختلف می‌باشد [۲۲] و [۱۱] و [۲۴] و [۱۷].

در ایران نیز از یک طرف با توجه به تصویب قانون دانش-بنیان و گسترش نهادهای حمایتی شاهد گسترش و ایجاد شرکت‌های نوپای دانش‌بنیان زیادی در سال‌های اخیر می‌باشیم که این شرکت‌ها دارای مشکلات مالی و عدم تامین از طرق سنتی مانند بانک‌ها می‌باشند [۱]. به همین دلیل در سال‌های اخیر تلاش‌هایی برای ارائه روش‌های متنوع و جدید برای حمایت از شرکت‌های نوپای دانش بنیان مانند ایجاد صندوق‌های سرمایه‌گذاری ریسک‌پذیر، بازار فرابورس، مراکز رشد و شتاب‌دهنده‌ها و خیرین فناوری صورت گرفته است [۷]. از طرف دیگر در سال‌های اخیر زیرساخت‌های اینترنت در ایران رشد زیادی داشته و استفاده از شبکه‌های اجتماعی، کارآفرینی از این طریق و فعالیت رسمی و غیررسمی در آن رشد فزاینده ایی داشته است.

در ایران غیر از دو محرک که در دنیا سبب ایجاد و توسعه تامین مالی جمعی شده است استفاده از مشارکت مردمی در جمع آوری کمک‌های مالی دارای سابقه طولانی می‌باشد بعنوان مثال گلریزان یا صندوق‌های قرض الحسنه محلی، وقف و خیرین (مدرسه ساز) نمونه‌هایی از این مشارکت می‌باشند. اما تامین مالی جمعی به شکل امروزی و مبتنی بر اینترنت در ایران تجربه‌ایی جدید است که در چند سال اخیر شاهد آن بوده‌ایم. بطوریکه در این چند سال بیش از ۱۴ پلتفرم فعال با عملکردهای متفاوت ایجاد شده است.

همراه با توسعه و گسترش روش تامین مالی جمعی در دنیا و ایران تحقیقات مختلفی درباره شناخت و استخراج عوامل موثر بر این روش صورت گرفته است که بیشتر مبتنی بر مطالعات تجربی[۳] و با رویکرد اکتشافی بوده اند و در موارد معدودی از نظرات خبرگان یا ذینفعان استفاده شده است [۳۳].

مثلا [۲۰] به طراحی مدل کسب و کار تامین مالی جمعی پرداخته‌اند. [۲۸] ضمن شناخت ذینفعان درگیر در تامین مالی جمعی، خلق ارزش هر یک و انتظارات آن‌ها را از طریق داده‌های تجربی مورد مطالعه قرار داده اند. [۲۴] با مطالعه-ای جامع به دنبال درک کلان از پویایی‌های تجربی تامین مالی در پروژه‌های مختلف بوده است. [۱۳] نحوه انتخاب کارآفرینان و موسسین از دو نوع مدل پیش سفارش و سودمحور را مورد مطالعه قرار داده اند. [۳۳] اثر شبکه اجتماعی کارآفرینان در تامین مالی جمعی را تحلیل و مورد بررسی قرار داده اند. محققانی هم به بررسی مدل‌های مختلف تامین مالی جمعی و اثرات و نقش‌های آن‌ها پرداخته اند [۲۲] و [۱۶] و [۱۴]، ۲۰۱۵؛ [۳۱]. [۲۵] به اهمیت وضع قوانین برای افزایش اعتماد به روش تامین مالی جمعی و گسترش آن پرداخته‌اند. همه این مقالات به بررسی تجربی یافته‌های حاصل از کمپین‌ها و پروژه‌ها پرداخته‌اند. اما در این میان، چو و همکاران[۴] (۲۰۱۶) رضایت اسپانسرها در تامین مالی جمعی را با استفاده از پرسشنامه و از طریق پیمایش مورد بررسی قرار داده اند [۱۶].

[2] Belleflamme et al.

[3] Empirical

[4] Xu et al.

در ایران نیز مطالعات زیادی در این زمینه صورت نگرفته است و مطالعات موجود بیشتر در راستای معرفی این روش می‌باشد با این وجود قربانی، حبیب‌نژاد و سمیع زاده (۱۳۹۳) به شناسایی فاکتورهای مؤثر بر سرمایه‌گذاری جمعی و اولویت‌بندی آن‌ها از نظر سرمایه‌گذار پرداخته‌اند. این فاکتورها شامل منطقه جغرافیای طرح ابتکاری، ارزش مشترک بین کارآفرین و افراد، مبلغ اولیه برای سرمایه-گذاری، مدت زمان بازگشت نتیجه پروژه، میزان و نوع پاداش بازگشتی، اعتماد به کارآفرین، کیفیت طرح پروژه، شبکه اجتماعی مورد استفاده می‌باشد [۸]. چیت‌سازان، کولجی و باقری (۱۳۹۸) وجود مزایای مادی در کنار غیرمادی را از ویژگی های تامین مالی جمعی بر می‌شمرند.

ریاحی و همکاران (۱۳۹۳) برنامه ایی با هدف طراحی و بکارگیری ابزارهای نوین تامین مالی کارآفرینی متکی بر مشارکت آحاد جامعه تدوین کرده‌اند [۷].

بهاری‌فر، شهرابی و جلال (۱۳۹۵) در گزارش خود ضمن معرفی و گونه شناسی مفهوم تأمین مالی جمعی به عنوان یک مفهوم نوین در ادبیات سرمایه گذاری ایران به تشریح چهار نمونه از این روش به طور کامل پرداخته اند. سپس توصیه‌هایی در مورد الزامات قانونی پیشنهادی مرتبط به این مفهوم در کشور بیان کرده اند [۱].

حسن زاده سروستانی، تمیزی فر و سیمیاری (۱۳۹۷) روابط حقوقی حاکم بر تامین مالی جمعی بر جذب مشارکت خیرین را با استفاده از عقود وکالت و هبه تبیین نموده‌اند.

حاجی غلام سریزدی و همکاران (۱۳۹۷) به ارائه مدل بوم کسب و کار تامین مالی جمعی در ایران پرداخته‌اند.

همانطور که قبلا نیز بیان شد بررسی ادبیات موضوع و تحقیقات پیرامون این روش چه در دنیا و چه در ایران نشان می‌دهد که بیشتر تحقیقات بصورت اکتشافی و تجربی بوده و تنها یکی از ابعاد این روش را مورد بررسی قرار داده است. اما از آنجا که موفقیت یا شکست روش‌های جدید خصوصا در عرصه تامین مالی جمعی که با مشارکت مردم صورت می گیرد ارتباط تنگاتنگی با بینش‌ها و رفتارهای اجتماعی دارد بنابراین نیاز است تا با استفاده از نظر مردم و با رویکردی جامع، شناخت درستی از این روش‌ها و کنش‌های مردم درباره آن‌ها ایجاد شود. لذا در این تحقیق از طریق مدل‌سازی جمعی[5] و با استفاده از نظرات خبرگان، ذینفعان و بازیگران این روش (شامل سرمایه گذاران و مردم، سرمایه خواهان و صاحبان کسب وکار حوزه IT، و پلتفرم‌های تامین مالی جمعی) به تعیین عوامل موثر بر تامین مالی جمعی در ایران در راستای حمایت از شرکت‌های نوپای دانش بنیان حوزه IT پرداخته شد.

مدل‌سازی جمعی رویکردی مبتنی بر مدل‌سازی پویایی شناسی سیستم‌ها می‌باشد که با استفاده از هوش جمعی از طریق شبکه‌های اجتماعی و وب ۲ به بررسی پدیده‌های جدید می‌پردازد. در این نوع مدل‌سازی بینش و نظرات جمع و به عبارت دیگر مردم ملاک می‌باشد و هدف از این مدل‌سازی ضمن شناخت دقیق از ابعاد مختلف پدیده مورد مطالعه، اطلاع‌رسانی و آموزش آن نیز در بین مردم می‌باشد. بنابراین در این تحقیق از روش مدل‌سازی جمعی جهت شناخت بهتر تامین مالی جمعی برای حمایت از شرکت‌های نوپای دانش بنیان حوزه IT و همچنین اطلاع رسانی به مردم استفاده شده است.

در ادامه مقاله ابتدا روش تامین مالی جمعی معرفی شده است. سپس روش تحقیق این مقاله که مدل‌سازی جمعی می‌باشد تبیین شده و گام‌های مختلف آن معرفی و تشریح می گردد. بعد از آن نتایج مدل‌سازی جمعی و خروجی‌های مرتبط با عوامل موثر بر تامین مالی جمعی ارائه می گردد.

۲-کسب و کارهای نوپای دانش بنیان IT

در اقتصاد دانش‌بنیان، شرکت‌های دانشی بعنوان یکی از مهمترین نهادها و فناوری اطلاعات بعنوان یکی از پیشران‌های مهم مطرح هستند. لذا کسب و کارهای نوپای دانش‌بنیان حوزه فناوری اطلاعات با دارا بودن هر دو وجه فوق از اهمیت بالایی برخوردارند و بررسی ویژگی‌های آن‌ها در مدیریت شرایط ایجاد و حمایت از این نوع کسب و کارها و تامین منابع مورد نیاز از جمله منابع مالی ضروری می‌باشد. در ادامه ویژگی‌های شرکت‌های نوپای دانش‌بنیان حوزه فناوری اطلاعات آورده شده است.

[5] Crowd Model Building(CMB)

• محصول این شرکت ها، کالا یا خدمات با فناوری بالا در حوزه فناوری اطلاعات حداقل در حد نمونه آزمایشگاهی می‌باشد.

• مهمترین دارایی این شرکت‌ها، نیروی انسانی دانشی (تخصص) و ایده های خلاقانه است که نیاز به حفظ حقوق مالکیت فکری می‌باشد.

• بدلیل ماهیت این شرکت‌ها مانند کوچک بودن و عدم قدرت جذب منابع و ضعف مدیریتی، نیاز به تامین منابع بخصوص منابع مالی (جذب سرمایه) و خدمات مشاوره در ابتدای شکل‌گیری دارند.

• سرعت بالای شکل گیری، بلوغ و شکست و همچنین عدم اطمینان بازار از ویژگی‌های مهم شرکت‌های دانش بنیان فناوری اطلاعات است.

• جذابیت‌های ذاتی صنعت فناوری اطلاعات به ویژه نرخ بازگشت سرمایه و سرعت بازگشت سرمایه نسبتا بالاتر در مقابل سایر صنایع و موفقیت کسب و کارهای نوپای این حوزه در کشور در سال‌های اخیر، باعث روی آوردن بسیاری از نهادهای سرمایه گذار دولتی، عمومی، خصوصی و سرمایه گذاران شخصی به این حوزه شده است.

• عدم آشنایی تخصصی سرمایه‌گذاران سنتی با صنعت فناوری اطلاعات و کسب و کارهای این حوزه.

• تفاوت ماهیت کسب و کارها و ایده‌های حوزه فناوری اطلاعات هم از جهت ارزیابی و هم قیمت‌گذاری باعث عدم درک مناسب و افزایش ریسک سرمایه‌گذاری در این حوزه برای سرمایه‌گذاران سنتی شده است [۱۵] و [۶].

• از طرفی هر چند روش‌های مختلف تامین مالی در مراحل مختلف چرخه عمر کسب و کار متفاوت می‌باشد ولی مهمترین مرحله بخصوص برای شرکت‌های دانش بنیان تامین سرمایه اولیه است. لذا با توجه به ویژگی‌های شرکت‌های نوپای دانش بنیان حوزه فناوری اطلاعات می‌توان گفت که این شرکت‌ها نیاز به تامین مالی در ابتدای شکل گیری خود دارند و این نیاز توسط سرمایه‌گذاران سنتی تامین نمی‌شود و از طرفی آشنایی این کسب و کارها با فناوری اطلاعات بستر خوبی برای استفاده از تامین مالی جمعی است.

۳- معرفی روش تامین مالی جمعی

روش تامین مالی جمعی برگرفته از جمع سپاری[6] می‌باشد. جمع سپاری به استفاده از ظرفیت و منابع مردمی برای انجام کار گفته می‌شود. در تامین مالی جمعی منبع مورد نظر، پول و کار مورد نظر، سرمایه گذاری می‌باشد.

در تعریف تامین مالی جمعی بلفلیم و همکاران (۲۰۱۴) و لامبرت و اسچوینباچر[7] (۲۰۱۰) بطور مشابه مبتنی بر انواع مدل‌های تامین مالی جمعی این روش را اینگونه تعریف کرده اند: "یک دعوت باز اینترنتی برای تامین منابع مالی یا بصورت کمک بلاعوض (هدیه)[8] یا در ازای نوعی از پاداش و یا حق رای[9] به منظور حمایت از طرح‌ها و پروژه‌های با مقاصد خاص". همچنین مولیک (۲۰۱۴) و ولنسین و جیگلوسیوت (۲۰۱۴) براساس فرایند و محتوا تامین ملی جمعی تعریفی نزدیک به هم ارائه کرده اند که عبارتست از: "تامین مالی جمعی اشاره به تلاش کارآفرینان (فرهنگی، اجتماعی و انتفاعی) برای تامین مالی فعالیت‌های مخاطره‌آمیزشان از مشارکت افراد زیاد با سرمایه کم با استفاده از اینترنت بدون واسطه‌ها و الزامات مالی استاندارد دارد" [۲۴] و [۲۸].

تامین مالی جمعی حداقل سه بازیگر اصلی دارد: سرمایه گذاران که همان انبوه مردم مشارکت کننده می‌باشند، پلتفرم‌ها و سایت‌های ارائه دهنده خدمات که درگاه واسط می‌باشند، سرمایه خواهان و متقاضیان وجه که صاحبان ایده و کسب و کارهای نوپا می‌باشند.

هر چند استفاده از مشارکت مردم در جذب منابع مالی به سال‌های دور بر می‌گردد ولی تامین مالی جمعی به شکل امروزی قدمتی ۱۵ ساله دارد. اما مهمترین اتفاق ایجاد سایت Kickstarter بعنوان معروفترین پلتفرم در سال ۲۰۰۹ بود و بعد از آن پلتفرم‌های مختلفی در سراسر دنیا راه اندازی شد. بطوری‌که با اینکه روش تامین مالی جمعی روشی نوظهور می‌باشد ولی دارای رشد چشمگیری از نظر

[6] Crowd Sourcing
[7] Lambert T, Schwienbacher A.
[8] Donation
[9] Reward and/or voting rights

تعداد پلتفرم‌ها، تعداد کمپین‌ها، نرخ موفقیت، میزان سرعت و حجم در جمع‌آوری سرمایه، ارائه مدل‌های مختلف و حوزه‌های تحت پوشش بوده است. طبق آمار مس سلوشن[10] (۲۰۱۳) تعداد پلتفرم‌های تامین مالی جمعی بشدت در حال افزایش است و دارای رشد نمایی می‌باشد بطوریکه پیش بینی‌ها نشان می‌دهد که تعداد آن‌ها از سال ۲۰۰۹ که ۵۳ پلتفرم بوده است به بیش از ۱۰۰۰ پلتفرم در سال‌های اخیر افزایش یافته است. همچنین مبلغ سالیانه مجموع سرمایه‌گذاری‌های صورت گرفته از طریق تامین مالی جمعی در دنیا از ۰,۵۳ میلیارد دلار به ۳۴,۴ میلیارد دلار رسیده است به طوری که در سال ۲۰۱۴ شرکت‌های تجاری و کارآفرینان با جذب ۴۱/۳ درصد از منابع مردم بیشترین سهم را در حوزه‌های مختلف این نوع تأمین مالی داشته اند [۱]. در کنار رشد کمی تعداد پلتفرم‌ها و حجم سرمایه تامین شده، شاخص نرخ موفقیت کمپین‌های تامین مالی جمعی که بیانگر درصد کمپین‌هایی که موفق به رسیدن به هدف مالی می‌شوند می‌باشد رو به افزایش است [۲۴].

در ایران تامین مالی جمعی به شکل امروزی و مبتنی بر وب در چند سال اخیر دارای حدود ۱۴ پلتفرم که بعضی موفق و بعضی ناموفق بودند می‌باشد. از ۱۴ پلتفرم تامین مالی جمعی در ایران، ۳ پلتفرم با مدل اهدا (بخشش)، ۴ پلتفرم با مدل پاداش، ۱ پلتفرم با مدل وام (بدهی)، ۳ پلتفرم با مدل سهام، ۲ پلتفرم با مدل اهدا و پاداش و ۱ پلتفرم با مدل وام و سهام فعالیت دارند. همچنین از ۱۴ پلتفرم تامین مالی جمعی در ایران، ۱۲ پلتفرم با سیاست همه یا هیچ و تنها ۲ پلتفرم با سیاست حفظ سرمایه فعالیت دارند. همچنین تعداد کل کمپین‌ها حدود ۷۳۴ کمپین با نرخ متوسط موفقیت ۵۵ درصد می‌باشد که این نرخ به متوسط جهانی نزدیک می‌باشد.

همانطور که مشخص است در ایران نیز چندین پلتفرم برای تامین مالی جمعی با مدل‌های مختلف در حوزه‌های مختلف به فعالیت پرداخته اند که بعضی موفق بوده بطوریکه نرخ موفقیت آن‌ها در جذب منابع تا ۹۰ درصد می‌باشد و بعضی با عملکرد متوسط و با نرخ موفقیت در حدود ۵۰ درصد و کمتر و بعضی دیگر کاملا ناموفق بوده و شکست خورده اند. بنابراین با توجه به اینکه از یک طرف این پدیده در جهان و همچنین در ایران جدید می‌باشد و دارای جنبه‌های مختلف و پیچیدگی‌های خاص خود است و نیاز به درگیر کردن مردم و جامعه یکی از مولفه‌های اصلی آن می‌باشد و از طرف دیگر از آنجا که تامین مالی جمعی به شکل امروزی در ادبیات موضوع و در مقالات از سال ۲۰۰۶ مطرح شده است که در این مدت تحقیقات و مقالات کمی در این زمینه چه در ایران و چه در دنیا صورت گرفته است که همه مبتنی بر داده‌های تجربی و با رویکرد تجربی و اکتشافی می‌باشد و هیچکدام به بررسی بینش‌ها و نظرات افراد درگیر در این روش (سرمایه گذار، سرمایه خواه، پلتفرم‌های تامین مالی جمعی) نپرداخته اند. بنابراین در این مقاله با استفاده از مدل‌سازی جمعی به تعیین عوامل موثر بر تامین مالی جمعی در ایران در راستای حمایت از شرکت‌های نوپای دانش بنیان حوزه IT پرداخته شده است.

۳- روش تحقیق

در این مقاله از رویکرد پویایی شناسی سیستم‌ها و با استفاده از روش مدل‌سازی جمعی به شناسایی این پدیده پرداخته شده است. در این تحقیق بر اساس مدل ذینفعان ارائه شده توسط ولنسین و جیگلوسیوت (۲۰۱۴) سوالات طراحی شد بصورتی که سوالات بتواند هر چهار ذینفع یعنی سرمایه گذار (مردم عادی)، سرمایه خواه (صاحبان شرکت‌های نوپای دانش بنیان حوزه IT)، پلتفرم‌های تامین مالی جمعی و محیط (سایر ذینفعان) را پوشش دهد. بنابراین روش گردآوری اطلاعات از نوع میدانی و از طریق مصاحبه جمعی و بصورت آنلاین می‌باشد. مصاحبه نیز بصورت نیمه ساختارمند است بصورتی که در هر مرحله سوالی مطرح و براساس آن مباحث آغاز می گردید تا به اشباع نظری برسد و نهایتا جمع بندی از آن سوال صورت گرفته و سپس سوال بعدی مطرح می شد.

در ادامه به تشریح مدل‌سازی جمعی پرداخته شده است و سپس گام‌های مدل‌سازی جمعی بیان و در نهایت نتایج حاصل از این روش ارائه شده است.

[10] Massolution

۴- معرفی مدل‌سازی جمعی

رویکرد پویایی شناسی سیستم‌ها توسط فارستر[11] برای بررسی و مطالعه مسائل درون سیستم‌های پیچیده پویا ایجاد گردید. از ظهور رویکرد پویایی شناسی سیستم‌ها تاکنون این رویکرد از جنبه‌های مختلفی چون حوزه‌های کاربرد، روش‌ها و ابزارهای مورد استفاده، نرم افزارها و منابع اطلاعاتی و روش‌های گرداوری اطلاعات توسعه یافته است (آکرمن و ونیکس، ۱۹۹۶). بعنوان مثال در زمینه توسعه نرم افزارهای مورد استفاده در پویایی شناسی سیستم‌ها امروزه نرم افزارهای آنلاین و تحت وب یا سایت‌هایی که تحت وب قادر به مدل‌سازی پویا می‌باشند توسعه یافته است. با گسترش و توسعه رویکرد پویایی شناسی سیستم‌ها روش‌های متفاوتی برای مدل‌سازی مبتنی بر روی نحوه کسب اطلاعات از منابع مدل‌های ذهنی، پایگاه داده‌های عددی و پایگاه مکتوب [۱۹] برای ساخت مدل ارائه شده است [۲۶].

در زمینه روش‌های مدل‌سازی در ابتدا بیشتر تمرکز بر مدل ذهنی مدلساز و مبتنی بر منابع اطلاعاتی و کتابخانه ایی بود و مدل‌سازی بصورت فردی انجام می‌شد. ولی از آنجا که مسائل امروزی از توان یک فرد برای درک و حل آن خارج بوده و در حوزه مسئولیت یک فرد هم نیست (روس[12]، ۱۹۹۷) مدل‌سازی مشارکتی توسط دانشمندان پویایی شناسی سیستم‌ها ارائه شده است. مدلساز مشارکتی می‌تواند بصورت مدل‌سازی گروهی یا فردی صورت گیرد [۳۰].

در این راستا از یک طرف در مدل‌سازی پویایی شناسی سیستم‌ها بر مشارکت افراد و استفاده از مدل‌های ذهنی آن‌ها تاکید شده است و از طرف دیگر با گسترش اینترنت و ظهور شبکه‌های اجتماعی بستر خوبی جهت استفاده از جمع برای کارهای مختلفی از جمله تسهیم دانش و ایده از طریق جمع سپاری، تسهیم منابع مالی از طریق تامین مالی جمعی و ... در حال افزایش است. بنابراین از طریق شبکه‌های اجتماعی می‌توان به مدل‌سازی مشارکتی بصورت جمعی پرداخت.

منظور از مدل‌سازی جمعی استفاده از وب ۲ و شبکه‌های اجتماعی برای گردآوری جمع کثیری از افراد برای شناسایی مسئله پویای جدید و اکتشافی، گردآوری دیدگاه‌های مختلف، تحلیل، مدل‌سازی و ارائه نتایج آن می‌باشد. بنابراین در مدل‌سازی جمعی تاکید بر مشارکت اجتماعی می‌باشد بصورتی که این مدل‌سازی شامل گروه خیلی بزرگ، نامحدود و نامشخص از تمامی افراد مرتبط و متخصص یا نامرتبط می‌باشد. معمولا تیم مدلساز یک فراخوان عمومی از طریق اینترنت برای دعوت افراد به مدل‌سازی برگزار می‌کند. دیدگاه این روش واگرا و برای مسائل جدید و تحقیق اکتشافی که نیاز به جمع آوری نظرات جامعه و هوش جمعی دارد می‌باشد. البته معمولا از این روش برای موضوعاتی که متخصص خاصی برای آن وجود ندارد استفاده می شود. در اینجا درگیرکردن کاربران به دو دلیل کسب دانش و انتشار دانش[13] اهمیت دارد [۳۰].

۵- مدل‌سازی جمعی

برای مدل‌سازی جمعی ۷ گام باید صورت گیرد که در ادامه به شرح این گام‌ها که در این تحقیق انجام شده است پرداخته شده است:

گام اول - تعیین بستر فناوری مدل‌سازی: در این گام بایستی بستر فناوری جهت استفاده برای مدل‌سازی انتخاب شود. که در این تحقیق شبکه اجتماعی تلگرام بدلیل گستردگی آن در ایران و استفاده از این شبکه در محافل علمی و فناوری جهت مدل‌سازی انتخاب شد. طبق نتایج نظرسنجی‌ها شبکه اجتماعی تلگرام پرطرفداترین شبکه اجتماعی در ایران می‌باشد[۵].

گام دوم – آماده سازی بستر فناوری مدل‌سازی انتخاب شده: در این مرحله بایستی بستر انتخاب شده جهت مدل‌سازی آماده گردد. به همین منظور ابتدا در تلگرام ابرگروه[14] با نام تامین مالی استارت آپ‌ها ایجاد گردید و مطالب و سوالات برای شروع مدل‌سازی آماده شد. این مطالب شامل معرفی گروه و هدف از ایجاد آن، معرفی تامین

[11] Forrester
[12] Roos
[13] Knowledge acquisition and diffusion
[14] Super group

مالی جمعی و ارائه منابع مفید مانند فیلم‌ها، گزارش‌ها، مقالات و آدرس سایت‌ها و پلتفرم‌های معروف (داخلی و خارجی) در جهت اطلاع رسانی و آشنایی افراد، معرفی مدل‌سازی جمعی و روش کار و همچنین سوالات تحقیق بود.

گام سوم – تعیین جامعه هدف اولیه: پس از ایجاد بستر فناوری نیاز است تا گروه‌های مختلفی که بیشترین ارتباط را با پدیده یا مسئله مد نظر دارند یا کسانی که تیم مدلساز علاقمند است تا آن‌ها در مدل‌سازی حضور داشته باشند بعنوان جامعه هدف اولیه شناسایی شوند. جامعه آماری در جهت استفاده از روش مدل‌سازی جمعی در این تحقیق شامل تمام افرادی که در بازه زمانی جمع آوری داده در این فرایند شرکت می‌کنند می‌باشد. این افراد در برگیرنده مدیران شرکت‌های نوپای دانش بنیان حوزه IT (شرکت‌های مستقر در پارک‌های علم و فناوری و یا تحت مجوز دانش بنیان معاونت علمی و فناوری ریاست جمهوری)، مدیران پلتفرم‌های ارائه دهنده تامین مالی جمعی و سرمایه گذاران (مردم عادی) می‌باشد.

گام چهارم - اطلاع رسانی و عضوگیری: از آنجا که مدل‌سازی جمعی به مشارکت عموم نیازمند است بایستی بعد از شناسایی جامعه هدف اولیه اطلاع رسانی لازم هم برای جامعه هدف اولیه و هم عموم از طریق شبکه‌های اجتماعی، اینترنت یا سایر وسایل ارتباط جمعی و همچنین مکاتبات رسمی صورت گیرد. بدین منظور از دو طریق رسمی (ارسال نامه رسمی به روسای پارک‌های علم و فناوری (مراکز رشد فناوری اطلاعات)، سرپرست امور شرکت‌ها و موسسات دانش بنیان معاونت علمی و فناوری ریاست جمهوری، رئیس صندوق نوآوری و شکوفایی، رئیس انجمن مدیریت فناوری ایران، دبیر ستاد خیرین فناوری، هیئت امنا بنیاد حامیان علم و فناوری، مدیر عامل انجمن خیرین توسعه علم و فناوری ایرانیان، رئیس صندوق پژوهش و فناوری استان یزد) و غیر رسمی (گروه‌ها و کانال‌های مختلف تلگرام (مانند گروه و کانال فناوری پارک علم و فناوری یزد، کانال دونیت، کانال دانشگاه یزد، کانال انجمن مدیریت فناوری ایران، کانال صندوق پژوهش و فناوری استان یزد، کانال شرکتهای دانش بنیان مربوط به معاونت علمی ریاست جمهوری و همچنین در شبکه لینکدین و همچنین از طریق ایمیل نیز به افراد مختلف خصوصا صاحبان پلتفرم‌های تامین مالی جمعی در ایران) اطلاع رسانی صورت گرفت.

گام پنجم – انجام مدل‌سازی: با عضویت تعداد قابل قبولی از اعضا در پلتفرم فناوری مورد نظر فرایند مدل‌سازی آغاز می گردد. در این گام بهتر است ابتدا معرفی و شرح مختصری از مدل‌سازی جمعی و مسئله مدنظر ارائه شود. سپس سوالات مرتبط با پدیده یا مسئله در جمع مطرح شود و نظرات جمع گردآوری گردد. مدل‌سازی هنگامی پایان می پذیرد که یا به اشباع نظری برسیم و یا اینکه بازه زمانی پروژه اتمام شده باشد. در این تحقیق یک روز بعد از ایجاد گروه تعداد اعضا به حد قابل قبول رسید و در تاریخ ۲۵ / ۶ / ۱۳۹۵ و به مدت یک هفته هدف از گروه و همچنین مدل‌سازی جمعی و تامین مالی جمعی با ارائه منابعی که در گام دوم آماده شده بود معرفی گردید و به سوالات در این زمینه پاسخ داده و گروه آماده پاسخ به سوالات شد.

با توجه به اینکه هدف این تحقیق بررسی عوامل موثر بر تامین مالی جمعی برای حمایت از شرکت‌های دانش بنیان نوپا (استارت‌آپ‌ها) خصوصا در حوزه IT در ایران می‌باشد در طراحی سوالات بایستی هر چهار طرف یعنی سرمایه گذار، سرمایه خواه، پلتفرم‌های تامین مالی جمعی و محیط مد نظر قرار می گرفت و عوامل رشد دهنده و چالش‌های هریک مورد سوال واقع شود و همچنین سیاست‌ها یا راهکارهایی جهت توسعه این روش در ایران پیشنهاد می شد بنابراین سوالات ۹ گانه زیر با این هدف طراحی گردید.

۱- چه عواملی (مرتبط با سرمایه گذار یا همان مردم) سبب رونق روش تامین الی جمعی و سرمایه گذاری بیشتر و مشارکت بیشتر افراد برای سرمایه گذاری می شود؟

۲- چه عواملی (مرتبط با سرمایه گذار یا همان مردم) بعنوان مانعی بر سر راه این روش وجود دارد؟

۳- چه عواملی (مرتبط با سرمایه خواه یا همان شرکت‌های دانش بنیان نوپای حوزه IT) سبب رونق این روش و سرمایه

گذاری بیشتر و مشارکت بیشتر افراد برای سرمایه گذاری می شود؟

۴- چه عواملی (مرتبط با سرمایه خواه یا همان شرکت‌های دانش بنیان نوپای حوزه IT) بعنوان مانعی بر سر راه این روش وجود دارد؟

۵- چه عواملی (مرتبط با پلتفرم یا همان سایت‌های ارائه دهنده تامین مالی جمعی) سبب رونق این روش و سرمایه گذاری بیشتر و مشارکت بیشتر افراد برای سرمایه گذاری می شود؟

۶- چه عواملی (مرتبط با پلتفرم یا همان سایت‌های ارائه دهنده تامین مالی جمعی) بعنوان مانعی بر سر راه این روش وجود دارد؟

۷- چه عواملی در محیط (مانند دولت، قوانین و مقررات، شرایط کسب و کار، زیرساخت‌های فناوری اطلاعات و ...) سبب رونق این روش و سرمایه گذاری بیشتر و مشارکت بیشتر افراد برای سرمایه گذاری می شود؟

۸- چه عواملی در محیط (مانند دولت، قوانین و مقررات، شرایط کسب و کار، زیرساخت‌های فناوری اطلاعات و ...) بعنوان مانعی بر سر راه این روش وجود دارد؟

۹- چه سیاست‌ها یا راهکارهایی در ایران برای توسعه تامین مالی جمعی جهت حمایت از کسب وکارهای نوپای دانش بنیان حوزه IT می توان ارائه نمود؟

در رابطه با نحوه برگزاری مدل‌سازی و ارائه و جمع آوری پاسخ‌ها بدین گونه عمل شد که در تاریخ ۰۲ / ۷ / ۱۳۹۵ سوال اول و دوم مطرح و مدت زمان اولیه یک هفته جهت پاسخ تعیین گردید. با این حال با رسیدن به اشباع نظری پاسخ در روز شش ام به پایان رسید و در روز هفتم به بحث و جمع بندی نتایج پرداخته شد و به همین منوال سوالات بعدی مطرح و پاسخ‌ها احصا و جمع بندی آن‌ها صورت گرفت.

گام ششم – جمع بندی و نتیجه گیری: با اتمام زمان مدل‌سازی نتایج حاصل از مباحث مدل‌سازی جمع بندی و به جمع ارائه می گردد تا تایید شود. سپس نتایج حاصل از مدل‌سازی در فرمت مناسب تدوین می گردد. برای این منظور

نتایج نهایی در گروه ارائه شد و به مدت یک هفته جهت اصلاح و یا تکمیل آن‌ها زمان گذاشته شد که در این مدت افراد ضمن

تایید نتایج، پیشنهاداتی جهت اصلاح عوامل موثر ارائه کردند که با تایید و اجماع اعضا مورد بررسی و اعمال قرار گرفت. در جدول ۱ فعالیت‌های صورت گرفته و توضیحات آن ارائه شده است

جدول ۱: فعالیت‌های صورت گرفته و تشریح آن

ردیف	نام فعالیت		زمان	توضیحات
۱	ایجاد گروه تامین مالی استارت آپ‌ها در تلگرام		۲۴ / ۶ / ۱۳۹۵ الی ۱۷ / ۸ / ۱۳۹۵	این گروه با هدف بحث درباره سوالات مشخص تحقیق ایجاد شد.
۲	آماده سازی و تهیه مطالب مورد نیاز		۲۰ / ۶ / ۱۳۹۵ الی ۱ / ۷ /۱۳۹۵	مطالب مرتبط با معرفی گروه، معرفی مدل‌سازی جمعی، معرفی تامین مالی جمعی و سوالات تحقیق آماده گردید.
۳	اطلاع رسانی و عضوگیری		۲۴ / ۶ / ۱۳۹۵ الی ۱۷ / ۸ / ۱۳۹۵	از طریق رسمی و غیررسمی اطلاع رسانی صورت گرفت.
۴	انجام مدل‌سازی جمعی	معرفی گروه	۲۵ / ۶ / ۱۳۹۵ الی ۰۲ / ۷ / ۱۳۹۵	مطالبی که در مرحله آماده سازی تهیه شده بود در گروه بارگذاری گردید.
		سوال اول و دوم	۰۲ / ۷ / ۱۳۹۵ الی ۰۹ / ۷ / ۱۳۹۵	سوالات در روز ۲ / ۷/ ۹۵ مطرح شد و پاسخ به آن تا اشباع نظری یا پایان مهلت پاسخ(یک هفته) ادامه یافت.
		سوال سوم و چهارم	۱۰ / ۷ / ۱۳۹۵ الی ۲۳ / ۷ / ۱۳۹۵	سوال سوم و چهارم در ساعت ۱۹ روز ۱۰ / ۷/ ۹۵ مطرح شد.
		سوال پنجم و ششم	۲۴ / ۷ / ۱۳۹۵ الی	سوال سوم و چهارم در ساعت ۲۱ روز ۲۴ / ۷/ ۹۵ مطرح

			۱۳۹۵ / ۷ / ۲۹	شد.
		سوال هفتم و هشتم	۳۰ / ۷ / ۱۳۹۵ الی ۶ / ۸ / ۱۳۹۵	سوال سوم و چهارم در ساعت ۲۳ روز ۳۰ / ۷/ ۹۵ مطرح شد.
		سوال نهم	۷ / ۸ / ۱۳۹۵ الی ۱۰ / ۸ / ۱۳۹۵	سوال نهم در ساعت ۲۳ روز ۳۰ / ۷/ ۹۵ مطرح شد.
۵	جمع بندی و نتیجه گیری		۱۰ / ۸ / ۱۳۹۵ الی ۱۷ / ۸ / ۱۳۹۵	-

۶- نتایج حاصل از مدل‌سازی جمعی

در این قسمت نتایج حاصل از مدل‌سازی جمعی در استخراج عوامل و متغیرهای موثر بر تامین مالی جمعی برای حمایت از شرکت‌های نوپای دانش بنیان حوزه IT به تفکیک هر ذینفع آورده شده است. ذینفعان براساس مدل ذینفعان ارائه شده توسط ولنسین و جیگلوسیوت(۲۰۱۴) شامل سرمایه-گذار (مردم عادی)، سرمایه‌خواه (صاحبان شرکت‌های نوپای دانش بنیان حوزه IT)، پلتفرم‌های تامین مالی جمعی و محیط (سایر ذینفعان) می‌باشند.

۶-۱- عوامل موثر بر سرمایه گذاران

در جداول زیر عواملی که سبب رونق و گسترش تامین مالی جمعی و همچنین مانع رشد آن در طرف سرمایه گذار می‌باشد آورده شده است. در طرف سرمایه گذار ۲۴ متغیر شناسایی شد که ۱۷ متغیر مثبت و ۷ متغیر منفی می‌باشند. در جدول ۲ عوامل مثبت که در طرف سرمایه گذار سبب رشد تامین مالی جمعی می شوند آورده شده است.

جدول شماره ۲: عوامل مثبت و رشد دهنده طرف سرمایه گذار

ردیف	نام متغیر	فراوانی
۱	انگیزه‌های مالی و غیر مالی و نیت‌های خیرخواهانه	۲۲
۲	تاثیر میزان موفقیت کمپین‌ها بر مردم	۲۰
۳	برنامه‌های ترویجی و استفاده از افراد مورد اعتماد	۱۶
۴	فعالیت گسترده در شبکه‌های اجتماعی	۱۲
۵	تمایل به کارآفرینی در جامعه	۱۰
۶	قدرت شبکه سازی و دسترسی به نزدیکان و حلقه معتمد	۶
۷	وجود تحصیلکرده‌های بیکار	۴
۸	ترکیب جمعیتی جوان کشور به همراه ایده‌های نوین	۴
۹	آمادگی ذهنی جامعه	۴
۱۰	دسترسی راحت برای جامعه	۴
۱۱	تطابق خروجی دانش بنیان با موضوعات فرهنگی – عقیدتی مردم	۴
۱۲	حذف واسطه‌های سرمایه گذاری و ایجاد ارتباط مستقیم سرمایه گذار در طرح	۴
۱۳	همراهی مردم از کمپین‌ها و اهداف خیرخواهانه با تبلیغات مناسب	۲
۱۴	بازدهی نامناسب بانک‌ها برای سرمایه‌های اندک	۲
۱۵	پشتیبانی سرمایه گذار از پروژه‌های سرمایه گذاری شده	۲

۱۶	نیاز به ایجاد هیجان در مردم	۲
۱۷	علاقه‌مندی به مشارکت داشتن در پیشبرد پروژه‌ای مشخص	۲

در جدول ۳ عوامل منفی که در طرف سرمایه گذار از رشد تامین مالی جمعی جلوگیری می کنند آورده شده است.

جدول۳: عوامل منفی و کند کننده طرف سرمایه گذار

ردیف	نام متغیر	فراوانی
۱	عدم اعتماد به روش‌های جدید برای انجام امور مالی و نیاز به تضمین سرمایه گذاری توسط سایر نهادها	۳۴
۲	عدم آگاهی مردم درباره تامین مالی جمعی و مزایای آن	۲۲
۳	فرهنگ نادرست استفاده از شبکه‌های اجتماعی در ایران	۱۸
۴	عدم ریسک پذیری مردم	۱۴
۵	عدم روحیه مشارکت پذیری	۶
۶	حضور غیر موثر در شبکه‌های اجتماعی	۴
۷	سواد مالی پایین جامعه	۴

همانطور که از جداول فوق مشخص است دو متغیر اول با بیشترین فراوانی مربوط به عوامل منفی و جز چالش‌های تامین مالی جمعی می‌باشند که به ترتیب با فراوانی ۳۴ و ۲۲ می‌باشند و در رتبه سوم متغیر مثبت و رشد دهنده با فراوانی ۲۲ می‌باشد.

۶-۲-عوامل موثر بر سرمایه‌خواهان

در جداول زیر عواملی که سبب رونق و گسترش تامین مالی جمعی و همچنین مانع رشد آن در طرف سرمایه خواه(شرکت‌های نوپای دانش بنیان حوزه IT) می‌باشد آورده شده است. در طرف سرمایه خواه ۲۲ متغیر شناسایی شد که ۱۵ متغیر مثبت و ۷ متغیر منفی می‌باشند. در جدول ۴ عوامل مثبت که در طرف سرمایه خواه سبب رشد تامین مالی جمعی می شوند آورده شده است.

جدول ۴: عوامل مثبت و رشد دهنده طرف سرمایه خواه

ردیف	نام متغیر	فراوانی
۱	داشتن ایده یا طرحی نو و کاربردی و مردم پسند	۶۴
۲	معرفی موثر و اطلاع رسانی مناسب کسب و کار و ایده جهت جذب حامی در طول اجرای طرح	۶۴
۳	مشکلات مالی شرکت‌های نوپای دانش بنیان	۴۸
۴	اشنایی با توانمندی‌ها و اگاهی بیشتر درباره روش‌های جدید مبتنی بر IT	۴۸
۵	اعتماد بیشتر به روش‌های مبتنی بر IT	۴۰
۶	داشتن روحیه فعالیت‌های تیمی و منفعت رسانی به جمع	۴۰
۷	جذب حامی مالی و غیرمالی	۳۲
۸	معرفی و تبلیغ طرح به مردم	۲۴
۹	حضور و فعالیت گسترده در شبکه‌های اجتماعی	۱۶
۱۰	شناسایی بازار هدف و بازاریابی بر مبنای نیاز مشتریان	۱۶
۱۱	اطمینان بالا از موفقیت ایده و طرح کسب و کار خود	۱۶
۱۲	کشش بازار برای راه اندازی کسب و کاری نوپا در حوزه IT با سرمایه نسبتا اندک	۸

۱۳	اعتبار شرکت و داشتن رزومه مناسب	۸
۱۴	ویژگی‌های شخصیتی سرمایه خواه	۸
۱۵	حوزه ی طرح (مانند حوزه‌های اجتماعی که در شرایط فعلی استقبال بیشتری دارد)	۸

در جدول ۵ عوامل منفی که در طرف سرمایه خواه از رشد تامین مالی جمعی جلوگیری می کنند آورده شده است.

جدول ۵: عوامل منفی و کند کننده طرف سرمایه‌خواه

ردیف	نام متغیر	فراوانی
۱	مساعد نبودن شرایط کشور برای وفاداری به کسب و کارهای نو	۳۲
۲	تمرکز بر حوزه فنی و عدم تقویت حوزه مدیریتی کسب وکار خصوصا مالی (عدم سواد مالی)	۱۶
۳	آشکار شدن ایده برای رقبا	۱۶
۴	غرور ناشی از کارافرینی	۸
۵	عدم اشنایی با تفکر صحیح جهت حل مسئله و برون فکنی	۸
۶	ارائه تصویر غلط از شرکت و کاهش اعتبار آن	۸
۷	ناملموس بودن برخی از طرحهای حوزه IT	۸

همانطور که مشخص است متغیرهای با فراوانی زیاد بیشتر مربوط به عوامل مثبت و رشد دهنده بوده و بدلیل شناخت و آگاهی شرکت‌های‌های حوزه IT از روش تامین مالی جمعی که مبتنی بر IT است و همچنین مشکلات آن‌ها در تامین منابع از سایر روش‌ها می‌باشد.

۶-۳- عوامل موثر بر پلتفرم‌های تامین مالی جمعی

در جداول زیر عواملی که سبب رونق و گسترش تامین مالی جمعی و همچنین مانع رشد آن در طرف پلتفرم می‌باشد آورده شده است. در طرف پلتفرم‌ها ۱۷ متغیر شناسایی شد که متغیر ۱۰ مثبت و ۷ متغیر منفی می‌باشند. در جدول ۶ عوامل مثبت که در طرف پلتفرم تامین مالی سبب رشد تامین مالی جمعی می شوند آورده شده است.

جدول ۶: عوامل مثبت و رشددهنده طرف پلتفرم

ردیف	نام متغیر	فراوانی
۱	طراحی دقیق و حرفه ای پلترفرم (محتوا و ظاهر) با عنوان جذاب و راحت سایت	۷۲
۲	سهولت در ارائه خدمات و دسترسی سریع و مقید نبودن به زمان و مکان	۶۴
۳	استمرار در ارائه خدمات و به روز بودن از جنبه‌های مختلف	۴۸
۴	امکانات فضای مجازی (مانند ارائه گزارشات تحلیلی، دسترسی به سوابق و ...)	۱۶
۵	قابلیت یکپارچه سازی با فعالیت‌های مالی افراد در بستر وب	۱۶
۶	استفاده از تکنولوژی‌های روز (اپلیکیشن‌های مفید)	۱۶
۷	بومی بودن سایت	۸
۸	تبلیغ مراکز رشد و پارک‌ها در این زمینه	۸
۹	انتخاب نام دامنه سایت متناسب با فعالیت اقتصادی	۸
۱۰	قابلیت ردگیری سرمایه توسط سرمایه گذار	۸

در جدول ۷ عوامل منفی که در طرف پلتفرم از رشد تامین مالی جمعی جلوگیری می کند آورده شده است.

جدول ۷: عوامل منفی و کند کننده طرف پلتفرم

ردیف	نام متغیر	فراوانی
۱	اطلاع رسانی ضعیف و عدم شناخت کافی از این پلتفرم ها	۸۰
۲	عدم اعتماد به سایت‌ها و نیاز به ایجاد امنیت	۵۶
۳	عدم شبکه سازی قوی و بلندمدت بین بازیگران مختلف تامین مالی جمعی و ایجاد مکانیزم‌های بازخورد	۴۸
۴	عدم حضور فعال در شبکه‌های اجتماعی واقعی و مجازی	۴۰
۵	ملاحظات حقوقی در خصوص فعالیت‌های تجاری در فضای مجازی	۲۴
۶	عدم انجام مطالعات اولیه و عدم شفافیت در اهداف	۱۶
۷	ضعیف بودن فرهنگ سرمایه گذاری از این طریق	۸

همانطور که مشخص است متغیرهای با فراوانی زیاد بیشتر مربوط به عوامل مثبت و رشد دهنده می‌باشد. همچنین عدم اعتماد به آن‌ها و عدم شبکه سازی قوی آن‌ها در دنیای مجازی و واقعی از ضعف‌ها و چالش‌های پلتفرم‌های تامین مالی جمعی در ایران می‌باشد.

۶-۴- عوامل محیطی موثر بر تامین مالی جمعی

در جداول زیر عواملی که سبب رونق و گسترش تامین مالی جمعی و همچنین مانع رشد آن در طرف محیط می‌باشد آورده شده است. در طرف محیط ۱۰ متغیر شناسایی شد که متغیر ۷ مثبت و ۳ متغیر منفی می‌باشند. در جدول ۸ عوامل مثبت که در طرف محیط سبب رشد تامین مالی جمعی می شوند آورده شده است.

جدول ۸: عوامل مثبت و رشد دهنده طرف محیط

ردیف	نام متغیر	فراوانی
۱	فرهنگ‌سازی استفاده از سرمایه گذاری جمعی توسط صاحبان کسب و کار و ترویج نمونه‌های موفق	۴۸
۲	زیرساخت فناوری (اینترنت و شبکه‌های اجتماعی)	۴۰
۳	حمایت دولت از سرمایه گذاری جمعی	۴۰
۴	مشکلات اقتصادی و رکود (تحریم)	۳۲
۵	نقدینگی در درست مردم	۲۴
۶	ورود موج در بازار و تغییر بازارهای سرمایه پذیر و ناامنی در آن‌ها(ریسک سرمایه گذاری)	۲۴
۷	مشکلات در تامین سرمایه از سایر روش‌ها (بروکراسی و ضمانت‌های زیاد در بانک‌ها و محدودیت آن‌ها و ...)	۱۶

در جدول ۹ عوامل منفی که در طرف محیط از رشد تامین مالی جمعی جلوگیری می کند آورده شده است.

جدول ۹: عوامل منفی و کند کننده طرف محیط

ردیف	نام متغیر	فراوانی
۱	عدم آشنایی نهادهای حامی فناوری و جامعه از روش‌های جدید	۴۰
۲	عدم وجود قوانین حمایت و شفاف (زیرساخت حقوقی)	۳۲
۳	فرهنگ و تجربه مردم بصورت سنتی و در مقابل عدم اعتماد به این روش و تجربه در اینترنت	۱۶

همانطور که مشخص است اطلاع رسانی و اگاهی بخشی، زیرساخت‌های فناوری و حمایت دولت از عوامل رشد و نبود قوانین مرتبط از عوامل منفی می‌باشد.

۶-۵-عوامل محیطی موثر بر تامین مالی جمعی

سوال نهم بدنبال شناسایی راهکارها و سیاست‌های پیشنهادی جهت توسعه و گسترش تامین مالی جمعی در ایران است. همانطور که در جدول ۱۰ مشخص است بدلیل عدم آشنایی مردم و شرکت‌های نوپای دانش بنیان حوزه IT با روش تامین مالی جمعی و جدید بودن آن نیاز است تا این روش با تبلیغات مناسب به مردم خصوصا خیرین اطلاع رسانی و آموزش داده شود. همچنین قوانین و سازوکارهای

اعتماد افزا و حمایت دولت از این پلتفرم‌ها سبب توسعه و گسترش این روش تامین مالی جمعی در ایران می شود.

جدول ۱۰: راهکارها و سیاست‌های پیشنهادی

ردیف	نام متغیر	فراوانی
۱	تبلیغات موثر و آموزش (خصوصا در رابطه با خیرین)	۷۲
۲	تدوین قوانین مرتبط با تامین مالی جمعی	۴۸
۳	ایجاد سازوکارهای تضمین تامین مالی جمعی	۴۰
۴	حمایت دولت	۴۰

۷- نتیجه‌گیری

نیاز به تامین مالی در مراحل اولیه تاسیس شرکت‌های نوپای دانش بنیان و پیشرفت در فناوری اطلاعات سبب شده است تا تامین مالی جمعی در جهان و ایران ایجاد و توسعه یابد. بررسی ادبیات موضوع این روش نشان می‌دهد که مطالعات پیرامون آن بیشتر با رویکرد اکتشافی و تجربی به بررسی یکی از جنبه‌های این روش پرداخته اند.

در ایران نیز تامین مالی جمعی به شکل امروزی و مبتنی بر وب تجربه ایی جدید می‌باشد که در چند سال اخیر حدود ۱۴ پلتفرم ایجاد شده اند که بعضی موفق و بعضی ناموفق بوده اند.

اما از آنجا که روش تامین مالی جمعی مبتنی بر مشارکت مردم می‌باشد و ارتباط تنگاتنگی با بینش‌ها و رفتارهای اجتماعی دارد بنابراین نیاز است تا با استفاده از نظر مردم و با رویکردی جامع، شناخت درستی از این روش‌ها و کنش‌های مردم درباره آن‌ها ایجاد شود. لذا در این تحقیق از طریق مدل‌سازی جمعی و با استفاده از نظرات خبرگان، ذینفعان و بازیگران این روش به استخراج عوامل موثر بر تامین مالی جمعی در ایران در راستای حمایت از شرکت‌های نوپای دانش بنیان حوزه IT پرداخته شد.

منظور از مدل‌سازی جمعی استفاده از وب ۲ و شبکه‌های اجتماعی برای گردآوری جمع کثیری از افراد برای شناسایی مسئله پویای جدید و اکتشافی، گردآوری دیدگاه‌های مختلف، تحلیل، مدل‌سازی و ارائه نتایج آن می‌باشد. در این تحقیق گام‌های مدل‌سازی جمعی از طریق شبکه اجتماعی تلگرام در
طول ۲ ماه انجام شد.

با استفاده از مدل‌سازی جمعی عوامل موثر بر تامین مالی جمعی برای حمایت از شرکت‌های نوپای دانش بنیان استخراج شد که در ۴ دسته ذینفعان (سرمایه گذاران، سرمایه خواهان، پلتفرم‌ها و محیط) دسته بندی شدند که در هر دسته به ترتیب
۲۴، ۲۲، ۱۷ و ۱۰ متغیر در قالب دو دسته عوامل مثبت و رشد دهنده و عوامل منفی و کند کننده توسعه تامین مالی جمعی شناسایی شد. در انتها نیز براساس عوامل استخراج شده و نظر اعضا، راهکارها و سیاست‌هایی جهت توسعه این روش در ایران پیشنهاد شد که عبارتند از: تبلیغات موثر و آموزش (خصوصا در رابطه با خیرین)، تدوین قوانین مرتبط با تامین مالی جمعی، ایجاد سازوکارهای تضمین تامین مالی جمعی و حمایت دولت.

منابع

۱. ح، بهاری فر، م. ج، محسن شهرابی، الگوی تأمین مالی بنگاههای کوچک و متوسط. تأمین مالی جمعی: مفاهیم، مدلها و ملاحظات قانونگذاری، دفتر مطالعات اقتصادی (گروه مطالعاتی محیط کسب وکار) معاونت پژوهشهای اقتصادی.

۲. ه، چیت سازان، س، کولجی، ا، باقری. شناسایی عوامل پیش‌برنده استفاده از تأمین مالی جمعی توسط کارآفرینان (مورد مطالعه: صنعت گردشگری). فصلنامه علمی پژوهشی توسعه کارآفرینی، ۱۲(۱)، ص ۲۱-۴۰.

۳. ع، حاجی غلام سریزدی، ع، رجب زاده قطری، ع، مشایخی، ع، حسن زاده، ارائه مدل بوم شناسی کسب و کار تامین مالی جمعی در ایران، دو فصلنامه علمی پژوهشی کاوش‌های مدیریت بازرگانی یزد، سال دهم، شماره ۱۹، بهار و تابستان ۱۳۹۷، صفحه ۳۰۷-۳۴۲.

۴. ح، حسن زاده سروستانی، م، تمیزی فر، م، سیمیاری، تأمین مالی جمعی، الگویی مناسب جهت جذب مشارکتهای خیّرین. فصلنامه حقیقات مالی اسلامی، سال هشتم، شماره اول (پیاپی ۱۵)، پاییز و زمستان ۱۳۹۷، ص ۵۹ – ۹۰.

۵. خبرگذاری دانشگاه آزاد اسلامی: آنا، آمار جالب کاربران تلگرام در ایران، (http://www.ana.ir/news/75610).

۶. م، خیاطیان، س. ح، طباطبائیان، م، امیری، م، الیاسی، تحلیلی بر عوامل موثر بر رشد و پایداری شرکتهای دانش بنیان در ایران، فصلنامه نوآوری و ارزش افرینی، سال سوم، شماره ششم، پایز و زمستان ۱۳۹۳.

۷. پ، ریاحی، و همکاران، طراحی شیوه‌های مختلف تامین مالی مردم نهاد کارآفرینی، معاونت توسعه کارآفرینی و اشتغال وزارت تعاون، کار و رفاه اجتماعی، اسفند ۱۳۹۳.

۸. ن، قربانی، ح، حبیب نژاد، ر، سمیع زاده، مطالعه تحقیقی در تعیین اهمیت فاکتورهای موثر بر سرمایه گذاری در طرح‌های تامین مالی جمعی با استفاده از شبکه‌های عصبی، کنفرانس بین المللی ابزار و تکنیک‌های مدیریت، ۵ و ۶ اسفند، ۱۳۹۳، تهران.

9. . A., Agrawal, C., Catalini, A., Goldfarb, Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions, *Journal of Economics & Management Strategy*, Volume 24, Number 2, Summer 2015, 253–274.

10. H., Akkermans, J., Vennix, Clients' Opinions on Group Model-Building: An Exploratory Study, The 14th International Conference of the System Dynamics Society, 1996 Cambridge, Massachusetts, USA.

11. T. H., Allison, et al. Crowdfunding in a Prosocial Microlending Environment: Examining the Role of Intrinsic Versus Extrinsic Cues, Entrepreneurship Theory and Practice, 39(1), 53–73.

12. N., Barasinska, D., Schafer, Is Crowdfunding Different? Evidence on the Relation between Gender and Funding Success from a German Peer-to-Peer Lending Platform, German Economic Review, 15(4): 436–452.

13. P., Belleflamme, et al., "Crowdfunding: Tapping the right crowd." Journal of Business Venturing 29(5): 585-609.

14. P., Belleflamme, N., Omrani, M., Peitz, The economics of crowdfunding platforms, Information Economics and Policy, 33 (2015) 11–28.

15. B., Broers, The Characteristics of a Fast-Growing Start-Up. Nijmegen school of management master thesis in business administration organization design & development, November, 2016.

16. M., Cholakova, B., Clarysse, Does the possibility to make equity investments in crowdfunding projects crowd out reward-based investments?, Entrepreneurship Theory and Practice, 39(1), 145–172.

17. L., Corazzini, C., Cotton, P., Valbonesi, Donor coordination in project funding: Evidence from a threshold public goods experiment, Journal of Public Economics 128 (2015) 16–29.

18. A., Cosh, D., Cumming, A., Hughes, Outside entrepreneurial capital. Econ. J. 119, 1494–1533.

19. J., Forrester, (1980), Information Sources for Modeling the National Economy, *Journal of the American Statistical Association*, 75 (371): 555-574.
20. N., Ibrahim, J., Verliyantina, The Model of Crowdfunding to Support Small and Micro Businesses in Indonesia Through a Web-based Platform, Procedia Economics and Finance 4(0): 390-397.
21. T., Lambert A., Schwienbacher, An empirical analysis of crowdfunding. Available at: http://www.crosnerlegal.com/images/47770544_An_Empirical_Analysis_of_Crowdfunding.pdf.
22. S. A., Macht, J., Weatherston, The Benefits of Online Crowdfunding for Fund-Seeking Business Ventures, Strategic Change, 23: 1–14 (2014), Published online in Wiley Online Library.
23. Massolution, Crowdsourcing and Crowdfunding …A Global Industry Perspective, Digital Malaysia: National Crowdsourcing Conference, 10 June 2013, Kuala Lumpur, Malaysia.
24. E., Mollick, "The dynamics of crowdfunding: An exploratory study." *Journal of Business Venturing* 29(1): 1-16.
25. M. J., Renwick, E., Mossialos, Crowdfunding our health: Economic risks and benefits. Social Science & Medicine, Volume 191, October 2017, Pages 48-56.
26. G. P., Richardson, J., Vennix, D. M. Anderson, J., Rohrbaugh. W.A. Wallace, Eliciting Group Knowledge for Model-Building, The 7th International Conference of the System Dynamics Society, 1989 Stuttgart, Germany.
27. E., Roos, Group Model Building with a Client Using System Dynamics Modeling, The 15th International Conference of the System Dynamics Society, August 19-22, 1997, Istanbul, Turkey.
28. L., Valančienė, S. Jegelevičiūtė "Crowdfunding for Creating Value: Stakeholder Approach." Procedia - Social and Behavioral Sciences 156(0): 599-604.
29. J., Vennix, Group Model Building: Facilitating Team Learning Using System Dynamics. London: John Wiley & Sons, 1 edition (August 1996).
30. J., Vennix, J. W., Gubbels, D., Post, H. J., Poppen, A Structured Approach to Knowledge Acquisition in Model Development, The 6th International Conference of the System Dynamics Society, 1988 La Jolla, CA USA.
31. X., Walthoff-Borm, T., Vanacker, V., Collewaert, Equity crowdfunding, shareholder structures, and firm performance. Corp Govern Int Rev. 2018; 26: 314– 330. https://doi.org/10.1111/corg.12259.
32. B., Xu, H.,, Zheng, Y., Xu, T., Wang, Configurational paths to sponsor satisfaction in crowdfunding, Journal of Business Research, 69 (2016) 915–927.
33. H., Zheng, et al. "The role of multidimensional social capital in crowdfunding: A comparative study in China and US." Information & Management 51(4): 488-496.

فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷
صص: ۱۷ - ۳۴

ارائه متدولوژی انتخاب استراتژی مدیریت پروژه‌های فناوری اطلاعات

[*] مونا جامی‌پور [**] سید محمدباقر جعفری [***] مرضیه نصراللهی
* استادیار گروه مدیریت، دانشگاه حضرت معصومه (س)، قم، ایران
** استادیار گروه مدیریت صنعتی و مالی، دانشکده مدیریت و حسابداری، پردیس فارابی دانشگاه تهران، قم، ایران
*** کارشناس ارشد، مدیریت اجرایی، دانشکده مدیریت و حسابداری، پردیس فارابی دانشگاه تهران، قم، ایران
تاریخ دریافت: ۱۳۹۸/۰۳/ ۲۱ تاریخ پذیرش: ۱۳۹۹/۰۳/۱۰

چکیده

رویکرد استراتژیک به مدیریت پروژه، نگرش جدیدی است که در دهه اخیر، بیش از پیش، مورد توجه اندیشمندان کسب وکار قرار گرفته است. داشتن رویکرد استراتژیک در حوزه پروژه‌های فناوری اطلاعات، مدیران را ملزم می‌سازد تا در گام نخست مدیریت پروژه‌ها، به انتخاب استراتژی مناسب با درنظر گرفتن عوامل متعدد تأثیرگذار بر انتخاب استراتژی مبادرت ورزند. رویکردهای منطقی و تطبیقی دو استراتژی غالب در مدیریت پروژه می‌باشند که هریک ابزارها، منابع و رویه‌های متفاوتی را می طلبند. علیرغم افزایش سرمایه‌گذاری‌ها در پروژه‌های فناوری اطلاعات و وابستگی سازمان‌ها به چنین فناوری‌هایی، هنوز مطالعه‌ای که به مدیران فناوری اطلاعات در انتخاب استراتژی مناسب برای چنین پروژه‌هایی یاری رساند صورت نپذیرفته است. لذا، هدف اصلی پژوهش حاضر، شناسایی رویکرد انتخاب استراتژی مناسب در مدیریت پروژه‌های فناوری اطلاعات می‌باشد. به منظور دستیابی به این هدف در مرحله اول به منظور شناسایی عوامل تأثیرگذار بر انتخاب استراتژی مدیریت پروژه‌های فناوری اطلاعات پس از مرور جامع ادبیات در حوزه مربوطه به بهره‌مندی از نظرات خبرگان در گروه کانونی پرداخته شده است. در مرحله دوم، به منظور اعتبارسنجی عوامل استخراج شده، و در مرحله سوم، به منظور وزن‌دهی هریک از عوامل در انتخاب دو رویکرد غالب منطقی و تطبیقی، به نظرسنجی از خبرگان با استفاده از روش پیمایش پرداخته شده است. در متدولوژی توسعه‌یافته عواملی که بر انتخاب استراتژی تأثیر گذارند عبارتند از: عوامل فرایندی، عوامل انسانی، عوامل مرتبط با سیستم و عوامل زمینه‌ای. استراتژی نهایی مدیریت پروژه‌های فناوری اطلاعات به صورت طیفی از استراتژی منطقی و استراتژی تطبیقی بدست می‌آید.

واژه‌های کلیدی: مدیریت پروژه، پروژه فناوری اطلاعات، استراتژی مدیریت پروژه، رویکرد منطقی، رویکرد تطبیقی

۱. مقدمه

امروزه بهره‌گیری از فناوری اطلاعات در حوزه‌های مختلف فعالیت‌های سازمانی، بسیار فراگیر شده و به جزئی از فرایندهای اصلی و اساسی سازمان‌ها تبدیل شده است. علی‌رغم مزایای فراوان به‌کارگیری فناوری اطلاعات و دهه‌ها تلاش برای بهبود مدیریت پروژه‌های فناوری اطلاعات، نرخ موفقیت این پروژه‌ها همچنان کم است [۲]. تحقیقات بسیاری هم در زمینه‌ی علل و عوامل شکست پروژه‌ها انجام شده است و تأکید بر چشم‌انداز مدیریت پروژه و افزایش اثربخشی، بهره‌وری و گزارش‌دهی افزایش‌یافته است [۱۷]؛ باوجود این، آمارها همچنان

نویسندهٔ عهده‌دار مکاتبات: مونا جامی پور monajami@ut.ac.ir

حکایت از شکست اغلب پروژه‌ها دارد [۲۳]. این در حالی ست که مزایای استفاده از کاربردهای فناوری اطلاعات تنها در صورت پیاده‌سازی موفقیت‌آمیز پروژه‌های فناوری اطلاعات انجام می‌شود [۵]. به منظور کاهش و کنترل چالش‌ها در پروژه‌ها، بایستی رویکرد مدیریت پروژه از مدیریت سنتی به مدیریت استراتژیک تغییر یابد تا پاسخگوی بهینه‌سازی عملکرد و موفقیت پروژه باشد. مدیریت استراتژیک پروژه، مجموعه‌ای از شیوه‌ها، رویه‌ها، فرایندها، ابزارها و رفتارهایی است که به‌طور کلی مشخص می‌کند تا چه حد یک سازمان پیوند مؤثری بین شیوه‌های مدیریت پروژه و مدیریت کسب‌وکار ایجاد کرده است [۲۷]. مدیریت استراتژیک پروژه، بر ایجاد مزیت رقابتی برای سازمان تمرکز می‌کند [۴۱]؛ در نتیجه مدیران پروژه به منظور دستیابی به مزیت رقابتی و عملکرد بهتر باید یک چارچوب برنامه‌ریزی و اجرایی مطابق با رویکرد راهبردی داشته باشند [۳۱].

مدیریت استراتژیک پروژه بر بهترین شکل استفاده از منابع و هماهنگی این منابع برای دستیابی به چشم‌انداز و اهداف سازمان پروژه محور، تأکید دارد [۵۰]. در رویکرد استراتژیک پروژه، سبک مدیریت پروژه با شرایط خاص هر پروژه مطابقت دارد و پروژه‌ها با رویکرد مدیریت انطباقی اداره می‌شوند [۴۱]؛ بنابراین مدیریت استراتژیک پروژه برای موفقیت آن بسیار حیاتی است [۴۴]. ویژگی‌های برنامه‌ریزی استراتژیک می‌تواند به‌طور مؤثر در چارچوب مدیریت پروژه گنجانیده شود و بینش مفیدی در رابطه با مدیریت پروژه و موفقیت احتمالی آن به دست آید (پاپک شیلد و بایررایت، ۲۰۱۷).

اخیراً پشتیبانی از رویکرد استراتژی «منطقی- تطبیقی» در برنامه‌ریزی مدیریت پروژه دیده شده است. یافته‌ها نشان می‌دهد که مدیریت پروژه توسط درجات مختلفی از رویکردهای منطقی- تطبیقی گرفته شده که با موفقیت آن و استفاده از ابزار/ تکنیک‌های مدیریت پروژه ارتباط مثبت دارد. با توجه به تفاوت‌ها در چارچوب، می‌توان گفت که ادبیات مدیریت پروژه دارای تشابه قابل‌توجه با تحقیقات برنامه‌ریزی استراتژیک است. فرایند مدیریت پروژه، انجام آن از طریق برنامه‌ریزی و اجرا و ارتباط با موفقیت پروژه یک تمرکز مستمر تحقیقات مدیریت پروژه است [۳۸]. علی‌رغم اهمیت مباحث استراتژیک در مدیریت پروژه در حوزه مدیریت پروژه‌های فناوری اطلاعات، بررسی‌ها در مقالات و کتب مختلف نشان می‌دهد که این مباحث مورد غفلت واقع شده است. لذا هدف از انجام این پژوهش ارائه متدولوژی انتخاب استراتژی مناسب در مدیریت پروژه‌های فناوری اطلاعات می‌باشد. سؤال اصلی پژوهش آن است که متدولوژی مناسب ایجاد استراتژی در مدیریت پروژه‌های فناوری اطلاعات چیست؟ سؤالات فرعی پژوهش عبارتند از: عوامل مؤثر بر انتخاب رویکردهای مناسب ایجاد استراتژی در مدیریت پروژه‌های فناوری اطلاعات در یک سازمان کدامند؟ وزن هر یک از این عوامل در هر یک از رویکردها چقدر است؟

۲- مروری بر ادبیات پژوهش

۲-۱- پروژه‌های فناوری اطلاعات

پروژه‌های فناوری اطلاعات را می‌توان در زمره مهم‌ترین سرمایه‌گذاری سازمانی در نظر گرفت [۳۰] و [۲۹]. هدف از سرمایه‌گذاری در فناوری اطلاعات، بهبود بهره‌وری عملیاتی یک سازمان است تا هزینه‌ها کاهش یافته و سطح سود افزایش یابد [۳۰]. بر طبق گزارش موسسه گارتنر (۲۰۱۹) میزان سرمایه گذاری در حوزه فناوری اطلاعات در سال ۲۰۱۹ نسبت به سال قبل ۰,۴٪ رشد داشته است. بودجه مربوط به فناوری اطلاعات در سال ۲۰۲۰ به ۳,۴ تریلیون دلار خواهد رسید و پیش بینی شده است در سال ۲۰۲۱ حدود ۳,۸ ٪ رشد خواهد داشت. با این وجود، یک آمار چشمگیری از این پروژه ها به شکست می‌انجامند و میلیاردها دلار اتلاف منابع در اثر شکست چنین پروژه‌هایی برجای خواهد ماند. مدیوم (۲۰۱۹) پروژه های فناوری اطلاعات را پروژه های بدنام به دشواری مدیریت کردن آنها بیان می کند و به استناد به مجله کسب و کار هارواد بیان می دارد که ۲۷٪ از پروژه های فناوری اطلاعات بیش از بودجه هایشان هزینه داشته اند (تاوریزین، ۲۰۱۹). حداقل یکی از شش پروژه فناوری اطلاعات هزینه اش بیش از ۲۰۰٪ و برنامه زمانبندی پروژه بیش از ۷۰٪ از تخمین‌های اولیه خواهد شد. باید بیان کرد اگرچه فناوری اطلاعات باعث سرعت بخشیدن، افزایش قابلیت اطمینان و هزینه پایین تر انجام تعاملات کسب و کارها می‌گردد ولی پروژه های فناوری اطلاعات هزینه، ریسک و پیچیدگی بالایی دارند. پروژه‌های فناوری اطلاعات معمولاً پروژه‌های منحصر به‌فردی برای یک سازمان هستند و به‌صورت روتین انجام

نمی‌شوند [۲۴]. این به پیچیدگی پروژه‌های فناوری اطلاعات می‌افزاید. علاوه بر این مدیریت پروژه فناوری اطلاعات دارای گستره وسیعی است، زیرا شامل دو موضوع مجزا، یعنی مدیریت فناوری اطلاعات و مدیریت پروژه است [۵] و مجموعه‌ای گسترده از فعالیت‌های سازمانی را پوشش می‌دهد؛ از نگهداری سیستم‌های قدیمی گرفته تا ایجاد ایده‌های خلاقانه که از طرح‌های نو نظیر چاپ سه‌بعدی و یا رایانش ابری پشتیبانی می‌کند [۲۹]. بروکس (۲۰۱۵) نیز تفاوت پروژه‌های حوزه فناوری اطلاعات و دیگر پروژه‌ها را این‌چنین عنوان می‌کند:

نامرئی بودن: پیشرفت ساختارهای فیزیکی مثل پل یا جاده می‌تواند دیده شود، در حالی‌که پیشرفت پروژه‌های نرم‌افزاری بلافاصله قابل مشاهده نیست.

پیچیدگی: محصولات نرم‌افزاری پیچیده‌تر از سایر مصنوعات مهندسی است.

انعطاف پذیری: سهولتی که در تغییر نرم افزار وجود دارد، معمولاً به عنوان یکی از نقاط قوت آن دیده می‌شود.

بنابراین در زمان مدیریت این‌گونه پروژه‌ها به یک مجموعه قوی فنی و غیر فنی به همراه مهارت‌های متنوعی نیاز است تا بتوان پروژه‌های فناوری اطلاعات را با موفقیت به سرانجام رساند [۲۹].

زمانی که یک سازمان محصول، سرویس و یا راه‌حل‌های جدید مبتنی بر فناوری اطلاعات را ایجاد و یا پیاده‌سازی می‌نماید، متعهد به صرف زمان، پول و سایر منابع مورد نیاز برای پروژه است و بدیهی است که این انتظار را داشته باشد که بتواند پس از خاتمه پروژه، ارزشی را برای سازمان به دنبال داشته باشد (اسچوالب، ۲۰۱۵). این وظیفه مدیر و تیم پروژه است که آن ارزش را برای سازمان ارائه نماید [۲۹]. در طول این سال‌ها مشخص‌شده است که مدیریت پروژه ابزاری مؤثر برای مقابله با وظایف پیچیده‌ای مانند درگیر شدن در پروژه‌های فناوری اطلاعات است، به‌طوری که در سال‌های اخیر به‌سرعت تکامل یافته است [۵۱]. ساختار انعطاف‌پذیر سازمان‌های پروژه محور به آن‌ها اجازه می‌دهد تا با اجرای پروژه‌های منطبق بر برنامه‌ریزی و اهداف استراتژیک کارایی بالاتری را تجربه نمایند؛ بنابراین می‌توان گفت امروزه یکی از مهم‌ترین مزیت‌های رقابتی سازمان‌ها موفقیت در پروژه‌ها است [۲].

۲-۲- استراتژی مدیریت پروژه

در محیط نظامی، استراتژی به معنای چگونگی برنامه‌ریزی برای پیروزی است. همین اصل باید در پروژه‌ها اعمال شود؛ بنابراین، استراتژی پروژه در مورد برنده شدن است. برنده شدن در نبرد بازار، با محصول یا خدمات خاص تولید شده توسط استراتژی خوب پروژه اتفاق می‌افتد؛ محصولات یا خدماتی که مزیت رقابتی ایجاد می‌کند. ازاین‌رو، استراتژی پروژه راه خاصی است که در محیط رقابتی امروز، برنده شدن پروژه را در جنگ تجاری موجب می‌شود.

در جدول (۱) تفاوت مدیریت کلاسیک و مدیریت استراتژیک پروژه نشان داده شده است.

جدول ۱ : تفاوت مدیریت پروژه و مدیریت استراتژیک پروژه

	مدیریت پروژه (دیدگاه سنتی)	مدیریت استراتژیک پروژه (دیدگاه نوین)
پارادایم اساسی	پروژه‌ها مجموعه فعالیت‌هایی هستند که باید مطابق زمان، بودجه و الزامات اجرا شوند.	پروژه‌ها فرایندهای راهبردی سازمانی هستند که برای دستیابی به اهداف کسب‌وکار انجام می‌شوند.
کانون توجه	کارایی	اثربخشی و کارایی
دیدگاه	عملیاتی	راهبردی و عملیاتی
نقش مدیر پروژه	انجام کار مطابق زمان، بودجه و با توجه به مشخصات	دستیابی به نتایج کسب‌وکار و برد در بازار
سبک مدیر پروژه	یک رویکرد برای تمام پروژه‌ها مناسب است	رویکرد تعدیلی با توجه به نوع پروژه
تعریف پروژه	محدوده پروژه، چه چیزی باید انجام شود؟	محصول، مزیت رقابتی، راهبرد، محدوده
برنامه‌ریزی	فعالیت‌ها، زمان‌بندی، بودجه	نتایج نهایی، ابعاد موفقیت، فعالیت‌ها

	مدیریت پروژه (دیدگاه سنتی)	مدیریت استراتژیک پروژه (دیدگاه نوین)
بازنگری پروژه	پیشرفت، وضعیت کنونی، نقاط عطف پروژه (مایلستون‌ها) ، بودجه	نیازهای مشتری، راهبرد، ابعاد موفقیت، وضعیت کنونی
جنبه انسانی	تیم‌ها، حل تعارض	معنا بخشی، انگیزش

برگرفته از مقاله (Petrović, 2012)

استراتژی پروژه جهت گیری پروژه می‌باشد که در موفقیت آن در محیط پروژه نقش اساسی دارد [۶]. استراتژی پروژه اشاره به استراتژی یک پروژه منفرد می‌باشد که بیشتر بر برنامه و اهداف پروژه تأکید دارد [۵۶]. استراتژی پروژه به‌عنوان چشم‌انداز پروژه، موقعیت و دستورالعمل‌ها با هدف دستیابی به بالاترین مزیت رقابتی و بهترین ارزش در پروژه است.

اگرچه رویکرد مدیریت استراتژیک پروژه محبوبیت فراوانی یافته است، ولی هنوز به یک رویکرد صریح و گسترده‌ی مورد استفاده در اجرای پروژه‌ها تبدیل نشده است [۳۹] و یکی از دلایل اصلی شکست پروژه ها، استراتژی‌های ضعیف آن‌ها می‌باشد [۵۶].

مکاتب رویکرد استراتژیک به مدیریت پروژه به سه دسته تقسیم می‌شوند: ۱- ماهیت تجویزی (طراحی، برنامه‌ریزی و موقعیت‌یابی)؛ ۲- ماهیت تطبیقی (کارآفرینی، شناخت، یادگیری، قدرت، فرهنگ و محیط)؛ ۳- ماهیت ترکیبی که به دنبال ادغام رویکردهای تجویزی و تطبیقی است [۷]. به اعتقاد پاپک‌شیلد و بایررایت (۲۰۱۷) پروژه‌ها اغلب به عنوان بخشی از فرایند برنامه‌ریزی استراتژیک سازمان‌ها آغاز می‌شوند. استفاده از رویکرد استراتژیک در برنامه‌ریزی و اجرای پروژه‌ها با موفقیت پروژه ارتباط مثبتی دارد و از آنجا که یکی از بزرگ‌ترین عوامل شکست پروژه‌های فناوری اطلاعات عدم پیاده‌سازی جامع استراتژی و برنامه‌ریزی است [۵]، رویکرد استراتژیک در مدیریت پروژه باید در شروع پروژه مورد بحث و تصمیم‌گیری قرار گیرد. احتمال موفقیت پروژه را می‌توان با انتخاب آگاهانه یک رویکرد مدیریت پروژه در راه‌اندازی پروژه جدید افزایش داد. بسته به رویکرد انتخاب‌شده مدیریت پروژه (تجویزی یا تطبیقی)، عوامل موفقیت متفاوتی وجود دارد که باید توسط سازمان مورد توجه قرار گیرد و موفقیت نیز بستگی به عوامل زیادی دارد که ممکن است از پروژه‌ای به پروژه دیگر و از سازمانی به سازمان دیگر تغییر کند [۴۵].

انتخاب استراتژی در ابتکارات سازمانی و پروژه‌های متفاوت از اهمیت بسزایی برخوردار می‌باشد که تحقیقات متعددی نیز در حوزه‌های مختلف در رابطه با متدولوژی-های انتخاب استراتژی صورت گرفته است. جدول (۲) پژوهش‌های صورت گرفته در زمینه ارائه متدولوژی انتخاب استراتژی در حوزه‌های مختلف را نشان می‌دهد.

جدول (۲): پژوهش های صورت گرفته در حوزه متدولوژی انتخاب استراتژی

محققین (سال) / هدف	انواع استراتژی مورد بررسی	عوامل تأثیرگذار بر استراتژی مورد نظر
کاپالدو و ریپا (۲۰۰۹) ارائه متدولوژی انتخاب استراتژی پیاده‌سازی سیستم برنامه ریزی منابع شرکت	پیاده سازی محلی، بیگ بنگ، راهکارهای مبتنی بر سیستم های موروثی، راهکارهای برنامه‌ریزی منابع شرکت جزئی، پیاده سازی تدریجی	استانداردسازی، شخصی سازی، گرایش مهندسی مجدد فرایند، کاربرپسندبودن
زایم و همکاران (۲۰۱۲) رویکرد انتخاب استراتژی نگهداری	نگهداری قابل پیش بینی، نگهداری دوره‌ای، نگهداری اصلاحی	پیاده سازی، امنیت، هزینه، ارزش افزایی
چن و همکاران (۲۰۱۲) ارائه رویکرد انتخاب استراتژی تجاری مدیریت زنجیره تأمین سبز	استراتژی مبتنی بر ریسک، استراتژی مبتنی بر کارایی، استراتژی مبتنی بر نوآوری، استراتژی حلقه بسته	طراحی سبز، تولید سبز، خرید سبز، بازاریابی و خدمات سبز
پاپک‌شیلد و بایررایت (۲۰۱۷) شناسایی ویژگی‌های استراتژیک مدیریت پروژه	رویکرد تطبیقی، رویکرد منطقی	رسمیت، جامعیت، مشارکت، شدت

محمدی مقدم و همکاران (۱۳۹۴) ارائه متدولوژی استراتژی مدیریت دانش بر اساس مارپیچ دانش سازمان	اجتماعی کردن، بیرونی کردن، درونی کردن، ترکیب	ساختار سازمانی، فرهنگ سازمانی، فناوری اطلاعات، راهبردهای منابع انسانی و فرآیندهای تسهیم، حفظ، کسب و خلق دانش
شعبان الهی و همکاران (۱۳۸۹) ارائه یک متدولوژی برای ایجاد استراتژی مدیریت دانش مطالعه و بررسی سه سازمان نمونه	سیستم گرا، انسان گرا	استراتژی عمومی کسب و کار سازمان، ساختار سازمانی، عوامل فرهنگی و فرایندهای خلق و انتشار دانش در ناحیه دانش بر سازمان

همانطور که در جدول (۲) نشان داده شده است پژوهشی تاکنون در حوزه انتخاب استراتژی پروژه‌های فناوری اطلاعات صورت نگرفته است تنها پژوهش صورت گرفته در حوزه کلی مدیریت پروژه مربوط به پژوهش پاپک‌شیلد و بایررایت (۲۰۱۷) می‌باشد که با توجه به ویژگی‌های استراتژیک مدیریت پروژه به انتخاب استراتژی مناسب پرداخته است. بنابراین نیاز به توسعه متدولوژی انتخاب استراتژی پروژه‌های فناوری اطلاعات کاملاً مشهود می‌باشد. همانطور که با مطالعه پژوهش‌های مرتبط با انتخاب استراتژی آشکار می‌گردد نیاز است تا عوامل مؤثر بر انتخاب استراتژی شناسایی گردد که مراحل پژوهش با الهام از پژوهش‌های حوزه انتخاب استراتژی در بخش روش شناسی بیان شده است.

۲-۲- استراتژی‌های مدیریت پروژه‌های فناوری اطلاعات

بعضی از مدیران پروژه در محقق سازی استراتژی اولیه‌ی پروژه‌های فناوری اطلاعات شکست می‌خورند درحالی‌که مهم‌ترین فاز در هر پروژه فناوری اطلاعات برنامه‌ریزی استراتژی پروژه و پیاده‌سازی آن است [۵]. بورگان و بورگان (۲۰۱۴) دو رویکرد قابل پیش بینی (برنامه محور) و تطبیقی (تغییر محور) را استراتژی‌های غالب در مدیریت پروژه بیان می‌کند که رویکرد اول در محیطی که عدم اطمینان پایین است مناسب می‌باشد و تمام شرایط در کنترل مدیر پروژه خواهد بود. در رویکرد دوم، چرخه حیات پروژه در شرایطی که نیازمند تعامل ذینفعان و کاربران می‌باشد مناسب خواهد بود.

پروژه‌های فناوری اطلاعات باید با در نظر گرفتن رویکردهای توسعه نرم‌افزاری برنامه‌ریزی شوند [۱۱]. استراتژی‌های پروژه‌های فناوری اطلاعات شامل تنظیم افق برنامه‌ریزی بلندمدت برای سرمایه‌گذاری در زمینه‌ی منابع انسانی، تخصص، فناوری، سخت‌افزار و نرم‌افزار مورد نیاز برای بهره‌برداری از فرصت‌هایی ست که ممکن است در جریان فرایند مدیریت پروژه فناوری اطلاعات به‌وجود آید [۷]. واضح است که برنامه‌نویسی و مدیریت پروژه‌های نرم‌افزاری بدون استفاده از یک رویکرد، بسیار سخت و حتی غیرممکن است [۵۵]. به همین منظور، سیستم‌های نرم‌افزاری شامل رویکردهای «ساخت‌یافته و چابک» هستند [۲۵]. مارچیوکا و همکاران (۲۰۱۶) دو رویکرد رایج ساخت‌یافته (تجویزی) و چابک (تطبیقی) را استراتژی‌های رایج در مدیریت پروژه‌های فناوری اطلاعات بیان می‌کند. رویکردهای ساخت‌یافته رویکردی منطقی و سیستماتیک در پروژه‌های فناوری اطلاعات دارا می‌باشند و رویکرد چابک رویکرد پذیرای تغییرات حین پروژه‌های فناوری اطلاعات می‌باشد تا محصولی بهتر را به کاربران ارائه دهد. هر یک از این روش‌ها از مجموعه‌ای از ابزارها، مدل‌ها و مفاهیم خاص استفاده می‌کند؛ بنابراین آن‌ها در برخی زمینه‌های خاص موفق و قابل اجرا هستند و انتخاب روش مناسب در یک زمینه و سیستم خاص، یک فرایند بسیار مهم و دشوار است [۵۵].

در روش‌های ساختاریافته که هم‌اکنون نیز در توسعه سیستم‌ها استفاده می‌شوند بخش بزرگی از فعالیت‌های فرایند توسعه از ابتدای پروژه برنامه‌ریزی می‌شود. این روش‌ها با عناوین سنتی یا سنگین‌وزن نامیده می‌شوند. این متدولوژی نیاز به یک فرایند مشخص برای توسعه سیستم‌ها، بر اساس مستندات جامع دارد تا این فعالیت را قابل پیش‌بینی‌تر و کارآمدتر کند. بخش بزرگی از فرایند نرم‌افزاری برای مدت زمان طولانی برنامه‌ریزی شده است. روش‌های سنتی را می‌توان تنها برای سیستم‌های توسعه نرم‌افزاری که نیازهای آن از ابتدای طرح مشخص‌شده است، با موفقیت مورد استفاده قرار داد [۲۵]. فرایندهای ساخت‌یافته با توجه به عملیات به‌طور دقیق تعریف‌شده، مستندات گسترده و برنامه‌ریزی و مدیریت دقیق مشخص می‌شوند. انتقاد اصلی در برابر روش‌های ساخت‌یافته، این است که آن‌ها منجر به افزایش سطح مستندات و درنتیجه بوروکراسی می‌شوند [۱۱].

اخیراً، موج فرایندهای توسعه چابک برای غلبه بر بعضی از محدودیت‌ها و جنبه‌های نامطلوب فرایندهای ساخت‌یافته ارائه شده است. فرایندهای مرسوم در توسعه چابک، بر اهمیت برقراری ارتباط غیررسمی مؤثر میان توسعه‌دهندگان و تکرار بهبود پیاده‌سازی‌ها به‌وسیله سناریوهای مورد استفاده و تیم‌های توسعه تأکید دارد [۲۲]. استفاده از روش چابک به‌عنوان رویکرد مدیریت پروژه‌ها در طول چندین سال گذشته به‌طور چشمگیری افزایش یافته است. طبق گزارش سی آی او [۴۹]حدود ۷۱ درصد از سازمان‌ها از رویکرد چابک در مدیریت پروژه‌های فناوری اطلاعات استفاده کرده‌اند. ترونگ و جیتبیپون (۲۰۱۶) بیان می‌کنند که رویکرد چابک به پروژه‌های فناوری اطلاعات استراتژی رو به رشدی می‌باشد که می‌تواند به موفقیت پروژه‌های فناوری اطلاعات بینجامد.

در ادامه، پیشینه‌ی تجربی پژوهش حاضر مطابق جدول (۳) ارائه می‌شود.

جدول ۳ : خلاصه‌ی تحقیقات انجام شده در حوزه عوامل مؤثر بر رویکرد/ استراتژی مدیریت پروژه

محققان (سال) و عنوان پژوهش	عوامل مؤثر
گیمباسو و همکاران (۲۰۱۱)	عواملی که بر تصمیم‌گیری در انتخاب روش‌های توسعه نرم‌افزار عنوان شده‌اند عبارتند از: تخمین دقیق هزینه‌های اولیه و زمان توسعه، هزینه‌های توسعه، مدت‌زمان تحویل سیستم نهایی، پیچیدگی سیستم، اندازه تیم توسعه.
استلر و همکاران (۲۰۱۴)	اهمیت پروژه، انگیزه تیم، صرفه جویی، ارتباطات بلادرنگ، ارتباطات غیرهمزمان، تفاوت فرهنگی، مدیریت پروژه، فقدان دانش فنی، مهارت کارگر، خواندن و نوشتن مستندات، تضاد کارکنان، برنامه زمانبندی پروژه، مالکیت معنوی
ستزینگر و همکاران (۲۰۱۱)	میزان مشخص بودن نیازمندی ها و الزامات سیستم، ریسک فنی
والاسیچ و همکاران (۲۰۱۵)	اندازه، کلیدی بودن پروژه، پویایی، پرسنل و فرهنگ
وینکار و همکاران (۲۰۰۶)	اندازه، پویایی بودن و کلیدی بودن پروژه
هارب و همکاران (۲۰۱۵)	سبک توسعه، مشارکت مشتری، سطح مستندسازی، دوره زمان اصلاحات، فرهنگ، ساختار و شیوه‌های مدیریت، گارانتی موفقیت و زمان اجرا.
کرام و مارابلی (۲۰۱۸)	میزان سازگاری نیازمندی ها، ترکیب تیم، اعتماد بین اعضا، قابلیت های مدیریت، یادگیری مستمر، میزان استاندارد بودن و رسمی بودن آموزش، میزان تمرکز بر نوع دانش خروجی پروژه

با بررسی ادبیات حوزه مدیریت پروژه فناوری اطلاعات می‌توان انتقادات زیر را مطرح کرد: اول، در مدیریت پروژه، رویکرد استراتژیک اهمیت زیادی دارد و علی‌رغم اهمیت آن، در مطالعات مدیریت پروژه پژوهش‌های محدودی به آن پرداخته‌اند [۳۸]؛ [۴۵]؛ [۵۰]؛ [۴۱]، در پروژه‌های فناوری اطلاعات مورد غفلت واقع شده و با توجه به اهمیت متفاوت پروژه‌های فناوری اطلاعات نیاز به دید استراتژیک نسبت به این پروژه‌ها وجود دارد. دوم، عمده مطالعات به بیان انواع رویکردها و استراتژی‌های مدیریت پروژه پرداخته‌اند که اکثریت آنها دو رویکرد تطبیقی (چابک) و منطقی (ساخت یافته) را استراتژی غالب مدیریت پروژه های فناوری اطلاعات بیان کرده‌اند ولی چگونگی انتخاب هریک از استراتژی‌ها و میزان تمرکز بر آن‌ها حلقه مفقوده در ادبیات مدیریت پروژه فناوری اطلاعات می‌باشد. سوم، در برخی از مطالعات به عوامل مؤثر بر انتخاب استراتژی‌ها اشاره شده است ولی در رابطه با میزان تأثیر آن‌ها بر انتخاب دو رویکرد رایج بحثی در ادبیات توسط محققان یافت نشده است.

۳- روش‌شناسی پژوهش

هدف اصلی این پژوهش توسعه متدولوژی مناسب برای انتخاب استراتژی مدیریت پروژه های فناوری اطلاعات است که می توان آن را از نظر هدف در زمره تحقیقات توسعه ای-کاربردی و از نظر روش و ابزار گردآوری داده‌ها

در حوزه تحقیقات توصیفی دانست. دستیابی به هدف پژوهش در سه مرحله زیر صورت گرفته است.

مرحله اول: در ابتدا به استخراج عوامل تأثیرگذار بر انتخاب استراتژی مدیریت پروژه فناوری اطلاعات از طریق مرور جامع ادبیات در حوزه استراتژی مدیریت پروژه، مدیریت پروژه فناوری اطلاعات و رویکردهای غالب مدیریت پروژه فناوری اطلاعات پرداخته شده است. سپس از گروه کانونی به منظور اغنای نظرات و طبقه‌بندی آن‌ها بهره گرفته شد. گروه کانونی برای پاسخ به سؤالات در مصاحبه‌های عمیق در بستر اجتماعی است. این تکنیک به درک تجارب مشارکت کنندگان، درک بهتر معانی و چرایی نگرش‌های آنها در رابطه با مسئله تحقیق می-پردازد. یکی از اهداف گروه کانونی درک نگرش‌ها و عقاید مشارکت‌کنندگان در رابطه با موضوع مورد بررسی می-باشد [۳۴]؛ [۲۰]. به نظر می‌رسد که نوعی توافق بر این امر وجود دارد که تعداد اعضای هر گروه نباید متجاوز از ده تا دوازده نفر شرکت‌کننده باشد [۳۷] و از طرف دیگر اندازه ایده‌آل گروه نیز، تعدادی بین پنج تا هفت نفر است [۸]. گروه کانونی پژوهش حاضر شامل ۵ نفر از خبرگان حوزه فناوری اطلاعات، مدیریت پروژه و برنامه‌ریزی استراتژیک بوده است. بعد از شناسایی افراد شرکت کننده، به صورت رسمی از آن‌ها دعوت به‌عمل آمد تا در مباحث گروه شرکت کنند. در این دعوت به صورت کلی مواردی برای آن‌ها بیان گردید که عبارتند از: هدف از بحث‌های گروه کانونی چیست؟ این بحث‌ها در کجا و با چه کسانی صورت می‌گیرد و محققان چه کسانی هستند و به دنبال ارائه چه چیزی هستند؟ گفتگوها ضبط و مستندسازی‌ها جهت رسیدن به توافق برای خروجی‌ها صورت گرفت و با ارائه یافته‌ها به گروه مشارکت‌کننده از درستی تحلیل‌ها اطمینان حاصل شد.

مرحله دوم: در این مرحله با استفاده از روش پیمایش به نظرسنجی از خبرگان در رابطه با عوامل تأثیرگذار بر انتخاب استراتژی مدیریت پروژه‌های فناوری اطلاعات پرداخته شده است. پرسشنامه توسعه یافته برمبنای یافته‌های حاصل از مرحله اول پژوهش بوده است که از خبرگان خواسته شده تا نظر خود را در رابطه با هریک از عوامل و شاخص‌های مربوطه بیان کنند. پرسشنامه اول پژوهش، حاوی بخش‌های اطلاعات جمعیت شناختی، راهنمای پرسشنامه و سؤالات مربوط به عوامل تأثیرگذار بر انتخاب استراتژی مدیریت پروژه فناوری اطلاعات بوده است. در این پرسشنامه از خبرگان خواسته شده بود تا نظراتشان را در رابطه با اینکه "آیا شاخص‌های ارائه‌شده می‌توانند به عنوان عوامل تأثیرگذار بر انتخاب استراتژی مدیریت پروژه فناوری اطلاعات درنظر گرفته شوند؟" بیان کنند. برای این منظور از طیف لیکرت ۵ گزینه‌ای استفاده شد. داده‌های حاصل از این پرسشنامه با استفاده از آزمون علامت مورد تحلیل قرار گرفته است.

مرحله سوم: در مرحله آخر با استفاده از پرسشنامه، وزن‌دهی به هر یک از عوامل مؤثر انجام گردید. پرسشنامه دوم پژوهش، به منظور تعیین نقش هریک از عوامل در انتخاب استراتژی‌های تطبیقی و منطقی مورد استفاده قرار گرفته است. در این پرسشنامه از خبرگان خواسته شد تا میزان اهمیت هریک از عوامل را در انتخاب استراتژی منطقی و استراتژی تطبیقی در قالب طیف لیکرت از بسیار کم (۱) تا بسیار زیاد (۵) بیان کنند. سؤالات پرسشنامه دارای دو طیف لیکرت مربوط به استراتژی منطقی و استراتژی تطبیقی بوده‌اند.

جامعه آماری در مرحله اول پژوهش، شامل متخصصان و خبرگانی بودند که دارای سابقه و تجربه در حوزه مدیریت پروژه فناوری اطلاعات می‌باشند. ملاک انتخاب افراد اساتید دانشگاهی در حوزه فناوری اطلاعات و با سابقه مدیریت حداقل دو پروژه در این حوزه و نیز مدیران فناوری اطلاعات با سابقه بیش از ۵ پروژه بوده است. در نهایت تعداد افراد انتخاب شده برای گروه کانونی در این پژوهش ۵ نفر بوده است که حاضر به همکاری در پروژه شده‌اند. در مرحله دوم و سوم پژوهش نیز جامعه آماری شامل متخصصان و خبرگان حوزه مدیریت پروژه فناوری اطلاعات و مدیریت استراتژیک فناوری اطلاعات بوده‌اند که براساس روش قضاوتی به نمونه‌گیری پرداخته شده است. ملاک قضاوت در این دو مرحله داشتن حداقل چهار سال سابقه تدریس در حوزه مدیریت پروژه فناوری اطلاعات، داشتن کتاب، پایان نامه یا حداقل دو مقاله علمی معتبر در حوزه پژوهش، داشتن حداقل پنج سال تجربه در پروژه‌های فناوری اطلاعات بوده است. در مرحله دوم ۳۴ پرسشنامه از ۷۶ پرسشنامه‌ی ارسال شده از طریق ایمیل و تماس حضوری، دریافت گردید و مورد تحلیل قرار گرفت و در مرحله سوم ۲۱ پرسشنامه جمع-آوری شد. به منظور اطمینان از اعتبار یافته‌های حاصل از

گروه کانونی از استراتژی بازخور مشارکت‌کنندگان و نیز دریافت نظرات همکاران بهره گرفته شده است بدین صورت که تفسیرها و نتایج به مشارکت کنندگان و نیز دیگر همکاران و محققان ارائه شده و موارد بد درک شده، اصلاح گردیده است.

جهت تعیین روایی پرسشنامه از روش اعتبار محتوا استفاده شده است و پرسشنامه طراحی شده در اختیار ۴ نفر از اساتید دانشگاه در حوزه فناوری اطلاعات قرار گرفت و اصلاحات لازم صورت پذیرفت. به منظور تعیین پایایی پرسشنامه با استفاده از نرم‌افزار SPSS، ضریب آلفای کرونباخ (جداول ۴ و ۵) محاسبه گردید که نتایج آن در جدول (۴) ارائه شده است.

جدول ۴: ضریب آلفای کرونباخ برای عوامل پرسشنامه اول

متغیرها	مقدار آلفای کرونباخ	تفسیر	تعداد آیتم‌های موردبررسی
عوامل انسانی	۰,۸۶۱	پایایی بالا	۶
عوامل فرایندی	۰,۷۷۲	پایایی قابل‌قبول	۵
عوامل مرتبط با سیستم	۰,۸۵	پایایی بالا	۷
عوامل زمینه‌ای	۰,۷۶۳	پایایی قابل‌قبول	۷

همانطور که مشاهده می‌گردد تمامی عوامل دارای مقدار آلفای بالاتر از مقدار قابل قبول ۰/۶ [۱۳] می‌باشند.

۴- یافته ها

انتخاب استراتژی مدیریت پروژه‌های فناوری اطلاعات قبل از پیاده‌سازی و اقدام برای مدیریت پروژه، گام حساسی است که مدیران در نقش استراتژیست‌ها باید برای سازمان خود انجام دهند. از آنجایی که عوامل تأثیرگذار در انتخاب استراتژی مدیریت پروژه فناوری اطلاعات زیاد بوده و تجزیه‌وتحلیل و تصمیم‌گیری به عوامل متعددی بستگی دارد، انتخاب این استراتژی در اکثر سازمان‌ها با مشکل روبه‌رو است. به همین منظور، این پژوهش به دنبال توسعه متدولوژی جهت انتخاب استراتژی مناسب برای مدیریت پروژه‌های فناوری اطلاعات بوده است. یافته‌های حاصل از تحلیل مباحث گروه کانونی و نیز مرور ادبیات که در مرحله اول پژوهش صورت گرفته در جدول(۵) نشان داده شده است.

جدول۵: عوامل مؤثر در انتخاب استراتژی پروژه‌های فناوری اطلاعات

ابعاد	شاخص‌ها	منبع	استخراج شده از گروه کانونی
عوامل انسانی	اندازه تیم	کوکبرن (۲۰۰۰)، گیمباسو و همکاران (۲۰۱۱)، هارب و همکاران (۲۰۱۵)، والاسیچ و همکاران (۲۰۱۵)، استلر و همکاران (۲۰۱۴)	*
	نیاز به مشارکت کاربر در توسعه پروژه	استلر و همکاران (۲۰۱۴)، المگرن (۲۰۱۴)	*
	اهمیت تعامل تیم پروژه در طول توسعه	اطلاعات کسب‌وکار بین‌المللی (۲۰۱۰)، گیمباسو و همکاران (۲۰۱۱)، هارب و همکاران (۲۰۱۵)، استلر و همکاران (۲۰۱۴)	*
	میزان مهارت و تجربه تیم پروژه درحال‌توسعه	مرکز خدمات پزشکی و درمانی (۲۰۰۸)	*
	میزان درک و آگاهی کاربر از پروژه درحال‌توسعه	مرکز خدمات پزشکی و درمانی (۲۰۰۸)، مدحا (۱۹۹۰)	*
	میزان ثبات تیم توسعه		*
عوامل فرایندی	امکان ارائه پروژه به‌صورت افزایشی و تکاملی (ماژولار بودن پروژه)	گیمباسو و همکاران (۲۰۱۱)، هارب و همکاران (۲۰۱۵)، دینگسویر (۲۰۱۲)	*
	مدت زمان تحویل سیستم نهایی		*

ابعاد	شاخص‌ها	منبع	استخراج شده از گروه کانونی
	میزان اهمیت مستندات	گیمباسو و همکاران (۲۰۱۱)، هارب و همکاران (۲۰۱۵) و استلر و همکاران (۲۰۱۴)	*
	میزان اهمیت صرفه‌جویی در هزینه‌ها در طول توسعه	استلر و همکاران (۲۰۱۴)، هارب و همکاران (۲۰۱۵)	*
عوامل مرتبط با سیستم	نقش استراتژیک (اهمیت) پروژه برای سازمان	رحیمیان و رامسین (۲۰۰۸)، والاسیچ و همکاران (۲۰۱۵)	*
	میزان اهمیت امنیت سیستم	کوکبرن (۲۰۰۰)، گیمباسو و همکاران (۲۰۱۱)، هارب و همکاران (۲۰۱۵)، والاسیچ و همکاران (۲۰۱۵)	*
	میزان تغییر نیازمندی‌ها	رحیمیان و رامسین (۲۰۰۸)، گیمباسو و همکاران (۲۰۱۱)، استلر و همکاران (۲۰۱۴)	*
	میزان پیچیدگی سیستم	لاسا و همکاران (۲۰۰۷)، گیمباسو و همکاران (۲۰۱۱)	*
	میزان عدم اطمینان نسبت به الزامات سیستم در طی فرایند توسعه	کوکبرن (۲۰۰۰)، گیمباسو و همکاران (۲۰۱۱)، مدحا (۱۹۹۰) و وکیاتو (۲۰۱۵)	
عوامل زمینه‌ای	اهمیت در دسترس بودن منابع		*
	میزان پایداری محیط	هارب و همکاران (۲۰۱۵)، والاسیچ و همکاران (۲۰۱۵)	*
	میزان پویایی صنعت	هارب و همکاران (۲۰۱۵)	
	میزان اهمیت فرهنگ خلاقیت/ یادگیری	هارب و همکاران (۲۰۱۵)، والاسیچ و همکاران (۲۰۱۵)، دینگسویر (۲۰۱۲) و راموس و موتا (۲۰۱۴)	*
	میزان کنترل پروژه‌ها در سازمان	کوکبرن (۲۰۰۰)، لاسا و همکاران (۲۰۰۷)، اطلاعات کسب‌وکار بین‌المللی (۲۰۱۰)	*
	اولویت پروژه	اطلاعات کسب‌وکار بین‌المللی (۲۰۱۰)، گیمباسو و همکاران (۲۰۱۱)	*

در مرحله دوم، به منظور ارزیابی شاخص‌ها و عوامل مرتبط با انتخاب استراتژی در پروژه‌های فناوری اطلاعات به نظرسنجی از خبرگان پرداخته شده است. داده‌های حاصل از پرسشنامه‌های گردآوری شده در این مرحله با استفاده از آزمون علامت مورد تحلیل قرار گرفته‌اند. سؤال اصلی در این مرحله این بوده است که آیا عوامل و شاخص‌های یادشده در پرسشنامه از دید خبرگان مورد تأیید می‌باشند یا خیر؟ نتایج حاصل از آزمون علامت در جدول (۶) ارائه شده است.

جدول ۶: نتایج آزمون علامت

بعد اصلی	نام متغیر	تعداد علامت‌های مثبت	سطح معناداری	مقدار آماره مشاهده‌شده	مقدار آماره جدول	نتیجه‌گیری
عوامل انسانی	میزان گستردگی تیم توسعه (اندازه تیم)	۳۲	۰,۰۰	۵,۲۳۴	۱,۶۴۵	تأیید
	نیاز به مشارکت کاربر در توسعه پروژه	۳۱	۰,۰۰۰	۵,۲۳	۱,۶۴۵	تأیید
	اهمیت تعامل تیم پروژه در طول توسعه	۳۳	۰,۰۰	۵,۳۴۲	۱,۶۴۵	تأیید
	میزان تجربه تیم پروژه‌ی در حال	۲۹	۰,۰۰	۴,۸۷	۱,۶۴۵	تأیید

بعد اصلی	نام متغیر	تعداد علامت‌های مثبت	سطح معناداری	مقدار آماره مشاهده‌شده	مقدار آماره جدول	نتیجه‌گیری
	توسعه					
	میزان درک و آگاهی کاربر از پروژه در حال توسعه	۲۸	۰,۰۰	۴,۸۲	۱,۶۴۵	تأیید
	میزان ثبات تیم توسعه	۲۷	۰,۰۰	۴,۶۷۴	۱,۶۴۵	تأیید
عوامل فرایندی	میزان بلوغ متدلوژی مدیریت پروژه فناوری اطلاعات	۳۱	۰,۰۰	۵,۱۴۲	۱,۶۴۵	تأیید
	امکان ارائه پروژه به‌صورت مرحله‌ای	۳۱	۰,۰۰	۵,۱۴۲	۱,۶۴۵	تأیید
	میزان اهمیت تحویل سریع پروژه	۳۲	۰,۰۰	۵,۲۳۸	۱,۶۴۵	تأیید
	میزان اهمیت مستندات در طول فرایند پروژه	۳۰	۰,۰۰	۴,۸۱۱	۱,۶۴۵	تأیید
	میزان اهمیت صرفه‌جویی در هزینه‌ها در طول توسعه	۲۸	۰,۰۰	۴,۷۶۴	۱,۶۴۵	تأیید
عوامل مرتبط با سیستم	نقش استراتژیک (اهمیت) پروژه برای سازمان	۳۲	۰,۰۰	۵,۲۳۴	۱,۶۴۵	تأیید
	میزان اهمیت امنیت سیستم	۳۱	۰,۰۰۰	۵,۲۳	۱,۶۴۵	تأیید
	تنوع کاربران سیستم	۳۳	۰,۰۰	۵,۳۴۲	۱,۶۴۵	تأیید
	میزان تغییر نیازمندی‌ها	۲۹	۰,۰۰	۴,۸۷	۱,۶۴۵	تأیید
	میزان گستردگی کارکردهای (نیازمندی‌ها) سیستم	۲۸	۰,۰۰	۴,۸۲	۱,۶۴۵	تأیید
	میزان پیچیدگی سیستم	۲۷	۰,۰۰	۴,۶۷۴	۱,۶۴۵	تأیید
	میزان عدم اطمینان نسبت به الزامات سیستم در طی فرایند توسعه	۲۹	۰,۰۰	۴,۹۵	۱,۶۴۵	تأیید
عوامل زمینه‌ای	اهمیت در دسترس بودن منابع	۳۰	۰,۰۰	۵,۱۲۳	۱,۶۴۵	تأیید
	میزان پویایی محیط سازمانی	۲۹	۰,۰۰۰	۴,۹۸	۱,۶۴۵	تأیید
	میزان پویایی صنعت فناوری اطلاعات	۳۲	۰,۰۰	۵,۳۴	۱,۶۴۵	تأیید
	میزان بلوغ صنعت پروژه در حال توسعه	۳۳	۰,۰۰	۵,۴۱۲	۱,۶۴۵	تأیید
	میزان اهمیت فرهنگ خلاقیت/ یادگیری	۳۱	۰,۰۰	۵,۱۲۳	۱,۶۴۵	تأیید
	میزان نظام‌مند بودن کنترل پروژه‌ها در سازمان	۳۱	۰,۰۰	۵,۱۲	۱,۶۴۵	تأیید
	اولویت پروژه	۳۲	۰,۰۰	۵,۳۵۴	۱,۶۴۵	تأیید

همان‌طور که در جدول (۶) مشخص شده است، از آنجا که مقدار آماره مشاهده شده از آماره جدول بیشتر می‌باشد می‌توان گفت با احتمال ۹۵٪ تمامی شاخص‌های مربوط به هر یک از عوامل انسانی، فرایندی، مرتبط با سیستم و زمینه‌ای مورد تأیید خبرگان می‌باشند.

در مرحله سوم، پرسشنامه مربوط به تعیین وزن هریک از عوامل و شاخص‌ها در استراتژی‌های مورد نظر مدیریت پروژه فناوری اطلاعات در اختیار خبرگان قرار گرفته است. در ادامه نتایج مربوط به اهمیت و وزن هریک از عوامل در انتخاب استراتژی های منطقی و تطبیقی مطرح شده است.

۴-۱- وزن شاخص‌های هر یک از عوامل در انتخاب رویکرد های منطقی و تطبیقی

بعد از محاسبه ارزش‌های محاسبه شده هر یک از شاخص‌ها، با توجه به این‌که باید مجموع اوزان شاخص‌های هر عامل برابر با یک شود، وزن هر شاخص با تقسیم ارزش محاسبه شده بر جمع ارزش‌ها محاسبه گردید که نتایج آن در جدول (۷) ارائه شده است.

جدول۷: وزن شاخص‌ها طبق رویکرد منطقی

شاخص‌های عوامل انسانی	ارزش محاسبه شده	وزن
میزان ثبات تیم توسعه	۳,۷۱۳	۰,۱۸۳
اهمیت تعامل تیم پروژه در طول توسعه	۳,۴۷۵	۰,۱۷۱
نیاز به مشارکت کاربر در توسعه پروژه	۳,۳۳۲	۰,۱۶۴
میزان تجربه تیم پروژه‌ی در حال توسعه	۳,۲۸۵	۰,۱۶۲
میزان درک و آگاهی کاربر از پروژه در حال توسعه	۳,۲۳۷	۰,۱۵۹
میزان گستردگی تیم توسعه	۳,۱۹۰	۰,۱۵۷
شاخص‌های عوامل فرایندی	**ارزش محاسبه شده**	**وزن**
میزان اهمیت مستندات در طول فرایند پروژه	۳,۸۵۶	۰,۲۰۴
میزان اهمیت تحویل سریع پروژه	۳,۷۶۴	۰,۱۹۹
امکان ارائه پروژه به‌صورت مرحله‌ای	۳,۷۶۱۳	۰,۱۹۸۱
میزان بلوغ متدلوژی مدیریت پروژه فناوری اطلاعات در سازمان	۳,۷۶۱۲	۰,۱۹۸۰
میزان اهمیت صرفه‌جویی در هزینه‌ها در طول توسعه	۳,۷۵۸	۰,۱۹۷
شاخص‌های عوامل مرتبط با سیستم	**ارزش محاسبه شده**	**وزن**
میزان اهمیت امنیت سیستم	۴,۲۸۵	۰,۱۷۳
نقش استراتژیک (اهمیت) پروژه برای سازمان	۳,۹۰۴	۰,۱۵۷
میزان گستردگی کارکردها	۳,۶۱۸	۰,۱۴۶
میزان پیچیدگی سیستم	۳,۵۲۲	۰,۱۴۲
تنوع کاربران سیستم	۳,۲۸۵	۰,۱۳۲
میزان تغییر نیازمندی‌ها	۳,۱۸۹	۰,۱۲۸
میزان عدم اطمینان نسبت به الزامات سیستم در طی فرایند توسعه	۲,۹۵۲	۰,۱۱۹
شاخص‌های عوامل زمینه‌ای	**ارزش محاسبه شده**	**وزن**
میزان نظام‌مند بودن کنترل پروژه‌ها در سازمان	۴,۳۳۲	۰,۱۷۶
اولویت پروژه	۳,۹۰۴	۰,۱۵۹
میزان پویایی صنعت فناوری اطلاعات	۳,۶۶۵	۰,۱۴۹
میزان بلوغ صنعت پروژه درحال‌توسعه	۳,۳۸	۰,۱۳۸
میزان پویایی محیط سازمانی	۳,۳	۰,۱۳۴
میزان اهمیت فرهنگ خلاقیت/ یادگیری	۳,۰۴۶	۰,۱۲۴
اهمیت در دسترس بودن منابع	۲,۸۵۶	۰,۱۱۶

جدول (۷) میزان وزن هر یک از عوامل را در انتخاب رویکرد منطقی نشان می‌دهد. همان‌طور که در این جدول ملاحظه می‌شود طبق رویکرد منطقی شاخص‌های میزان ثبات تیم توسعه در عوامل انسانی، میزان اهمیت مستندات در طول فرایند پروژه برای عامل فرایندی، میزان اهمیت امنیت سیستم در بعد عوامل مرتبط با سیستم و میزان نظام‌مند بودن کنترل پروژه‌ها در سازمان برای عوامل زمینه‌ای بیشترین وزن را دارا هستند. به عنوان مثال، در شرایط که ثبات تیم پروژه بالا باشد، امنیت در پروژه فناوری اطلاعات اهمیت داشته باشد، کنترل پروژه ها در سازمان نظام‌مند باشد تأکید بیشتر بر استراتژی منطقی می‌باشد تا استراتژی تطبیقی. جدول (۸) نتایج مربوط به وزن هریک از شاخص‌ها در انتخاب استراتژی تطبیقی ارائه شده است.

جدول ۸: وزن شاخص‌ها طبق رویکرد تطبیقی

شاخص‌های عوامل انسانی	ارزش محاسبه شده	وزن
میزان تجربه تیم پروژه‌ی در حال توسعه	۴,۱۹	۰,۱۷۹
اهمیت تعامل تیم پروژه در طول توسعه	۳,۹۹۹	۰,۱۷۱
میزان درک و آگاهی کاربر از پروژه در حال توسعه	۳,۹۵۱	۰,۱۶۹
نیاز به مشارکت کاربر در توسعه پروژه	۳,۸۵۶	۰,۱۶۵
میزان ثبات تیم توسعه	۳,۷۶۱	۰,۱۶۱
میزان گستردگی تیم توسعه	۳,۵۷	۰,۱۵۳
شاخص‌های عوامل فرایندی	**ارزش محاسبه شده**	**وزن**
میزان اهمیت صرفه‌جویی در هزینه‌ها در طول توسعه	۳,۶۱۸	۰,۲۱
میزان اهمیت تحویل سریع پروژه	۳,۵۷	۰,۲۰۷
امکان ارائه پروژه به‌صورت مرحله‌ای	۳,۵۲۱	۰,۲۰۴
میزان اهمیت مستندات در طول فرایند پروژه	۳,۲۸۵	۰,۱۹۱
میزان بلوغ متدلوژی مدیریت پروژه فناوری اطلاعات در سازمان	۳,۱۸۹	۰,۱۸۵
شاخص‌های عوامل مرتبط با سیستم	**ارزش محاسبه شده**	**وزن**
میزان عدم اطمینان نسبت به الزامات سیستم در طی فرایند توسعه	۴,۰۴۷	۰,۱۵۴
میزان پیچیدگی سیستم	۳,۹۹۹	۰,۱۵۲
میزان تغییر نیازمندی‌ها	۳,۹۵۲	۰,۱۵۰
تنوع کاربران سیستم	۳,۷۶۱	۰,۱۴۳
میزان گستردگی کارکردها	۳,۵۲۳	۰,۱۳۴
میزان اهمیت امنیت سیستم	۳,۴۷۵	۰,۱۳۲
نقش استراتژیک (اهمیت) پروژه برای سازمان	۳,۴۲۷	۰,۱۳۰
شاخص‌های عوامل زمینه‌ای	**ارزش محاسبه شده**	**وزن**
میزان اهمیت فرهنگ خلاقیت/ یادگیری	۴,۲۸۴	۰,۱۷۵
میزان پویایی صنعت فناوری اطلاعات	۴,۰۱	۰,۱۶۴
میزان پویایی محیط سازمانی	۳,۸۰۸	۰,۱۵۵
اهمیت در دسترس بودن منابع	۳,۶۶۵	۰,۱۵۰
اولویت پروژه	۳,۴۲۸	۰,۱۴۰
میزان بلوغ صنعت پروژه درحال‌توسعه	۳,۴۲۷	۰,۱۳۹
میزان نظام‌مند بودن کنترل پروژه‌ها در سازمان	۲,۸۰۸	۰,۱۱۴

مطابق با جدول (۸) در رویکرد تطبیقی میزان تجربه تیم پروژه‌ی در حال توسعه برای بعد انسانی، میزان اهمیت صرفه‌جویی در هزینه‌ها در طول توسعه در بعد فرایندی، میزان عدم اطمینان نسبت به الزامات سیستم در طی فرایند توسعه در بعد مرتبط با سیستم و میزان اهمیت فرهنگ خلاقیت/ یادگیری برای عوامل زمینه‌ای طبق نظر خبرگان بیشترین اولویت را دارند. همانطور که در جدول ۸ نشان داده شده است، به‌عنوان مثال از آن‌جا که وزن عدم اطمینان نسبت به الزامات سیستم، فرهنگ خلاقیت، صرفه‌جویی هزینه‌ها در طول پروژه در انتخاب این رویکرد بالا است، در صورتی که عدم اطمینان نسبت به الزامات سیستم بالا باشد، تأکید سازمان بر فرهنگ خلاقیت/ یادگیری باشد و صرفه‌جویی در هزینه‌ها در پروژه‌ی فناوری اطلاعات اهمیت زیادی داشته باشد تأکید بر استراتژی تطبیقی خواهد بود تا منطقی.

برای محاسبه وزن هر حوزه اصلی، ابتدا ارزش هر حوزه محاسبه می‌شود بدین‌صورت که مجموع ارزش محاسبه شده هر حوزه (مجموع ارزش شاخص‌های هر حوزه) بر تعداد شاخص‌های آن حوزه تقسیم می‌شود. با توجه به اینکه مجموع اوزان همه حوزه‌ها باید برابر با یک شود وزن هر حوزه با تقسیم ارزش محاسبه شده هر حوزه بر مجموع ارزش حوزه‌ها به دست می‌آید که نتایج آن در جدول (۹) ارائه شده است.

جدول ۹: وزن عوامل اصلی در رویکرد منطقی و تطبیقی

	حوزه اصلی	ارزش محاسبه شده هر حوزه	وزن هر حوزه
وزن عوامل اصلی در رویکرد منطقی	عوامل فرایندی	۳,۷۸	۰,۲۶۶
	عوامل مرتبط با سیستم	۳,۵۳۶	۰,۲۴۹
	عوامل زمینه‌ای	۳,۴۹۷	۰,۲۴۶
	عوامل انسانی	۳,۳۷۲	۰,۲۳۷
وزن عوامل اصلی در رویکرد تطبیقی	عوامل انسانی	۳,۸۸۷	۰,۲۶۴
	عوامل مرتبط با سیستم	۳,۷۴۰	۰,۲۵۴
	عوامل زمینه‌ای	۳,۶۳۲	۰,۲۴۷
	عوامل فرایندی	۳,۴۳۶	۰,۲۳۳

همان‌طور که در جدول (۹) نشان داده شده است در رویکرد منطقی اولویت ابعاد اصلی به ترتیب عبارتند از: عوامل فرایندی، عوامل مرتبط با سیستم، عوامل زمینه‌ای و درنهایت عوامل انسانی. درحالی‌که طبق رویکرد تطبیقی اولویت عوامل اصلی به ترتیب شامل عوامل انسانی، عوامل مرتبط با سیستم، عوامل زمینه‌ای و در نهایت عوامل فرایندی می‌باشد.

۴-۲- چگونگی کاربرد رویکرد انتخاب استراتژی غالب مدیریت پروژه فناوری اطلاعات

برای شناخت رویکرد مناسب استراتژی مدیریت پروژه فناوری اطلاعات در سازمان می‌توان از پرسشنامه طراحی شده در این پژوهش بهره برد. به منظور شناسایی رویکرد غالب در سازمان، ابتدا میانگین هر عامل در وزن به‌دست‌آمده آن عامل، در هر دو رویکرد منطقی و تطبیقی به‌صورت مجزا ضرب می‌شود تا میزان همسویی آن با استراتژی‌های منطقی و تطبیقی به دست آید. سپس از اعداد به‌دست‌آمده در هر رویکرد به‌طور جداگانه میانگین گرفته تا میزان نزدیکی با هر یک از انواع رویکردها مشخص شود. بدین ترتیب رویکرد غالب مدیریت پروژه‌های فناوری اطلاعات در سازمان مشخص می‌شود. فرمول‌های مورد استفاده به ترتیب زیر است:

$$LO_{Hum} = HUM * W_{humLO}$$
$$AD_{Hum} = HUM * W_{humAD}$$

$$Lo=(LO_{Hum} + LO_{Pro} + LO_{Sys} + LO_{Con})/4$$
$$Ad=(AD_{Hum} + AD_{Pro} + AD_{Sys} + AD_{Con})/4$$

فرمول‌های ذکر شده برای هر چهار عامل باید به صورت جداگانه محاسبه گردد.

۵- بحث و نتیجه‌گیری

اکنون، مدیریت پروژه به عنوان ابزاری شناخته می‌شود که از طریق آن برنامه‌های استراتژیک در شرکت‌ها مدیریت می‌شوند. بنابراین، آنها پل پایداری تجارت را تشکیل می‌دهند. از جمله تحولات مختلفی که در مدیریت پروژه ایجاد شده است، تعیین عوامل مهم موفقیت است[۴۰]. یکی از عوامل مهم موفقیت پروژه‌ها داشتن رویکرد استراتژیک به مدیریت پروژه است که مدیران پروژه را از تمرکز تنها بر روی کارایی به سمت اثربخشی و کارایی و دستیابی به نتایج کسب و کار و موفقیت پروژه‌ها سوق می‌دهد. در نتیجه مدیران پروژه‌های فناوری اطلاعات برای دستیابی به مزیت رقابتی و عملکرد بهتر باید چارچوب برنامه‌ریزی و اجرایی مطابق با رویکرد استراتژیک داشته باشند. در این مقاله تلاش شده است روشی عملی و کاربردی جهت انتخاب استراتژی مناسب مدیریت پروژه-های فناوری اطلاعات مدنظر قرار گیرد. از این رو، پس از بررسی جامع ادبیات موضوع و تشکیل گروه کانونی عوامل تأثیرگذار شناسایی شدند که شامل عوامل فرایندی، عوامل انسانی، عوامل مرتبط با سیستم و عوامل زمینه‌ای بوده‌اند.

در میان شاخص‌های مرتبط با عوامل انسانی، میزان ثبات تیم توسعه مؤثرترین شاخص در انتخاب رویکرد منطقی و میزان تجربه تیم پروژه مهمترین شاخص‌های مؤثر بر انتخاب رویکرد تطبیقی می‌باشد. در میان شاخص‌های مرتبط با عوامل فرایندی، میزان اهمیت مستندات در

طول فرایند پروژه موثرترین شاخص در انتخاب رویکرد منطقی و شاخص میزان اهمیت صرفه‌جویی در هزینه‌ها در طول توسعه مؤثرترین شاخص در انتخاب رویکرد تطبیقی می‌باشد. در بعد عوامل مرتبط با سیستم میزان اهمیت امنیت سیستم مؤثرترین شاخص در انتخاب رویکرد منطقی و میزان عدم اطمینان نسبت به الزامات سیستم در طی فرایند توسعه مهم‌ترین شاخص در انتخاب رویکرد تطبیقی بوده است. در بعد عوامل زمینه‌ای شاخص میزان نظام‌مند بودن کنترل پروژه‌ها مؤثرترین شاخص در انتخاب رویکرد منطقی و میزان اهمیت خلاقیت/ یادگیری مهمترین شاخص در انتخاب رویکرد تطبیقی می‌باشد. همانطور که در جدول ۵ ملاحظه می‌شود برخی از شاخص‌های استخراج شده در این تحقیق در پژوهش‌های پیشین وجود نداشت و از گروه کانونی استخراج شده است. ضمناً نوآوری اصلی این پژوهش مطرح شدن مباحث استراتژیک در مدیریت پروژه در حوزه مدیریت پروژه‌های فناوری اطلاعات می‌باشد که تا کنون مورد بررسی قرار نگرفته است. در مقاله پاپک‌شیلد و بایررایت (۲۰۱۶)، ویژگی‌های برنامه‌ریزی استراتژیک منتج شده از رویکردهای منطقی و تطبیقی در مدیریت پروژه به‌کار رفته شده است که ترکیب این رویکردها به موفقیت پروژه‌ها کمک شایانی نموده است ولی به‌طور خاص سخنی از مدیریت پروژه‌های فناوری اطلاعات به میان نیامده است. در پژوهش پیش رو بعد عوامل انسانی به‌عنوان فاکتوری مؤثر در انتخاب استراتژی پروژه‌های فناوری اطلاعات مورد تأیید خبرگان قرار گرفت و داگل (۲۰۰۶) و آبراهامسون و همکاران (۲۰۱۷) نیز از آن به عنوان یکی از مؤلفه‌های مؤثر در انتخاب رویکرد توسعه سیستم‌های اطلاعاتی یاد کرده‌اند. عوامل فرایندی در پژوهش هارب و همکاران (۲۰۱۵) و آبراهامسون و همکاران (۲۰۱۷) نیز از عوامل مهم در انتخاب رویکرد مناسب در پروژه‌های سیستم‌های اطلاعاتی نام برده شده است. عوامل مرتبط با سیستم نیز همانطور که در کتاب لسکواسکی (۲۰۱۱) و گیمباسو و همکاران (۲۰۱۱) مطرح گردیده است در پیشبرد موفق پروژه‌های فناوری اطلاعات تأثیرگذارند که در این پژوهش به طور جامع شاخص‌های مرتبط با آن شناسایی شده است. همچنین عوامل زمینه-ای در پژوهش داگل (۲۰۰۶) و هارب و همکاران (۲۰۱۵) در انتخاب رویکرد کلی توسعه سیستم‌های اطلاعاتی ذکر شده‌اند.

به طور کلی می توان مهمترین دانش‌افزایی‌های زیر را برای پژوهش حاضر بیان کرد:

اول، علیرغم اهمیت استراتژی در موفقیت مدیریت پروژه، مطالعات محدودی به حوزه انتخاب استراتژی مدیریت پروژه به طور کلی و مدیریت پروژه های فناوری اطلاعات به طور خاص پرداخته اند. پژوهش حاضر، با توسعه متدولوژی انتخاب استراتژی پروژه های فناوری اطلاعات در زمره اولین پژوهش ها در این حوزه می باشد.

دوم، استخراج لیست نسبتا جامعی از عوامل تاثیرگذار بر استراتژی سازی مدیریت پروژه های فناوری اطلاعات که دربرگیرنده عوامل انسانی، فرایندی، عوامل مرتبط با سیستم و عوامل زمینه‌ای می‌باشد از دیگر نوآوری‌های پژوهش می باشد. پژوهش های پیشین به طور پراکنده و مجزا به برخی از عوامل اشاره کرده اند. پژوهش حاضر، سعی در توسعه درک کل نگرانه و جامعی از عوامل انتخاب استراتژی پروژه‌های فناوری اطلاعات داشته است.

سوم، تحقیقات پیشین صرفا به بیان کلی تعدادی از عوامل تاثیرگذار بر انتخاب رویکرد مدیریت پروژه‌های فناوری اطلاعات پرداخته‌اند و فاقد روشی جهت تعیین استراتژی مناسب با شرایط پروژه بوده‌اند. متدولوژی پیشنهادی در پژوهش، علاوه بر شناسایی عوامل و وزن هریک از آنها در انتخاب دو استراتژی غالب، به چگونگی و نحوه انتخاب استراتژی نیز پرداخته است.

چهارم، توجه همزمان به دو رویکرد منطقی و تطبیقی در استراتژی انتخابی مدیریت پروژه های فناوری اطلاعات، بطوریکه استراتژی انتخابی در بازه‌ای از رویکرد منطقی و رویکرد تطبیقی جای می‌گیرد و صرفا بر یک استراتژی منفرد متمرکز نمی‌باشد.

در پایان باید گفت که علیرغم یافته‌های ارزشمند پژوهش، این تحقیق نیز همانند دیگر پژوهش‌ها با محدودیت‌هایی روبرو بوده است که براساس آنها می‌بایست پیشنهاداتی برای پژوهش‌های آتی ارائه گردد. با توجه به برخی از نتایج مهم و جدید تحقیق حاضر به محققان آتی توصیه می‌شود به کاربرد نتایج در سازمان‌های مختلف و مقایسه استراتژی‌های بدست آمده بپردازند. از آنجا که در این

پژوهش به طور کلی به تدوین استراتژی برای پروژه‌های فناوری اطلاعات پرداخته شده است پیشنهاد می‌شود با توجه به ماهیت متفاوت پروژه‌ها، مانند پروژه‌های نرم-افزاری، سخت افزاری، زیرساختی و ... به ارائه رویکرد انتخاب استراتژی مدیریت پروژه بپردازند. در نهایت پیشنهاد می‌شود با استفاده از رویکردهای کمی چون مدلسازی معادلات ساختاری و یا رویکرد نقشه ذهنی به تحلیل روابط بین عوامل تأثیرگذار بر انتخاب استراتژی مدیریت پروژه فناوری اطلاعات بپردازند که در حیطه این پژوهش نبوده و به آن پرداخته نشده است.

منابع

۱. ی، محمدی مقدم، ا، بهبودی، ر، حمزه و غ، خدایی. ارائه متدولوژی استراتژی مدیریت دانش بر اساس مارپیچ دانش سازمان. مطالعات مدیریت کسب و کار هوشمند، ۱۳۹۴، ۴(۱۴)، ۲۳-۵۰.

۲. ع، منوریان و ف، غزائیان. بررسی جایگاه نظام مدیریت پروژه در برنامه‌ریزی استراتژیک شرکت‌های پیمانکاری عمومی (مطالعه‌ی موردی شرکت بین‌المللی توسعه و مهندسی پارس). *فصلنامه علمی - پژوهشی مدیریت بازرگانی*، ۱۳۸۹ ۲، (۵).

۳. ش. الهی، ع. حسن زاده، آ. خدیور. ارائه یک متدولوژی برای ایجاد استراتژی مدیریت دانش مطالعه و بررسی سه سازمان نمونه. پژوهشهای مدیریت در ایران، ۱۳۸۹، ۱۴(۳)، ۲۴-۵۹.

4. P. Abrahamsson, O. Salo, J. Ronkainen & J. Warsta. Agile software development methods: Review and analysis. arXiv preprint arXiv:1709.08439, 2017.

5. K. Almgren. Information technology project management processes and practices: A comprehensive study for successful implementation of IT projects. International Journal of Business and Social Science, 2014, 5 (11).

6. K. Artto, J. Kujala, P. Dietrich & M. Martinsuo. What is project strategy?. International Journal of Project Management, 2008, 26(1), 4-12.

7. J. P. Assi–UNIMEP–joao, P. D. Antoniolli–UNIMEP & J. B. de Camargo Junior-UNIMEP. Proposal for Information Technology Strategic Planning Approach. European Journal of Business and Social Sciences, 2016, 5(02), 90-105.

8. M. Bloor, J. Frankland, M.Thomas and K. Robson. Focus Groups in Social Research. Sage, London, 2001.

9. F. Brooks. Difference between Software projects and other types of project, 2015. Retrieved 2018, Jan.19 from http://justinshalom.com/project.

10. S. C. Burgan & D. S. Burgan. One size does not fit all: Choosing the right project approach. Project Management Institute, 2014.

11. J. Cadle & D. Yeates (Eds.). Project management for information systems. Pearson education, 2008.

12. G. Capaldo & P. Rippa. A planned-oriented approach for EPR implementation strategy selection. Journal of Enterprise Information Management, 2009.

13. Centers for Medicare & Medicaid Services. Selecting a Development Approach, 2008. Retrieved 2016, Sept 18 from https://www. cms. gov/Research-Statistics-Data-and-Systems/CMS Information-Technology. XLC/ Downloads/SelectingDevelopmentApproach.pdf.

14. C. C. Chen, H. S. Shih, H. J. Shyur, & K. S.Wu. A business strategy selection of analytic network process. Computers & Mathematics with Applications, 2012, 64(8), 2544-2557.

15. A. Cockburn. Selecting a project's methodology. IEEE software, 2000, 17(4), 64-71.

16. W. A. Cram & M. Marabelli. Have your cake and eat it too? Simultaneously pursuing the knowledge-sharing benefits of agile and traditional development

approaches. Information & Management, 2018, 55(3), 322-339.
17. Dey, P. Chandra. Complete Collection of Project Management Statistics, (September 1, 2016). Retrieved 2017, March, 17, from https://www.planningproapp.com/blog/complete-collection-project-management-statistics-2016.
18. T. Dingsøyr, S. Nerur, V. Balijepally & N. B. Moe. A decade of agile methodologies: Towards explaining agile software development, 2012.
19. P. Duggal. Guidelines to support choice of development methodology. 2006, Information Systems BSc Session.
20. O. Einasto. Investigating e-service quality criteria for university library: a focus group study. New Library World, 2014, 115(1/2), 4-14.
21. ESI International, Inc, Successful Solutions Through Agile Project Management, 2010, Retrieved 2017, Nov 01 from https://www.projectsmart.co.uk/white-papers.
22. H.C. Estler, M. NordioC. A. Furia, B. Meyer & J. Schneider. Agile vs. structured distributed software development: A case study. Empirical Software Engineering, 2014, 19(5), 1197-1224.
23. S. Florentine. More than half of IT projects still failing, 2016. Retrieved 2017, Jul 10 from www.cio.com.
24. Gartner. Gartner Says Global IT Spending to Grow 3.7% in 2020 [site comment]. (2019, October 23). Retrieved 2020, May.21,from https://www.gartner.com/en/newsroom/press-releases/2019-10-23-gartner-says-global-it-spending-to-grow-3point7-percent-in-2020.
25. C.V. Geambasu, I. Jianu & A. Gavrila. Influence factors for the choice of a software development methodology. Accounting and Management Information Systems, 2011, 10(4), 479.
26. Y. Harb, C. Noteboom, & S. Sarnikar. Evaluating Project Characteristics for Selecting the Best-fit Agile Software Development Methodology: A Teaching Case. Journal of the Midwest Association for Information Systems, 2015, (1), 33.
27. G. Heerkens. Introducing the revolutionary strategic project management maturity model (SPM3). In annual North American meeting of the Project Management Institute (p. 1), 2007.
28. G. J. Hidding, J. M. Nicholas. A new way of thinking about IT project management practices: early empirical results. J.Organ. Comput. Electron. Commer. 2017, 27(1),81–95.
29. T. Marchewka. Information Technology Project Management: Providing Measurable Organizational Value. 5th Edition: Wiley Publishing, 2015.
30. K. S. Koong, L. C. Liu & X. Liu . A study of the demand for information technology professionals in selected internet job portals. Journal of Information Systems Education, 2020, 13(1), 4.
31. D. J. Laird. The impact of planning and other organizational factors on the success of small information technology projects. Doctoral dissertation, University of Pittsburgh, 2016.
32. J. Laskowski. Agile IT security implementation methodology. Packt Publishing Ltd, 2011.
33. J. Lassa, S. Kerkman, R. Decker, K. Rhoades, A. Lasee, D. Cullen, ... & J. Parisi. Information Technology Projects, 2007.
34. O.T. Massey. A proposed model for the analysis and interpretation of focus groups in evaluation research. Evaluation and program planning, 2011, 34(1), 21-28.
35. Medium. 16 Project Management Stats You Can't Ignore [2019], available at: https://medium.com/crowdbotics/hips-dont-lie-15-project-management-stats-you-can-t-ignore-6f655060ef30.

36. J. Modha, A. Gwinnett & M. Bruce. A review of information systems development methodology (ISDM) selection techniques. Omega, 1990, 18(5), 473-490.
37. D. L. Morgan. Focus groups as qualitative research (Vol. 16). Sage publications, 1996.
38. K. E. Papke-Shields, & K. M. Boyer-Wright. Strategic planning characteristics applied to project management. International Journal of Project Management, 2017, 35(2), 169-179.
39. P. Patanakul and A. J. Shenhar. "What project strategy really is: the fundamental building block in strategic project management", 2012, Project Management Journal, Vol. 43 No. 1, pp. 4-20.
40. A. R. Peña & F. A. Muñoz. Soft Skills as a Critical Success Factor in Project Management. In Handbook of Research on Project Management Strategies and Tools for Organizational Success, 2020 (pp. 376-392). IGI Global.
41. D. Petrović, M. Mihić & V. Obradović. Strategic project management–project strategy and measurement of success. Published as part of SYMORG 2012-XIII International Symposium on Innovative Management & Business Performance, 2012.
42. V. Rahimian & R. Ramsin. Designing an agile methodology for mobile software development: A hybrid method engineering approach. In processdings of Research Challenges in Information Science, 2008. RCIS 2008. Second International Conference, 337–342. IEEE.
43. P. Ramos & C. Mota. Perceptions of success and failure factors in information technology projects: a study from Brazilian companies. Procedia-Social and Behavioral Sciences, 2014, 119, 349-357.
44. A. Rodrigues, J. bowers. System dynamic in project management : a comprative analysis with traditional methods. System Dynamic Reviwe. 1996, 12(2),121-139.
45. A. Rolstadås, I. Tommelein, P. Morten Schiefloe & G. Ballard. Understanding project success through analysis of project management approach. International Journal of Managing Projects in Business, 2014, 7(4), 638-660.
46. J. W. Satzinger, R. B. Jackson & S. D. Burd. Systems analysis and design in a changing world. Cengage learning, 2011.
47. K. Schwalbe. Information technology project management. Cengage Learning, 2015.
48. U. Sekaran & R. Bougie. Research methods for business: A skill building approach. John Wiley & Sons, 2016.
49. F. Sharon. IT project success rates finally improving, 2017,Feb. 27, Retrieved 2020, May. 21, from https://www.cio.com/article/3174516/it-project-success-rates-finally-improving.html.
50. A. Shenhar. What is strategic project leadership?. Open Economics and Management Journal, 2015, 2(1).
51. O. Standing, S. Standing & E. Kordt. Explaining attribution in information technology projects. Journal of Systems and Information Technology, 2016, 18(2), 216-227.
52. D. Truong & T. Jitbaipoon. How can agile methodologies be used to enhance the success of information technology projects?. International Journal of Information Technology Project Management (IJITPM), 2016, 7(2), 1-16.
53. J. S. Valacich, J. F. George & J. A. Hoffer. Essentials of systems analysis and design. 2015, Pearson Education.
54. V. Vinekar, C. W. Slinkman & S. Nerur. Can agile and traditional systems development approaches coexist? An ambidextrous view. Information systems management, 2006, 23(3), 31-42.
55. M. Yaghini, A. Bourouni & R. H. Amiri. A framework for selection of information systems development

methodologies. Computer and information science, 2009, 2(1), 3.

56. L. R. Yang. Implementation of project strategy to improve new product development performance. International Journal of Project Management, 2012, 30(7), 760-770.
57. S. Zaim, A. Turkyılmaz, M. F. Acar, U. Al-Turki & O. F. Demirel. Maintenance strategy selection using AHP and ANP algorithms: a case study. 2012, Journal of Quality in Maintenance Engineering.

سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷
صص: ۳۵- ۴۴

فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

طراحی پردازنده مبتنی بر FPGA برای الگوریتم‌های رمزنگاری سری SHA-2

ندا صدق اهرابی * محمدعلی جبرئیل‌جمالی **
* کارشناس ارشد، گروه مهندسی برق، دانشکده فنی مهندسی، واحد تبریز، دانشگاه آزاد اسلامی، تبریز
** دکتری، گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، واحد شبستر، دانشگاه آزاد اسلامی، شبستر
تاریخ دریافت:۱۳۹۸/۰۴/۲۵ تاریخ پذیرش: ۱۳۹۹/۰۲/۱۳

چکیده

الگوریتم‌های درهم‌ساز ایمن، نوعی از الگوریتم‌های رمزنگاری هستند که اهمیت آن‌ها در جامعه امروزی با بروز کاربردهایی مانند استفاده از ابزارهای دیجیتالی شخصی در راستای حفظ محرمانگی پررنگ‌ترشده‌اند. از طرفی با پیشرفت تکنولوژی، لزوم پیاده‌سازی این الگوریتم‌ها روی بسترهای انعطاف‌پذیر، می‌تواند چالش‌برانگیز باشد. کاهش مساحت و افزایش سرعت اجرای عملیات، چالش‌های اساسی برای طراحی و پیاده‌سازی این دسته از الگوریتم‌ها هستند. در این مقاله یک معماری جدید برای پردازنده مبتنی بر FPGA برای الگوریتم‌های رمزنگاری سری SHA-2 پیشنهادشده است. در پردازنده پیشنهادی استفاده از واحدهای حافظه و مسیر داده چندپورته و به دنبال آن عملکرد موازی پردازنده باعث کاهش بکارگیری منابع و افزایش سرعت پردازش داده‌ها شده است. معماری پردازنده برای الگوریتم‌های رمزنگاری SHA-2 با زبان VHDL مدل‌سازی شده و پیاده‌سازی آن روی بستر FPGA در سری‌های Virtex توسط نرم‌افزار ISE انجام‌شده است. نتایج پیاده‌سازی نشان می‌دهند که پردازنده متراکم پیشنهادی در مقایسه با کارهای پیشین با اهداف مشابه، توانسته با ٪۲۵ افزایش فرکانس کاری برای الگوریتم رمزنگاری SHA-256 و اشغال ٪۵۵ مساحت کمتر برای الگوریتم رمزنگاری SHA-512 حد مطلوبی از توان عملیاتی و کارایی را نیز حفظ نماید. پردازنده پیشنهادی برای کاربردهایی مانند بسترهای سیار مورد اعتماد (TMP)، واحد پول دیجیتال (Bitcoin) و مسیریابی ایمن در شبکه روی تراشه (NoC) مناسب است.

واژه‌های کلیدی: الگوریتم‌های درهم‌ساز ایمن، الگوریتم‌های رمزنگاری سری SHA-2، پردازنده، FPGA،VHDL

۱- مقدمه

رشد سریع ارتباطات الکترونیکی حاکی از آن است که مسائلی در رابطه با امنیت اطلاعات و کاربردهای عملی آن‌هاروزبه‌روز بیشتر حائز اهمیت می‌باشند [۱]. امروزه ابزارهای محاسباتی فراگیر و ارتباطات بی‌سیم، بازگشای چالش‌های فراوانی هستند. اطلاعات حساس شخصی مانند اطلاعات پزشکی و مالی افراد در فضای وسیعی از ابزارهای دیجیتالی و محاسباتی در ارگان‌ها و سازمان‌ها، که با رابط‌های سیمی یا بی‌سیم در ارتباط با یکدیگرند، توزیع شده‌اند. بر همین اساس تحقیقات وسیعی مبنی بر بهبود مدیریت امنیت اطلاعات منطبق بر استانداردهای جهانی در تشکیلات گسترده، ارائه شده است[۲]. از سوی دیگر با رشد سریع تکنولوژی امنیت شبکه‌های کامپیوتری و لزوم آپلود تصاویر و اسناد شخصی در سامانه‌های مختلف، زمینه‌های تحقیقاتی وسیعی در راستای ارتقای امنیت اطلاعات در اینترنت را نیز فراهم ساخته است[۳].

نویسندهٔ عهده‌دار مکاتبات: محمدعلی جبرئیل جمالی m_jamali@itrc.ac.ir

از دیرباز الگوریتم‌های رمزنگاری[۱] به‌عنوان الگوریتم‌هایی با سرعت کم، منابع محاسباتی زیاد و پیاده‌سازی‌های ناکارآمد در پردازنده‌های همه‌منظوره[۲] موردتوجه بوده‌اند[۴]و [۵]؛ همین قضایا انگیزه‌ای برای پیشنهاد و پیاده‌سازی یک معماری جدید با سرعت بالا برای الگوریتم‌های رمزنگاری ایجادکرده است. همچنین امروزه به‌واسطه استفاده فراوان از ابزارهای سیار مانند ابزارهای دیجیتالی شخصی و گوشی‌های هوشمند، برنامه‌های کاربردی جدیدی ظهور کرده‌اند که به دنبال خود خطرهای احتمالی برای سیستم‌های مذکور خواهند داشت. در همین راستا و با توجه به این‌که پیشرفت تکنولوژی اهمیت پیاده‌سازی این الگوریتم‌ها روی بسترهای انعطاف‌پذیر مانند FPGA[۳]ها [۶]، CPLD[۴]ها و ASIC[۵]ها که سطوح امنیتی بالایی را تأمین کنند، پررنگ می‌کند[۹] و [۷]. ازآنچه گفته شد می‌توان نتیجه گرفت که پیاده‌سازی سرویس‌های امنیتی که قابل مصالحه ازنظر مساحت، فرکانس، توان عملیاتی ۶ و کارایی ۷ باشند، امری ضروری است. معماری پیشنهادی در قالب یک پردازنده جهت اجرای دستورالعمل‌های الگوریتم‌های رمزنگاری سری SHA-2 با رویکرد تخصیص حداقل منابع طراحی‌شده است؛ از طرفی پردازنده متراکم مذکور مبتنی بر FPGA بوده که با استفاده از واحدهای حافظه و مسیرداده چندپورته و استفاده از برخی امکانات موجود در بسترهای نامبرده کارهایی در راستای اجرای سریع دستورالعمل‌ها به کار گرفته‌شده است.

SHA مخفف عبارت Secure Hash Algorithm و به معنی الگوریتم درهم‌سازی ایمن است. این الگوریتم‌ها شامل توابعی هستند که عمل فشرده‌سازی را انجام می‌دهند؛ بدین معنی که برای ورودی‌های با طول متفاوت، کدهایی (خروجی‌هایی) با طول ثابت تولید می‌کنند که این کدها یک‌طرفه بوده و نمی‌توان از روی خروجی الگوریتم، ورودی آن را تشخیص داد[۱۰].

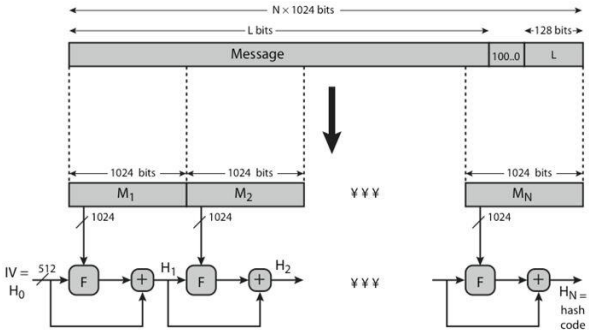
الگوریتم‌های درهم‌ساز ایمن مجموعه‌ای از توابع درهم‌سازی محسوب می‌شوند که از سال ۱۹۹۵ تاکنون در چهار نسخه توسط آژانس امنیت ملی ایالت متحده آمریکا (NSA[۸]) طراحی و توسط موسسه ملی فناوری و استانداردها (NIST[۹]) به‌عنوان استاندارد پردازش اطلاعات انتشاریافته‌اند. از سال ۲۰۱۰ نسخه‌های SHA-0 و SHA-1 با گونه‌های امن‌تر یعنی خانواده SHA-2 جایگزین شده‌اند[۱۱] و [۱۲]؛ ولی باوجود پیشنهاد الگوریتم SHA-3 در سال‌های اخیر [۱۵] و [۱۳]، خانواده SHA-2 هنوز هم یک سطح امنیتی کافی را برای ابزارها فراهم می‌کند.

به‌موازات پیشرفت تکنولوژی و بروز کاربردهای جدید از الگوریتم‌های درهم‌ساز ایمن، حجم وسیعی از تحقیقات حول پیاده‌سازی روی بسترهای FPGA به‌واسطه انعطاف‌پذیری آن‌ها انجام‌شده است. در ابتدا قالب این پیاده‌سازی‌ها با هدف کاهش تأخیر[۱۶] و افزایش سرعت اجرای الگوریتم به‌واسطه کاهش مسیرهای بحرانی [۱۷]، Unfolded [۱۵] و به کارگیری روش‌های موازی[۱۸] که از بروزترین تحقیقات در این راستا می‌توان به تابع هش شبکه موازی (PLHF) اشاره کرد[۱۹]، انجام گرفته است. از طرفی سابقه پیاده‌سازی الگوریتم‌های درهم‌ساز ایمن با رویکرد کاهش سطح[۲۳] و [۲۰] با به‌کارگیری روش‌های Folding و ساختار موازی مرحله پیش‌پردازش الگوریتم با هدف به‌کارگیری در کاربردهای مهمی همچون کاربردهای مبتنی بر وب (Web-based) [۲۴] و امنیت روترهای شبکه در پیشگیری از حمله غیرقابل پیش‌بینی هکرها [۲۵] انجام‌شده است. آنچه معماری جدید پردازنده را از کارهای قبلی متمایز می‌سازد به‌کارگیری واحدهای مسیر داده و حافظه چندپورته جهت فراهم‌سازی شرایطی برای عملکرد موازی الگوریتم SHA-2 و استفاده از ثبات‌هایی برای ذخیره موقت داده‌ها جهت کاهش میزان دسترسی به حافظه در قالب یک معماری ساده است که هدف از این طرح کاهش مساحت و افزایش سرعت پردازش داده‌ها به‌موازات حفظ حد مطلوبی از توان عملیاتی و کارایی در کاربردهای سیار است.

در ادامه مقاله در بخش ۲،الگوریتم SHA-2 به‌طور اجمالی تشریح شده و در بخش ۳، بلوک‌دیاگرام معماری پردازنده پیشنهادی همراه با ماژول‌های آن شرح داده‌شده است. بخش ۴، به ارزیابی طرح پیشنهادی و مقایسه نتایج حاصل از سنتز و پیاده‌سازی الگوریتم‌های رمزنگاری SHA-256 و SHA-512 پرداخته و در بخش ۵، نتیجه‌گیری کلی اعلام‌شده است.

۲-مفاهیم اولیه الگوریتم SHA-2

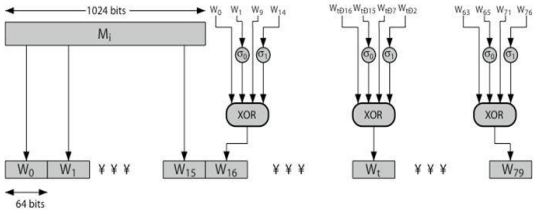
سری دوم الگوریتم‌های درهم‌ساز ایمن، مجموعه‌ای از توابع درهم‌ساز هستند که بر اساس طول چکیده (به تعداد بیت) در قالب‌های کلی به دو صورت SHA-256 و SHA-512 نام‌گذاری شده‌اند که محاسبات آن‌ها به ترتیب با کلمات ۳۲ و ۶۴ بیتی صورت می‌گیرد. این دو الگوریتم، ساختار نزدیک به هم دارند ولی در تعداد دورها و مقادیر اولیه به‌کاررفته متفاوت هستند که از بین آن‌ها استفاده از SHA-512 ازنظر میزان امنیت حاصله و سرعت آن روی سیستم‌های ۶۴ بیتی ارجحیت دارد. در شکل ۱ ساختار کلی الگوریتم SHA-512 نشان داده‌شده است که در ادامه عملیات اجرایی مربوط به شکل توضیح داده‌شده است.



شکل ۱: عملیات مربوط به الگوریتم رمزنگاری SHA-512 [۲۶]

۲-۱- مرحله پیش‌پردازش

طول پیام ورودی (Message) با افزودن عدد '۱' و به دنبال آن صفرهایی برای رسیدن به مضربی از ۵۱۲ در SHA-256 (۱۰۲۴ در SHA-512) افزایش می‌یابد. ۶۴ بیت آخر پیام در SHA-256 (۱۲۸ بیت آخر پیام در SHA-512) برای ذخیره طول پیام اصلی استفاده می‌شود. بعد از عمل افزونگی، پیام به‌دست‌آمده به بلوک‌هایی به طول ۵۱۲ بیتی برای SHA-256 (۱۰۲۴ بیتی برای SHA-512) به‌صورت $M1, M2, \ldots, M_N$ تقسیم می‌شود. هر بلوک از اطلاعات M_i به ترتیب با توابع اصلی *(F)* در طول ۶۴ دور برای SHA-256 و ۸۰ دور برای SHA-512 پردازش می‌شوند. تابع دور مربوط به الگوریتم SHA-512 در شکل ۲ نشان داده‌شده است.



شکل ۲: تابع دور الگوریتم SHA-512 [۲۶]

مقدار هش H_i از پردازش کامل بلوک اطلاعاتی M_i جاری به ترتیب ذیل به دست می‌آید.

۲-۱-۱-آماده‌سازی زمانبند پیام (W_t)

زمانبند پیام عبارت است از پیام‌هایی که از پیام اصلی جهت به‌کارگیری در عملیات مربوط به دورهای الگوریتم به دست می‌آیند؛ که در آن ۱۶ مقدار اول در ۱۶ دور اول مستقیماً از پیام ورودی حاصل می‌شود و برای بقیه دورها (دورهای ۱۷≤t≤۶۴ برای SHA-256 و دورهای ۱۷≤t≤۸۰ برای SHA-512) مقادیر Wt با استفاده از توابع σ0 و σ1 محاسبه می‌شوند[۱۲].

$$W_t = \begin{cases} M_t \\ \sigma_1(W_{t-2}) + W_{t-7} + \sigma_0(W_{t-15}) + W_{t-16} \end{cases} \quad (۱)$$

$$\sigma(x)_0^{256} = ROTR^7(x) \oplus ROTR^{18}(x) \oplus SHR^3(x) \quad (۲)$$

$$\sigma(x)_1^{256} = ROTR^{17}(x) \oplus ROTR^{19}(x) \oplus SHR^{10}(x) \quad (۳)$$

$$\sigma(x)_0^{512} = ROTR^1(x) \oplus ROTR^8(x) \oplus SHR^7(x) \quad (۴)$$

$$\sigma(x)_1^{512} = ROTR^{19}(x) \oplus ROTR^{61}(x) \oplus SHR^6(x) \quad (۵)$$

که در روابط فوق منظور از $ROTR^i(x)$، چرخش راست i واحدی عبارت x و منظور از عبارت $SHR^i(x)$، انتقال به راست i واحدی عبارت x است.

۲-۱-۲- مرحله فشرده‌سازی

در این مرحله حلقه اصلی پردازش بلوک اطلاعاتی Mi از ۸ متغیر h, g, f, e, d, c, b, a در قالب بافرهای حالت بوده و پس از طی شدن دورهای داخلی الگوریتم به‌عنوان مقدار هش میانی Htemp در نظر گرفته می‌شود. این مقدار هش با مقدار محاسبه‌شده برای بلوک اطلاعاتی قبلی جمع شده و

مقدار هش مربوط به پیام تا بلـوک اطلاعـات Mi را نتیجـه می‌دهد. وقتی بلوک اطلاعاتی اولیه پردازش می‌شـود، مقـدار هش اولیه H0 مورداستفاده قرار می‌گیرد. در اولین مرحله از محاسبه Hi+1 بایستی مقدار Hi وارد بافرهای حالت شوند. عملیات مربوط به این مرحله در رابطه (۶) نشان داده‌شده‌اند.

$$\begin{aligned} h &= g & ,g &= f \\ f &= e & ,e &= d + T_1 \\ d &= c & ,c &= b \\ b &= a & ,a &= T_1 + T_2 \end{aligned} \qquad (۶)$$

$$H_{temp} = a\,|\,b\,|\,c\,|\,d\,|\,e\,|\,f\,|\,g\,|\,h$$

$$H^{(i)} = H^{(i-1)} + H_{temp}$$

متغیرهای زمانی به‌کاررفته در الگوریتم در قالب رابطـه‌هـای (۷) و (۸) محاسبه می‌شوند[۱۲].

$$T_1 = h + \sum_1(e) + ch(e,f,g) + k_t + W_t \qquad (۷)$$

$$T_2 = \sum_0(a) + Maj(a,b,c) \qquad (۸)$$

مقادیر هش اولیه H_0 و ثابت K_t به‌کاررفتـه در روابـط فـوق، در مشخصـات و ویژگـی‌هـای توابـع SHA-256 و SHA-512 قابل‌دسترسی است.

تمامی عملیات درگیر در الگوریتم بر اسـاس روابـط (۹) الـی (۱۴) محاسبه می‌شوند [۲۶] و [۲۷].

$$ch(x,y,z) = (x \wedge y) \oplus (-x \wedge z) \qquad (۹)$$

$$Maj(x,y,z) = (x \wedge y) \oplus (x \wedge z) \oplus (y \wedge z) \qquad (۱۰)$$

$$\sum_0^{256}(x) = ROTR^2(x) \oplus ROTR^{13}(x) \oplus ROTR^{22}(x) \qquad (۱۱)$$

$$\sum_1^{256}(x) = ROTR^6(x) \oplus ROTR^{11}(x) \oplus ROTR^{25}(x) \qquad (۱۲)$$

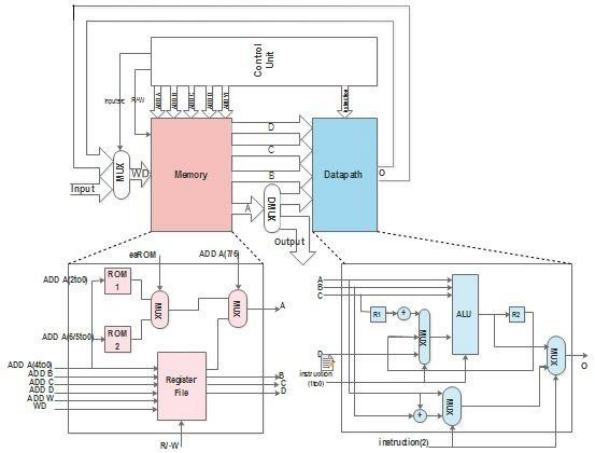
$$\sum_0^{512}(x) = ROTR^{28}(x) \oplus ROTR^{34}(x) \oplus ROTR^{39}(x) \qquad (۱۳)$$

$$\sum_1^{512}(x) = ROTR^{14}(x) \oplus ROTR^{18}(x) \oplus ROTR^{41}(x) \qquad (۱۴)$$

بعد از محاسبه بلوک اطلاعاتی نهایی M_N مقدار هش نهـایی H_N محاسبه می‌شود.

۳-پردازنده متراکم پیشنهادی

معماری پردازنده پیشنهادی از سه واحد مسیر داده، حافظـه و کنترل تشکیل‌شده است. ایده اصـلی معمـاری پیشـنهادی استفاده از مسـیر داده و حافظـه چندپورتـه جهـت عملکـرد موازی چند مرحلـه از الگـوریتم SHA-2، تغییـر در روابـط الگوریتم با حفظ ماهیت آن‌ها بـه‌منظـور کـاهش مسـیرهای بحرانی، استفاده از ثبات‌های موقت در واحد مسیر داده برای کم کردن میزان دسترسی به واحد حافظه و استفاده از بانک ثبات به‌منظـور اسـتفاده مجـدد از منـابع حافظـه و تسـریع دسترسی به اطلاعات است. همچنین امکـان نگاشـت واحـد محاسباتی و منطقی روی جداول جستجوی ۱۰ چهار ورودی در بسترهای FPGA، امکان پیاده‌سـازی طـرح پیشـنهادی روی این بسترها را تسهیل مـی‌کنـد. بلوکـدیاگرام پردازنـده هش در شکل ۳ نشان داده‌شده است.



شکل ۳: بلوک دیاگرام پردازنده هش پیشنهادی

۳-۱- ماژول‌های بکاررفته در پردازنده هش پیشنهادی

- مسیر داده

واحد مسیر داده همان‌طور کـه در شـکل ۱ نشـان داده‌شـده است از تعدادی ثبات به‌منظور ذخیره موقتی داده‌ها و از یک واحد محاسباتی و منطقی برای اجـرای عملیـات مربـوط بـه دورهای داخلی الگوریتم SHA-2 تشکیل‌شده است. در این پردازنده عملیات مربوط به دورهای داخلی الگـوریتم SHA-2 به‌صورت روابط T'،T و T'' (روابط ۱۶-۱۴) جهت کـاهش مسیرهای بحرانی بازنویسی شده‌اند.

$$T = \sum_1(e) + ch(e,f,g) + W_t \qquad ۱۵)$$

(

$$T' = h + k_t + d + T \qquad (۱۶)$$

$$T'' = \sum_0(a) + Maj(a,b,c) + (T' - d) \qquad ۱۷)$$

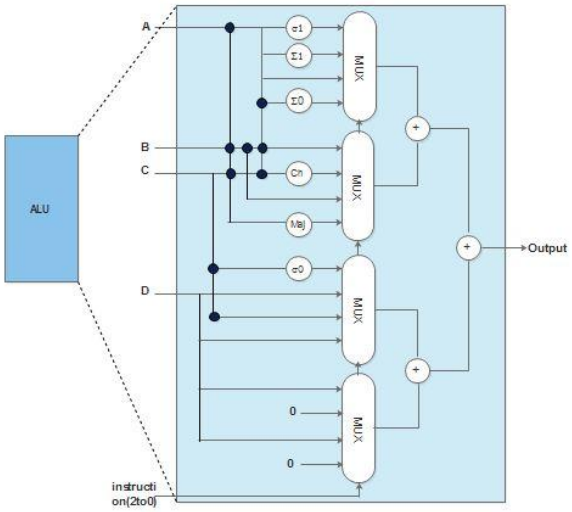
(

با توجه به ساختار الگوریتم و روابط بازنویسی شده می‌توان رابطه (۱۸) را نتیجه گرفت.

$$e = T' \qquad ,a = T'' \qquad ۱۸)$$

(

در شکل ۴ بلوک دیاگرام واحد محاسباتی و منطقی نشان داده‌شده است. ملاحظه می‌شود که به‌واسطه به‌کارگیری ثبات‌های موقت و بازنویسی روابط، میزان دسترسی به حافظه کم شده و از آدرس‌دهی غیرضروری به حافظه و عملیات منطقی مربوطه اجتناب شده است.



شکل ۴: بلوک دیاگرام واحد محاسباتی و منطقی

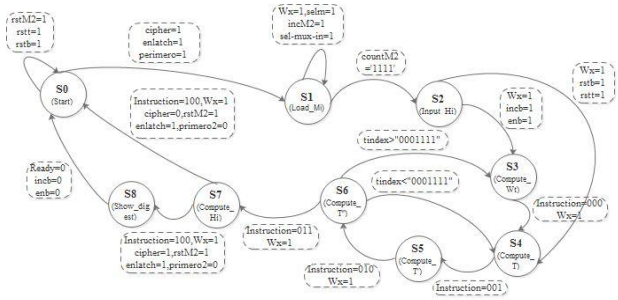
- واحد حافظه

واحد حافظه پردازنده پیشنهادی الگوریتم SHA-2 از یک بانک ثبات و دو حافظه فقط خواندنی (ROM) تشکیل‌شده است. بانک ثبات شامل ۳۲ ثبات ۳۲ بیتی برای SHA-256 و ۳۲ ثبات ۶۴ بیتی برای SHA-512است که برای خواندن اطلاعات از پیش ذخیره‌شده و نوشتن اطلاعات حاصل از عملیات به کار می‌رود. این اطلاعات شامل متغیرهای زنجیره‌ای، کلمات اولیه و پیام‌های زمان‌بندی محاسبه‌شده در دورهای مختلف توسط واحد مسیر داده هستند. حافظه فقط خواندنی ۸ خانه‌ای برای نگه‌داری متغیرهای زنجیره‌ای اولیه و ۶۴ خانه‌ای برای SHA-256 (۱۲۸ خانه‌ای برای SHA-512) به‌منظور ذخیره ثابت‌های دور K_t در واحد حافظه به کار گرفته می‌شوند.

- واحد کنترل

واحد کنترل از چند واحد شمارنده به‌منظورآدرس‌دهی بانک ثبات و ROMها تشکیل‌شده است؛ همچنین واحد کنترل با تولید تعدادی سیگنال کنترلی منابع اطلاعاتی واحد محاسباتی و منطقی را نیز هماهنگ می‌کند. در طول ۱۶ سیکل ساعت اول اطلاعات ورودی که شامل ۱۶ کلمه ۳۲ بیتی برای SHA-256 و ۱۶ کلمه ۶۴ بیتی برای SHA-512 هستند، جهت استفاده در توابع فشرده‌سازی به کار گرفته می‌شوند. پس از ورود کامل اطلاعات ۵۱۲ بیتی برای SHA-256و ۱۰۲۴ بیتی برای SHA-512 واحد کنترل مقادیر میانی Hi را به بافرهای حالت انتقال می‌دهد (ثبات‌های a-h). برای اولین بلوک مقادیر هش اولیه H_0 که در حافظه ثبات‌ها ذخیره‌شده به‌عنوان مقادیر هش میانی مورداستفاده قرار می‌گیرند و برای دورها و بلوک‌های اطلاعاتی بعدی، H_i توسط ثبات‌های a تا h و مقادیر هش به‌دست‌آمده از مرحله قبلی H_{i-1} محاسبه می‌شود.

واحد کنترل در پردازنده هش پیشنهادی یک FSM بوده که شامل ۹ حالت است. در شکل ۵ ماشین حالت محدود واحد کنترل نشان داده‌شده است.



شکل ۵: ماشین حالت محدود واحد کنترل پردازنده هش پیشنهادی

۴.ارزیابی نتایج

در این مقاله به‌منظور ارزیابی نتایج حاصل از پیاده‌سازی الگوریتم‌های هش SHA-256 و SHA-512 از یک معماری مشترک (شکل ۱) برای هرکدام از توابع استفاده‌شده است. پس از نوشتن معماری مذکور با زبان

توصیف سخت‌افزاری VHDL نتایج مربوط به متغیرها در هر دور با نرم‌افزار ISIM قابل‌دسترس بوده و سنتز و پیاده‌سازی آن برای تعدادی از FPGA های شرکت Xilinx توسط نرم‌افزار Xilinx ISE مورد بحث و بررسی قرار می‌گیرد.

FPGA های انتخابی جهت بررسی نتایج ذکرشده به ترتیب ذیل است:

۱- Virtex-4 (Xc4vlx100-12FF1148)

۲- Virtex-5 (Xc5vlx155t-3FF1136)

۳- Virtex-7 (Xc7vx330t-3FFG1157)

طرح VHDL معماری پردازنده پیشنهادی در هر مرحله (در هر دور از الگوریتم) ازنظر جریان داده و بررسی حالات هر متغیر به‌واسطه مقادیر به‌دست‌آمده از نتایج شبیه‌سازی توسط نرم‌افزار ISIM با نتایج ذکرشده در FIPS 180_2 [۱۱] مقایسه و اشکال‌زدایی شده و نهایتاً با این کار برای محاسبه مقادیر هش یک عبارت معتبرسازی می‌شود. در این مقاله کیفیت معماری پیشنهادی با معیارهای مساحت، فرکانس، توان عملیاتی و کارایی موردتوجه قرارگرفته است که پارامترهای مساحت و فرکانس مستقیماً از گزارش‌های حاصله از سنتز و پیاده‌سازی و توان عملیاتی و کارایی به‌صورت روابط (۱۹) و (۲۰) محاسبه می‌شوند[۲۹] و [۲۸].

(۱۹) توان عملیاتی =اندازه بلوک داده ×(زمان تأخیر/فرکانس)

(۲۰) کارایی=توان عملیاتی/مساحت

نتایج به‌دست‌آمده از سنتز و پیاده‌سازی الگوریتم‌های SHA-256 و SHA-512 برای پارامتر کارایی در جدول ۱ و برای پارامترهای فرکانس، مساحت و توان عملیاتی در جدول‌های ۲ و ۳ در مقایسه با کارهای پیشین نشان داده‌شده است.

جدول ۱: کارایی حاصله از پردازنده پیشنهادی

عنوان چیپ	کارایی(Mbps/Slice)
SHA-256 Virtex-4	۰/۲۷
SHA-256 Virtex-5	۱/۴۱
SHA-256 Virtex-7	۶/۹۸
SHA-512 Virtex-4	۰/۲۰
SHA-512 Virtex-5	۱/۱۷
SHA-512 Virtex-7	۳/۳۵

علی‌رغم دشوار بودن مقایسه پیاده‌سازی‌های مختلف روی FPGA ها به دلیل تکنولوژی‌های مختلف به‌کاررفته، در ادامه سعی بر مقایسه عادلانه نتایج حاصله از معماری پیشنهادی با کارهای پیشین در شرایط یکسان شده است.

۴-۱- نتایج به‌دست‌آمده و مقایسه‌های مربوط به الگوریتم SHA -256

تأخیر پردازنده SHA-256 برای یک بلوک داده ۵۱۲ بیتی ۲۸۰ سیکل ساعت است که به ترتیب ذیل در نظر گرفته می‌شود[۲۸]:

زمان لازم برای وارد شدن یک بلوک اطلاعاتی ۵۱۲ بیتی، ۱۶ سیکل زمانی است.

زمان لازم برای انتقال مقادیر هش میانی و بافرها (ثبات‌های a تا h) ۸ سیکل زمانی است.

زمان لازم جهت طی شدن ۶۴ دور داخلی با در نظرگرفتن دو حالت زیر محاسبه می‌شود:

برای ۱۶ دور اول، Wt مستقیماً از پیام ورودی نتیجه می‌شود که این مرحله فقط نیاز به ۳ سیکل زمانی دارد.

برای ۴۸ دور بعدی، Wt با استفاده از واحد محاسباتی و منطقی محاسبه‌شده و به ازای هر ۴ سیکل زمانی یک دور نتیجه می‌دهد.

بعد از محاسبات مربوط به دورهای داخلی، محاسبه مقادیر هش میانی نیاز به ۸ سیکل زمانی دارد.

نهایتاً زمان موردنیاز برای به‌دست آوردن خروجی (مقدار هش نهایی)۸ سیکل زمانی خواهد بود.

در جدول ۲ نتایج به‌دست‌آمده از معماری پیشنهادی برای SHA-256 با کارهای پیشین مقایسه شده است. مشاهده می‌شود که در مقایسه با معماری متراکم در [۳۰] علاوه بر کاهش ٪۳۰ استفاده از منابع، میزان تأخیر نیز به میزان قابل‌توجهی کاهش‌یافته است. نتایج حاصل از طرح متراکم در [۳۱] به دلیل پیاده‌سازی روی بستر دیگری از خانواده Virtex با معماری پیشنهادی قابل‌مقایسه نیست، ولی در حالت کلی امتیازهای این طرح از نتایج به‌دست‌آمده مشخص است. طرح پیشنهادی در مقایسه با معماری‌های [۳۲،۳۳] که با هدف کاهش تأخیر و افزایش توان عملیاتی

طراحی‌شده‌اند، توانسته به موازات کاهش قابل‌توجه مساحت به افزایش ٪۱۰ فرکانس نیز دست یابد. همچنین در مقایسه با [۲۸] علی‌رغم اشغال ٪۷ مساحت بیشتر در معماری پیشنهادی، ٪۲۵ افزایش فرکانس و به دنبال آن بهبود توان عملیاتی و کارایی در نتایج به‌دست‌آمده از پیاده‌سازی معماری پیشنهادی روی بستر Virtex-4 حاصل‌شده است.

جدول ۲: نتایج مقایسه حاصله از سنتز و پیاده‌سازی پردازنده SHA-256 پیشنهادی

منابع	بستر	مساحت (Slices)	فرکانس (MHZ)	تأخیر	توان عملیاتی MBPS)
[۳۰]	Virtex-4	۶۱۵	۱۰۲	۱۱۲۰	_
[۳۱]	Virtex-2	۱۲۱۰	۸۵	۳۵۵	۱۲۲/۶۰
[۳۲]	Virtex-5	۲۷۹۶	۱۷۹/۸	_	_
[۳۳]	Virtex-5	۱۸۸۵	۱۶۹	_	_
[۲۸]	Virtex-4 Virtex-5	۴۲۲ ۱۳۹	۵۰/۰۶ ۶۴/۴۵	۲۸۰ ۲۸۰	۹۱/۵۳ ۱۱۷/۸
طرح پیشنهادی	Virtex-4 Virtex-5 Virtex-7	۴۵۳ ۲۵۲ ۷۷	۶۷/۱۴۶ ۱۹۴/۹۲ ۲۹۳/۷۲	۲۸۰ ۲۸۰ ۲۸۰	۱۲۲/۷۸ ۳۵۶/۴۲ ۵۳۷/۰۹

۴-۲- نتایج به‌دست‌آمده و مقایسه‌های مربوط به الگوریتم SHA-512

تأخیر پردازنده SHA-512 برای یک بلوک داده ۱۰۲۴ بیتی ۳۴۴ سیکل ساعت است که به ترتیب ذیل در نظر گرفته می‌شود:

زمان لازم برای وارد شدن یک بلوک اطلاعاتی ۱۰۲۴ بیتی، ۱۶ سیکل زمانی است.

زمان لازم برای انتقال مقادیر هش میانی و بافرها (ثبات‌های a تا h) ۸ سیکل زمانی است.

زمان لازم جهت طی شدن ۸۰ دور داخلی با در نظرگرفتن دو حالت زیر محاسبه می‌شود:

برای ۱۶ دور اول، Wt مستقیماً از پیام ورودی نتیجه می‌شود که این مرحله فقط نیاز به ۳ سیکل زمانی دارد.

برای ۶۴ دور بعدی، Wt با استفاده از واحد محاسباتی و منطقی محاسبه‌شده و به ازای هر ۴ سیکل زمانی یک دور نتیجه می‌دهد.

بعد از محاسبات مربوط به دورهای داخلی، محاسبه مقادیر هش میانی نیاز به ۸ سیکل زمانی دارد.

نهایتاًزمان موردنیاز برای به‌دست آوردن خروجی (مقدار هش نهایی) ۸ سیکل زمانی است.

در جدول ۳ نتایج به‌دست‌آمده از معماری پیشنهادی برای SHA-512با کارهای پیشین مقایسه شده است. معماری متراکم پیشنهادی می‌تواند جایگزین مناسبی برای [۳۴] که با هدف افزایش توان عملیاتی طراحی و پیاده‌سازی شده، روی بستر Virtex-5 باشد. از نتایج به‌دست‌آمده در [۳۵] که در آن از روش موازی با رویکرد افزایش فرکانس استفاده‌شده، می‌شود که در طرح متراکم پیشنهادی علاوه بر بهبود ٪۲۰ فرکانس، از ٪۶۰ منابع کمتری روی Virtex-5 استفاده‌شده است. هسته هشت تجاری در [۳۶] برای اجرای الگوریتم SHA-512 از ۱/۷ برابر منابع کمتری نسبت به طرح پیشنهادی استفاده می‌کند ولی برای کاربردهایی که در آن‌ها نیاز به توان عملیاتی بالا وجود دارد، پیاده‌سازی طرح پیشنهادی روی بستر Virtex-5 به دلیل بهبود ۱۱ برابری توان عملیاتی ارجحیت دارد. مشاهده می‌شود که در معماری پیشنهادی کاهش ٪۶۸ پارامتر مساحت در سطح فرکانس مطلوب، نسبت به معماری [۳۷] روی بستر Virtex-4 حاصل‌شده است؛ همچنین در [۳۸] به‌کارگیری از روش‌های unrolling و pipeline، استفاده مجدد از منابع و پیش‌محاسبات زمانی با رویکرد افزایش فرکانس و توان عملیاتی انجام‌شده است؛ مشاهده می‌شود که معماری پیشنهادی پیاده‌سازی شده روی بستر Virtex-5 توانسته از٪۵۵ منابع کمتری در همان سطح فرکانسی استفاده کند.

جدول ۳: نتایج مقایسه حاصله از سنتز و پیاده‌سازی پردازنده SHA-512

منابع	بستر	مساحت (Slices)	فرکانس (MHz)	تأخیر	توان عملیاتی MBPS)(
[۳۴]	Virtex-2	۱۹۳۸	۸۱	_	۲۷۴
[۳۵]	Virtex-4 Virtex-5	۲۰۷۳ ۱۱۰۲	۱۰۶/۶۵ ۱۴۲/۸۸	_	_
[۳۶]	Hash Core	۲۵۱	۲۷۱	_	۴۶
[۳۷]	Virtex-4	۲۶۶۷	۸۱/۰۱	_	۱۰۲۴
[۳۸]	Virtex-6 Virtex-7	۹۸۶ ۱۰۲۱	۱۷۷/۹ ۱۹۸/۴	_ _	۴۵۵۴/۲ ۵۰۷۹/۱

طرح پیشنهادی	Virtex-4	۸۴۳	۵۸/۶۵۶	۳۴۴	۱۷۴/۶۰
	Virtex-5	۴۳۹	۱۷۳/۸۱۸	۳۴۴	۵۱۷/۴۱
	Virtex-7	۱۸۱	۲۰۳/۷۱۵	۳۴۴	۶۰۶/۴۱

با بررسی دقیق جدول‌های ۲ و ۳ و همچنین مطالعه دقیق آنچه که به‌تفصیل توضیح داده شد؛می‌توان نتیجه گرفت که با استفاده ازمعماری پیشنهادی به‌واسطه بکارگیری برخی راهکارها در راستای به حداقل رساندن ارجاع به حافظه در پروسه پردازشی و استفاده از روش‌های عملکرد موازی، توانسته در ضمن استفاده از حداقل منابع که منجر به کاهش مساحت می‌شود، از سرعت عملیاتی و فرکانس بالایی نسبت به تحقیقاتی با رویکردهای مشابه بهره برد.

۵- نتیجه‌گیری

در این مقاله معماری پیشنهادی پردازنده مبتنی بر FPGA برای الگوریتم‌های رمزنگاری سری SHA-2 معرفی‌شده است. این معماری به‌واسطه بکارگیری از واحدهای مسیرداده و حافظه چندپورته، استفاده از ثبات‌های موقت در واحد مسیر داده، استفاده از بانک ثبات به‌جای حافظه‌های RAM جهت ذخیره داده‌ها در واحد حافظه باعث عملکرد موازی و کاهش دفعات ارجاع به حافظه می‌شود؛ که نتیجه آن‌ها یک معماری متراکم ساده و یکپارچه با سرعت بالا است.

نتایج به‌دست‌آمده از پیاده‌سازی طرح پیشنهادی حاکی از آن است که پردازنده متراکم پیشنهادی در مقایسه با کارهای پیشین با اهداف مشابه، توانسته با افزایش فرکانس کاری ٪۲۵ برای SHA-256 و اشغال ٪۵۵ مساحت کمتر برای SHA-512 حد مطلوبی از توان عملیاتی و کارایی را نیز حفظ نماید. پردازنده SHA پیشنهادی برای کاربردهایی مانند بسترهای سیار مورد اعتماد (TMP)، واحد پول دیجیتال (Bitcoin) و مسیریابی ایمن در شبکه‌های روی تراشه‌ها (NoC) مناسب است.

با توجه به مزیت‌های طرح پیشنهادی می‌توان از ساختار ساده معماری پیشنهادی برای طراحی پردازنده جهت اجرای عملیات مربوط به الگوریتم SHA-3 نیز در راستای حفظ محرمانگی اطلاعات در تکنولوژی روز با کاربردهایی که در آن‌ها نرخ انتقال اطلاعات زیاد است، استفاده کرد.

منابع

۱. ذاکر حسینی ، ملکیان، امنیت داده‌ها، ویراسته ۱. باباخانی، ویرایش دوم، تهران، موسسه علمی-فرهنگی نص، ۱۳۸۷..

2.E. Kurniawan & I. Riadi, "Security Level analysis Of academic information systems based on standard ISO 27002:2003 using SSE-CMM'',vol. 16,no. 1, pp. 139-147,2018.

3.I. Riadi, E. I. Aristianto & A. Dahlan, "An Analysis of Vulnerability Web Against Attack Unrestricted Image File Upload'', Comput. Eng. Appl., vol. 5, no.1, PP. 19-28,2016.

4.H. Kim, M. Lee, D. K. Kim, S. K. Chung&K. Chung, "Design and implementation of crypto co-processor and its application to security systems", BerlinHeidelberg: Springer, Computational Intelligence and Security. Lecture Notes in Computer Science, vol. 3802, PP. 1104–9, 2005.

5.S. Reddy, R. Sakthivel&P. Praneet, "VLSI implementation of AES crypto processor for high throughput," In: (IJAEST) International Journal of Advanced Engineering Sciences and Technologies, vol. 6, PP. 2–6, 2011.

۶. دری، قیاسیان، سعیدی. «طراحی و پیاده‌سازی رمزنگار AES در بستر FPGA برای خطوط پرسرعت»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۶، شماره ۱، بهار ۱۳۹۵.

7.R. Glabb, L. Imbert, G. Jullien, A. Tisserand and N. Veyrat-Charvillon, " Multi-mode operator for SHA-2 hash functions",JSystArchit 53(2-3):127-38, 2007.

8.N. Sklavos& O. Koufopavlou, "On the hardware implementations of the SHA-2 (256,384,512) hash functions", In Proceeding of the 2003 International Symposium on Circuits and Systems, ISCAS'03.vol.5;2003.p.V-153-V-156, 2003.

9.K. Ting, S. Yuen,K.Lee and P. Leong, "An FPGA based SHA -256 processor", In: Glesner M, Zipf P, Renovell M, editors. Field-Programmable Logic and Applications: Reconfigurable Computing is Going Mainstream, Lecture notes in Computer Science, vol. 2438. Berlin/Heidelberg: Springer, p. 449–71, 2002.

10.http:// www.en.wikipedia.org/wiki/securehashalgorithm.

11.U.S. Department of Commerce National Technical Information Service,*FIPS180-2–Secure Hash Standard*,http://www.csrc.nist.gov/publications/fips/fips180-2/fips180-2, 2002.

12.N. Sklavosand & O.KouFOPAVLOU, "Implementation of the SHA-2 hash family standard using FPGAs", Springer Science+Business Media, Inc. Manufactured in the Netherlands, The Journal of supercomputing, vol. 31, pp. 227-248, 2005.

13.A. Regenscheid, R. Perlner, S. Chang, J. Kelsey,M. NandiandandS. Paul, Status Report on the First Round of the SHA-3 Cryptographic Hash Complete Competition, NIST, 2009.

14.E. Andreeva,B. Menninkand and B. Preneel, "Security reductions of the second round SHA_3 candidates," In: Proceedings of the 13thInternational Conference on Information Security, ISC'10. Berlin, Heidelberg: Springer-Verlag, PP. 39–53, 2011.

15.P. Kotewar, R. Mandavganeand & D. Khatri, "Review on area optimization and simulation of SHA-3,"Journal of Emerging Technologies and Innovative Research (JETIR), PP.290-292, 2014.

16.F. Crowe, A. Daly, T. Kerins and W.Marnane, "Single-chip FPGA implementation of acryptographic co-processor", In: Proceedings: IEEE International Conference on Field-Programmable Technology(IEEE Cat. No.04EX921), PP. 279–285,2005.

17.R. Lien, T. Grembowski and K. Gaj,A1Gbit/s Partially Unrolled Architecture of Hash Function SHA-1 & SHA-512, In: Topics in Cryptography a CT-RSA, pp.1995-1999, 2004.

18.A.P. Kakarountas,H. Michail,A. Milidonis, C.E. Goutis &G. Theodoridis, "High-speed FPGA implementation of secure hash algorithm for IPSec and VPN applications", Journal of Supercomputing, vol. 37, PP.179-195, 2006.

19.Y. Yang, F. Chen, Z. Sun, S. Wang, J. Li, J. Chen &Z. Ming, "Secure and efficient parallel hash function construction and its application on cloud audit", Soft Computing, vol. 23, pp. 8907-8925, 2019.

20.Y. K. Lee, H. Chan and I. Verbauwhede,"Iteration bound analysis and throughput optimum architecture of SHA-256(384,512) for hardware implementations", In:Proceedings of the 8th International Conference on Information Security Applications, vol 256, PP.102–114, 2007.

21.M. Kim,J. Ryouand &S. Jun, "Efficient hardware architecture of SHA-256 algorithm for trusted mobile computing," Information Security &Cryptology, Lecture Notes in Computer Science, Berlin Heidelberg: Springer, vol.5487, p. 240–52, 2009.

22.R. Chaves, G. Kuzmanor, L. Sousaand & S. Vassiliadis, "Cost efficient SHA hardware accelerators", IEEE Transaction on Very Large Scale Integration Systems, Vol.16,NO.8, PP.999-1008, 2008.

23.G. Feng, P. Jainand & K. Choi, "Ultra-low power and high speed design and implementation of AES and SHA1 hardware cores in 65 nanometer CMOS technology", In Electro/Information Technology,IEEE International Conference, PP.405-410, 2009.

24.M. Sumagita&I. Riadi, "Analysis of Secure Hash Algorithm (SHA) 512 for Encryption Process on Web Based Application", IJCSIS Int. J. Dig. Foren. Cyb. Secur., vol. 7, no. 4, pp. 373-381, 2018.
25.M.I.Mazdadi, I. Riadi&A. Luthfi, "Live Forensics on RouterOS API Services to Investigate Network Attacks", Int. J. Comput. Sci. Inf. Secur., Vol. 15, no. 2, pp.406-410, 2017.
26.W. Stallings, Cryptography and Network Security Principles and Practice, Fifth Edition, Pearson Education, Inc., Publishing as Prentice Hall, 2011.
27.A.L. Barkatullah& T. GailaniCelebi, Design and FPGA Implementation of Hash Processor, Master of Science thesis, Middle East Technical University, 2007.
28.R. Garcia, I. Algredo-Badillo, M.,Morales-Sandoval, C.Feregrino-Uribeand and R. Cumplido, "A compact FPGA-based processor for the secure hash algorithm SHA-256",Elsevier, Computers and Electrical Engineering, vol. 40, pp. 194-202, 2014.
29.H. E. Michail, G.S.Athanasiou,G. Theodoridisand & C. E. Goutis, "On the development of high-throuput and area-efficient multi-mode cryptographic hash designs in FPGA", Elsevier,Integration,The VLSI Journal, vol. 47, pp.387-407, 2014.
30.X. Cao,L.Luand & M. O'Neill, "A compact SHA-256 architecture for RFID tag*,"*,In: Proceedings of the 22nd IET irishSignals and Systems Conference, ISSC, Trinity CollegeDublin, 2011.
31.M. Kim, D. Lee and J. Ryou, "Compact and Unified Hardware Architecture for SHA-1 and SHA-256 of Trusted Mobile Computing", PersUbiquitComput 2012:1–12,http://www.dx.doi.org/10.1007/s00779-012-0543.
32.C. Jeong& Y. Kim, "Implementation of efficient SHA-256 hash algorithm forsecure vehicle communication using FPGA*"*, IEEE, ISOCC2014, PP. 224-226,2014.
33.H. Michail,"On the exploitation of a high-throughput SHA-256 FPGA design for HMAC, *"*ACM Trans on Reconfigurable Tech. and Sys.,vol. 5 no. 1,pp. 1- 28, 2012.
34.M.Zeghid, B. Bouallegue, A. Baganne, M. Machhoutand &R. Tourki,**"**A reconfigurable implementation of the new secure hash algorithm*",*Proc Second Int. Conf. Availability, Reliability and Security, (ARES2007), 10–13 April 2007.
35.P. Zawleski, M. Lukowiakand &S. Radziszowsk, *Case Study on FPGA Performance of Parallel Hash Function*,PrzegladElectrotechniczny/ElectricalReview, PP. 151-155,2010.
36.H. Technology, Efficient Tiny Hash Core Family for Xilinx FPGA Datasheet, Helion Technology Limited, 2010.
37.I. Algredo-Badillo,M. Morales-Sandoval, C. Feregrino-Uribe and R. Cumplido,"Throughput and efficiency analysis of unrolled hardware architectures for the SHA-512 hash algorithm",IEEE Computer Society Annual Symposium on VLSI, PP. 63- 68, 2012.
38.G. Athanasiou, H. Michail, G. Theodoridis and C. Goutis,"Optimising the SHA-512 cryptographic hash function on FPGAs*",* Published in IET Computers & DigitalTechniques, PP.70-83,2013.

سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷
صص: ۴۵- ۷۰

فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

یادگیری رتبه‌بندی محتوای فارسی وب بر مبنای برنامه‌نویسی ژنتیک چند لایه

امیر حسین کیهانی پور
استادیار، گروه مهندسی کامپیوتر، دانشکده مهندسی، پردیس فارابی، دانشگاه تهران

تاریخ دریافت:۱۳۹۸/۱۲/۱۵ تاریخ پذیرش: ۱۳۹۹/۰۳/۲۱

چکیده

یادگیری رتبه‌بندی، یک رویکرد نو ظهور به منظور رفع چالش‌های موجود و بهبود عملکرد جویشگرهای وب، بسیار امید بخش و کارآمد است. در عین حال عدم توجه جدی به سوابق تعاملات کاربران با جویشگر طی فرآیند جستجو و ارزیابی نتایج بدست آمده، یکی از معضلات جدی آن بشمار می‌رود. در عین حال حجم بسیار زیاد ویژگی‌های مورد نیاز از اسناد و پرس‌وجوهای کاربران نیز کاربردی بودن این رویکرد را در شرایط واقعی با ابهام مواجه ساخته است. استفاده از مدل اطلاعات کلیک از گذر داده‌ها و تولید ویژگی‌های کلیک از گذر داده، راهکار نوینی است که بر مبنای آن و با بکارگیری مدل برنامه‌نویسی ژنتیک چند لایه، مدل رتبه‌بندی مناسبی تحت عنوان MGP-Rank برای بازیابی اطلاعات انگلیسی وب، عرضه شده است. در این پژوهش این، با عنایت به ویژگی‌های خاص زبان فارسی، از طریق ارائه سناریوهای مناسب برای ایجاد ویژگی‌های کلیک از گذر داده این الگوریتم، این الگوریتم بومی‌سازی شده است. نتایج حاصل از ارزیابی عملکرد این الگوریتم در حوزه زبان فارسی با استفاده از مجموعه داده dotIR، حاکی از توانمندی قابل ملاحظه آن نسبت به روش‌های مرجع رتبه‌بندی اطلاعات است. این بهبود عملکرد، بخصوص در بخش ابتدایی فهرست نتایج جستجو که غالباً بیشتر مورد مراجعه کاربران است، قابل توجه است.

واژه‌های کلیدی: یادگیری رتبه‌بندی، مدل برنامه‌نویسی ژنتیک چند لایه، ویژگی‌های کلیک از گذر داده، محتوای فارسی وب، مجموعه داده dotIR

۱- مقدمه

استفاده گسترده کاربران محیط وب از جویشگرها به عنوان یکی از مهم‌ترین و پرکاربردترین سامانه‌های موجود در این محیط، موجب گردیده است تا پژوهش‌های زیادی به منظور تقویت و بهبود فرآیند جستجو و رتبه‌بندی اسناد مرتبط با پرس‌وجوهای کاربران، صورت گیرد. گرچه تحقیقات در زمینه انجام رتبه‌بندی مناسب نتایج جستجوهای کاربران جویشگرها، پیشینه قابل توجهی دارد، ولی در سال‌های اخیر با کاربرد گسترده‌تر این سامانه‌ها توسط کاربران، حجم قابل توجهی از اطلاعات مربوط به سوابق تعاملات کاربران با این سامانه‌ها، گردآوری شده است و این امر، امکان استفاده از الگوریتم‌های یادگیری ماشینی را برای ایجاد مدل‌های رتبه‌بندی کارآمد، فراهم نموده است. بطور مشخص،

نویسندهٔ عهده‌دار مکاتبات: امیر حسین کیهانی پور keyhanipour@ut.ac.ir

یادگیری رتبه‌بندی، به مجموعه روش‌های یادگیری ماشینی اطلاق می‌شود که در عملیات رتبه‌بندی، به انجام یادگیری و تشخیص مدل مناسب جهت رتبه‌بندی اطلاعات، می‌پردازند [۲۸]. یادگیری رتبه‌بندی، کاربردهای متعددی در زمینه بازیابی اطلاعات[۱]، پردازش زبان طبیعی[۲] و نیز داده‌کاوی[۳] دارد. در تحقیقات انجام شده در این خصوص، به این نکته توجه شده است که کاربران جویشگرها که بطور مستقیم با این سامانه‌ها درگیر هستند، قادر به ارایه بهترین قضاوت‌ها در مورد کیفیت عملکرد این سامانه‌ها می‌باشند. با این وجود، عدم تمایل اکثریت کاربران به ارایه قضاوت صریح از عملکرد جویشگرهای مورد استفاده، طراحان این سامانه‌ها را بر آن داشته است تا از سوابق تعاملات کاربران طی جستجوهای بعمل آمده، که اصطلاحاً «اطلاعات کلیک از گذر داده[۴]»، نامیده می‌شود، به منظور بررسی و بهبود عملکرد این سامانه‌ها بهره گیرند [۲۱]. مطالعات انجام شده، نشان می‌دهد که استفاده از اطلاعات کلیک از گذر داده، می‌تواند منجر به بهبود عملکرد جویشگرهای وب، شود [۳۰] و [۲۲] و [۳۱]. علیرغم این موضوع، فقدان اطلاعات کلیک از گذر داده در اغلب مجموعه‌های داده محک موجود در زمینه یادگیری رتبه‌بندی، موجب شده است تا اکثر الگوریتم‌های ارایه شده در این حوزه، از این اطلاعات ارزشمند، بی‌بهره باشند. در عین حال، استفاده اغلب روش‌های موجود از تعداد زیادی از ویژگی‌های اسناد و پرس‌وجوها، کاربرد آنها را در شرایط واقعی، دشوار می‌سازد.

در ادامه این نوشتار، ابتدا به مرور تحقیقات مرتبط پرداخته خواهد شد و بصورت تحلیلی، ضرورت انجام این پژوهش، مطرح خواهد شد. پس از آن، معرفی الگوریتم MGP-Rank صورت خواهد گرفت و نحوه عملکرد آن بیان خواهد شد. سپس نتایج حاصل از ارزیابی عملکرد الگوریتم یادگیری رتبه‌بندی MGP-Rank در مقایسه با روش‌های رتبه‌بندی مرجع، بر اساس شاخص‌های ارزیابی فرآیند بازیابی اطلاعات، بصورت تحلیل مورد قیاس قرار خواهد گرفت. بخش پایانی این نوشتار نیز به جمع‌بندی این پژوهش و نیز معرفی حوزه‌های تحقیقاتی آتی، اختصاص یافته است.

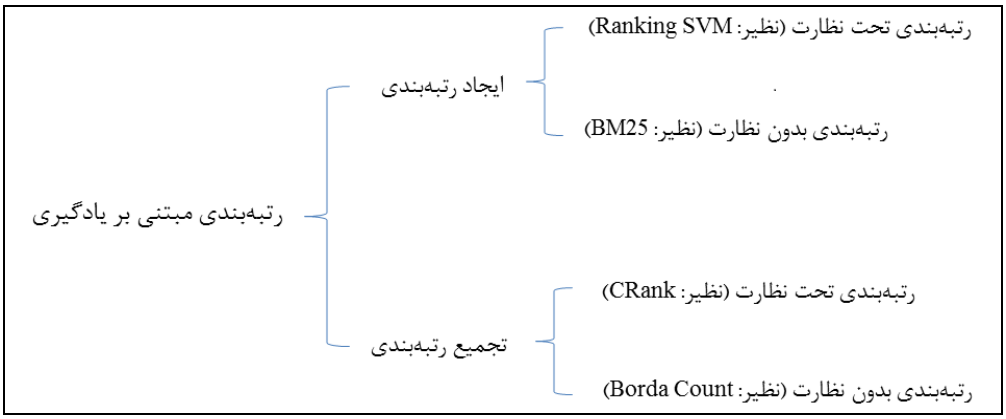
۲- مرور تحقیقات مرتبط

با توجه به نقش محوری رتبه‌بندی نتایج حاصل از جستجو در جویشگرهای وب، در سالهای گذشته تحقیقات گسترده‌ای در این در این زمینه صورت پذیرفته است. بصورت ویژه طی سالیان اخیر، موضوع یادگیری رتبه‌بندی، به عنوان یک روش نوین برای انجام رتبه‌بندی با استفاده از تکنیک‌های یادگیری ماشینی، مطرح شده است و در واقع، یادگیری رتبه‌بندی، به صورت ترکیبی از یادگیری ماشینی، بازیابی اطلاعات و نیز پردازش زبان طبیعی قابل توصیف است. بطور کلی، دو تعریف عمده برای مفهوم یادگیری رتبه‌بندی، ارائه شده است. در نگرش کلان، یادگیری رتبه‌بندی به هر نوع کاربرد تکنیک‌های یادگیری ماشینی در مساله رتبه‌بندی، اطلاق می‌شود. در مقابل و در یک نگرش جزیی‌تر، یادگیری رتبه‌بندی به آن دسته از روش‌های یادگیری ماشینی اشاره دارد که برای ایجاد مدل‌های رتبه‌بندی در مسایل «ایجاد رتبه‌بندی[۵]» و «تجمیع رتبه‌بندی[۶]»، مورد استفاده قرار می‌گیرند. در [۲۸]، یک دسته‌بندی کلی از انواع روش‌های مطرح در زمینه یادگیری رتبه‌بندی، در شکل شماره یک ارائه شده است.

در هر دو دسته تکنیک‌های یادگیری رتبه‌بندی را می‌توان جزء روش‌های یادگیری ماشینی تحت نظارت، بشمار آورد. بر این اساس، دادگان پایه یادگیری مشتمل بر مجموعه‌ای از اسناد و پرس‌وجوها و نیز قضاوت‌های کاربران خبره در مورد با میزان مرتبط بودن این اسناد و پرس‌وجوها در اختیار

[1] Information retrieval
[2] Natural Language Processing (NLP)
[3] Data mining
[4] Click-through data
[5] Rank Creation
[6] Rank Aggregation

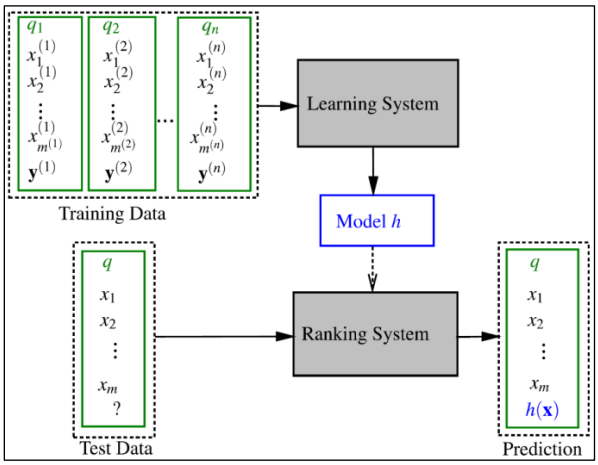
است. تولید مدل بر اساس قضاوت کاربران در رابطه با میزان مرتبط بودن زوج‌های اسناد - پرس‌وجوها صورت می‌گیرد. در روش‌های ایجاد رتبه‌بندی، بر اساس مجموعه داده یادگیری، مدلی به منظور رتبه‌بندی نتایج حاصل از پرس‌وجوهای کاربران، تولید می‌شود. در مقابل، در روش‌های تجمیع رتبه‌بندی، نتایج حاصل از رتبه‌بندی تعدادی از روش‌های رتبه‌بندی پایه، جهت تولید یک مدلی تجمیعی، با یکدیگر ادغام می‌شوند. بصورت خلاصه، می‌توان گفت در فرآیند «ایجاد رتبه‌بندی»، تهیه لیست مرتب، بر اساس ویژگی‌های درخواست و پیشنهادهای ممکن، صورت می‌گیرد، در حالی که در عملیات «تجمیع رتبه‌بندی»، از رتبه‌بندی‌های اولیه صورت گرفته، استفاده می‌شود. معمولاً هرگاه به مساله یادگیری رتبه‌بندی اشاره شود، منظور، ایجاد رتبه‌بندی بر اساس یادگیری تحت نظارت است [۲۸]. الگوریتم ارایه شده در این پژوهش یعنی MGP-Rank نیز از همین دسته روش‌های یادگیری رتبه‌بندی، بشمار می‌رود.



شکل ۱: طبقه‌بندی انواع روش‌های یادگیری رتبه‌بندی (Liu 2011)

شمای کلی عملیات روش‌های یادگیری رتبه‌بندی، در شکل شماره دو نشان داده شده است. بر اساس این ساختار، از آنجا یادگیری رتبه‌بندی، گونه‌ای از یادگیری تحت نظارت می‌باشد، نیاز به مجموعه داده آموزشی دارد. فرآیند تولید مجموعه داده آموزشی، شباهت زیادی با ایجاد داده آموزشی برای ارزیابی دارد. یک مجموعه داده آموزشی، عموماً شامل موارد زیر است:

- تعداد n پرس‌وجوی آموزشی $q_i (i = 1, \dots, n)$،
- مجموعه اسناد متناظر با آنها که به صورت بردار ویژگی‌ها نمایش داده می‌شوند $x^{(i)} = \left\{x_j^{(i)}\right\}_{j=1}^{m^{(i)}}$ (که در آن، $m^{(i)}$ برابر با تعداد اسناد متناظر پرس وجوی q_i است)، و
- قضاوت‌های میزان مرتبط بودن اسناد با پرس‌وجوها.



شکل ۲: شمای کلی چارچوب روش‌های رتبه‌بندی مبتنی بر یادگیری (Liu 2011)

بر این اساس، یک الگوریتم یادگیری به منظور تولید مدل رتبه‌بندی، استفاده می‌شود، به نحوی که این مدل، قادر باشد با توجه به تابع زیان[۷] مورد استفاده، میزان مرتبط بودن اسناد به پرس‌وجوها را با دقت بالا و حتی‌المقدور مشابه قضاوت‌های انسانی، پیش‌بینی نماید. در خلال فاز آزمون و با دریافت یک پرس‌وجوی جدید، مدل تولید شده طی فاز آموزش، روی مجموعه اسناد مورد نظر، اِعمال می‌شود تا فهرست رتبه‌بندی شده از آن اسناد، ایجاد شود. بر اساس این رویکرد، روش‌های یادگیری رتبه‌بندی را می‌توان به روش‌های نقطه‌ای[۸]، روش‌های جفتی[۹] و نیز روش‌های مبتنی بر فهرست[۱۰]، دسته‌بندی نمود. هر دسته از این روش‌ها، فرآیند یادگیری رتبه‌بندی را به اَشکال مختلف، مدل‌سازی می‌کنند. به عبارت دیگر، هر کدام از این انواع روش‌ها، تعاریف متفاوتی از فضاهای ورودی و خروجی بیان می‌کنند یا ممکن است از فرضیه‌های متفاوتی استفاده کنند و نیز توابع زیان متفاوتی را بکار گیرند.

بطور کلی، روش‌های نقطه‌ای، وابستگی بین اسناد را در نظر نمی‌گیرند و بنابراین موقعیت یک سند در لیست مرتب نهایی، در محاسبات، لحاظ نمی‌شود. فضای ورودی در روش‌های جفتی، زوج‌های اسناد می‌باشد که به‌صورت بردارهای ویژگی‌ها، نمایش داده می‌شوند. فضای خروجی نیز شامل اولویت بین این زوج‌های اسناد است که دامنه مقادیر آن از مجموعه $\{+1,-1\}$ می‌باشد. در برخی از روش‌های ارائه شده در این دسته، رتبه‌بندی به مساله رگرسیون، تقلیل می‌یابد [۱۰] و [۱۵] و [۴۵]. یک رویکرد دیگر، مدل‌سازی مساله رتبه‌بندی بصورت یک مساله رده‌بندی است. در تحقیقات مختلف، امکان استفاده از مدل رده‌بندی برای مساله رتبه‌بندی در حوزه بازیابی اطلاعات، بررسی شده است که از جمله آنها می‌توان به [۳۳] و [۴۴] و [۲۷]. اشاره نمود. در این پژوهش‌ها، ایده تبدیل رتبه‌بندی به مساله رده‌بندی دودویی مطرح شده است. در تحقیقات دیگری نظیر [۲۵] و [۳]. بکارگیری رده‌بندی چند کلاسه برای حل مساله رتبه‌بندی، پیشنهاد شده است. بکارگیری مدل رگرسیون ترتیبی، در دسته سوم روش‌های نقطه‌ای، دنبال شده است که به عنوان نمونه می‌توان به [۸] و [۹] اشاره نمود. با توجه به اینکه اغلب شاخص‌های ارزیابی در حوزه بازیابی اطلاعات مبتنی بر موقعیت اسناد و در سطح پرس‌وجوها محاسبه می‌شوند، روش‌های نقطه‌ای در شرایط واقعی، دچار محدودیت‌های جدی هستند.

فضای فرضیه در روش‌های جفتی یادگیری رتبه‌بندی، شامل توابع دو متغیره‌ای است که یک جفت سند را به عنوان ورودی، دریافت می‌کنند و ترتیب نسبی آن دو را در قالب خروجی، ارائه می‌دهند. ضمناً تابع زیان مورد استفاده در روش‌های جفتی، صرفاً ترتیب نسبی بین زوج‌های اسناد را در نظر می‌گیرد. در نتیجه، موقعیت نهایی اسناد در فهرست مرتب نهایی، به سختی قابل استحصال خواهد بود. از جمله ایده‌های مطرح شده در روش‌های جفتی رتبه بندی یادگیری، می‌توان به استفاده از شبکه‌های عصبی مصنوعی [۳۸] و [۱۶] و [۵۲] و [۲۴]. استفاده از مدل ماشین بردار پشتیبان[۱۱] [۲۱] و [۱۷]. و نیز بکارگیری الگوریتم‌های یادگیری ماشینی [۴] و [۵۱] و [۱۱] و [۲۰] و [۲۹]. اشاره نمود. ذکر این نکته در خصوص روش‌های جفتی، ضرورت دارد که با توجه به اینکه اغلب شاخص‌های ارزیابی در حوزه بازیابی اطلاعات، در سطح پرس‌وجوها و بر مبنای موقعیت نهایی اسناد در فهرست مرتب‌شده عمل می‌کنند، فاصله قابل توجهی، بین مدل عملکرد روش‌های جفتی و رتبه‌بندی مطلوب در مقوله بازیابی اطلاعات، مشاهده می‌شود.

فضای ورودی در روش‌های مبتنی بر فهرست، شامل مجموعه اسناد متناظر با پرس‌وجوی داده شده است. فضای

[7] Loss Function
[8] Pointwise
[9] Pairwise
[10] Listwise
[11] Support Vector Machine (SVM)

خروجی نیز عبارتست از فهرست مرتب‌شده اسناد یا به تعبیر دیگر، جایگشتی از مجموعه اسناد داده شده است. این دسته از الگوریتم‌های یادگیری رتبه‌بندی را بر اساس نوع تابع زیان مورد استفاده، به دو دسته، تقسیم نمود. دسته نخست، روش‌های بهینه‌سازی مستقیم[12] نامیده می‌شوند و شامل روش‌های رتبه‌بندی هستند که توابع زیان آنها، ارتباط مستقیمی با شاخص‌های استاندارد ارزیابی در حوزه بازیابی اطلاعات دارد [۴۰] و [۵۰] و [۴۶] و [۵۴] و [۵] و [۳۶] و [۷] و [۴۹]. در برخی دیگر از روش‌های اخیر، از رویکرد یادگیری عمیق[13] برای تخمین شاخص‌های ارزیابی استفاده شده است [۱] و [۳۷] و [۳۴] و [۳۹]. همچنین ایده استفاده از قابلیت‌های الگوریتم ژنتیک، به منظور بهینه‌سازی شاخص‌های ارزیابی، در برخی تحقیقات مورد توجه بوده است که از جمله آنها می‌توان به این موارد اشاره کرد: [۱۳] و [۴۱] و [۱۴] و [۴۸] و [۲۳]. در مقابل، در دسته دوم روش‌های فهرستی، الگوریتم‌های رتبه‌بندی قرار می‌گیرند که در آنها توابع زیان مورد استفاده، ارتباط مستقیمی با شاخص‌های استاندارد ارزیابی جستجو، ندارد. از جمله این روش‌ها می‌توان به [۴] و [۴۳] و [۳۴] و [۵۳] و [۴۴] اشاره نمود.

علیرغم مطالعات گسترده صورت گرفته در حوزه یادگیری رتبه‌بندی، این حوزه تحقیقاتی، با نواقص و محدودیت‌های جدی، همراه است. از مهم‌ترین این معضلات، عدم استفاده از اطلاعات مربوط به سوابق رفتارهای کاربران در حین انجام جستجوها، در فرآیندهای بازیابی و رتبه‌بندی نتایج است. در واقع، عدم اقبال کاربران برای ارایه بازخورد صریح از نحوه عملکرد جویشگر مورد استفاده، محققان را بر آن داشته است تا با پردازش سوابق رفتار کاربران، بازخورد ضمنی رفتار آنها را استخراج کنند. بنا بر نتایج تحقیقات انجام شده، اطلاعات حاصل از پردازش سوابق تعاملات و رفتارهای کاربران، که اصطلاحاً اطلاعات کلیک از گذر داده، نامیده می‌شود، بسیار ارزشمند بوده و بکارگیری آنها می‌تواند منجر به بهبود کیفیت فرآیند بازیابی اطلاعات، به میزان قابل توجهی شود [۲۲] و [۳۰] و [۳۱]. فقدان این قبیل اطلاعات، در اغلب مجموعه‌های داده محک[14] موجود در زمینه یادگیری رتبه‌بندی، باعث گردیده است تا اغلب روش‌های یادگیری رتبه‌بندی موجود، امکان استفاده از این قبیل اطلاعات را نداشته باشند که این امر، یکی از معضلات و محدودیت‌های اصلی پیش‌روی حوزه یادگیری رتبه‌بندی، بشمار می‌رود. در عین حال، عرضه انبوهی از ویژگی‌های مرتبط با پرس‌وجوها و نتایج و بکارگیری آنها در الگوریتم‌های موجود، کاربری آنها را در شرایط واقعی، دشوار می‌سازد. علت این امر آن است که در شرایط واقعی، تهیه داده‌های مربوط به این تعداد زیاد از ویژگی‌ها، فرآیندی دشوار و وقت‌گیر می‌باشد. بر این اساس و با الهام از پایه مفهوم «اطلاعات کلیک از گذر داده»، ایده جدیدی در سال‌های اخیر مطرح شده است که در آن، به تولید مجموعه‌ای از ویژگی‌های ثانویه بر پایه ویژگی‌های پایه مجموعه داده محک مورد استفاده پرداخته می‌شود. این مجموعه را اصطلاحاً «ویژگی‌های کلیک از گذر داده» می‌نامند. در [۲۳]، با اِعمال مدل برنامه‌نویسی ژنتیک چند لایه روی ویژگی‌های کلیک از گذر داده، الگوریتم جدیدی موسوم به MGP-Rank ارائه شده است که عملکرد قابل توجهی در رتبه‌بندی اطلاعات انگلیسی وب داشته است. توفیق این الگوریتم، الهام بخش این پژوهش بوده است تا با تطبیق ساختار این الگوریتم بر اساس ویژگی‌های زبان فارسی، الگوریتم رتبه‌بندی قدرتمندی برای دادگان فارسی وب، حاصل شود.

۳- معرفی روش پیشنهادی

این بخش به معرفی و بیان نحوه عملکرد الگوریتم MGP-Rank اختصاص یافته است و جزئیات گام‌های اجرای این الگوریتم، مورد بحث و بررسی قرار گرفته است.

[12] Direct optimization
[13] Deep Learning
[14] Benchmark Dataset

۳-۱-کلیات الگوریتم MGP-Rank

الگوریتم MGP-Rank با استفاده از مدل برنامه‌نویسی ژنتیک چند لایه، روی ویژگی‌های کلیک از گذر داده، عمل می‌کند[۲۳]. در این الگوریتم، ویژگی‌های کلیک از گذر داده، در یک چارچوب مدل برنامه‌نویسی ژنتیک چند لایه و چند جمعیتی، بکار گرفته می‌شوند تا بر اساس برخی شاخص‌های بازیابی اطلاعات، بهترین توابع رتبه‌بندی ممکن، شناسایی شوند. مهم‌ترین مزیت مدل برنامه‌نویسی ژنتیک چند لایه، قابلیت جستجوی سریع‌تر و فراگیرتر آن نسبت به مدل برنامه‌نویسی ژنتیک معمول است. بر این اساس، الگوریتم MGP-Rank، از این جهت با روش‌های مشابه، متفاوت است که در آن، از مفهوم اطلاعات کلیک از گذر داده [۲۱] به منظور تولید ویژگی‌های کلیک از گذر داده، استفاده شده است. فرآیند تولید ویژگی‌های کلیک از گذر داده، در هر دو حالت وجود یا عدم وجود اطلاعات مرتبط با کلیک از گذر داده در مجموعه داده محک مورد استفاده، امکان‌پذیر است. سپس از این ویژگی‌ها در قالب معماری LAGEP [۲۶]، به منظور یافتن توابع رتبه‌بندی مناسب، استفاده می‌شود که ترکیبی از ویژگی‌های مرتبط با محتوای اسناد، ویژگی‌های ساختاری صفحات وب و نیز ویژگی‌های مرتبط با نحوه جستجوی کاربران، خواهد بود. در تعیین توابع برتر، از شاخص‌های ارزیابی مرتبط با مقوله بازیابی اطلاعات نظیر دقت[۱۵]، صحت[۱۶]، فراخوانی[۱۷]، F-measure و نیز خاص بودن[۱۸] به عنوان توابع سازگاری، مورد استفاده قرار گرفته است [۳۲]. نتایج حاصل از آزمایش‌های صورت گرفته، نشان‌دهنده کارآمدی روش ارایه شده در مقایسه با الگوریتم‌های مطرح یادگیری رتبه‌بندی، بر اساس شاخص‌های دقت و NDCG است. این بهبود بازیابی، عمدتاً در بخش بالایی فهرست نتایج جستجوها قابل توجه است که غالباً بیشتر توجه کاربران را به خود، معطوف می‌دارد. در عین حال، بهبود حاصل از الگوریتم ارایه شده، در شرایطی که داده محک اولیه، در برگیرنده اطلاعات کلیک از گذر داده کاربران باشد، بیشتر خواهد بود.

۳-۲- مراحل الگوریتم MGP-Rank

الگوریتم MGP-Rank، متشکل از دو مرحله است که در گام نخست آن، بر اساس مفهوم کلیک از گذر داده، از اطلاعات موجود در مجموعه داده محک اولیه، هشت ویژگی ثانویه تولید می‌شود که آنها را ویژگی‌های کلیک از گذر داده می‌نامیم. برای این منظور، از تعدادی سناریوی متفاوت جهت تولید این ویژگی‌ها، استفاده می‌شود. نکته جالب توجه در خصوص این فرآیند، آن است که این کار، حتی در غیاب اطلاعات کلیک از گذر داده در مجموعه داده محک اولیه نیز امکان‌پذیر می‌باشد. در ادامه و طی گام دوم، از این ویژگی‌های ثانویه به منظور تولید توابع رتبه‌بندی، استفاده می‌شود که در ساختار مدل برنامه‌نویسی ژنتیک چند لایه موسوم به LAGEP [۲۵]، بر اساس توابع سازگاری مرتبط با حوزه بازیابی اطلاعات، تکامل می‌یابند تا اینکه نهایتاً توابع رتبه‌بندی مطلوب، شناسایی شوند.

۳-۲-۱- استخراج ویژگی‌های کلیک از گذر داده

هر چند تحقیقات مختلف نشان داده‌اند که بکارگیری اطلاعات کلیک از گذر داده در بهبود نتایج بازیابی، مفید می‌باشد [۲۲] و [۳۰] و [۳۱]، ولی این قبیل اطلاعات در بسیاری از مجموعه‌های داده ارائه شده برای مساله یادگیری رتبه‌بندی نظیر مجموعه داده LETOR مربوط به مایکروسافت [۳۵] و نیز مجموعه داده تهیه شده توسط شرکت یاهو [۶]، مورد استفاده قرار نگرفته است. از سوی دیگر، این مجموعه‌های داده، شامل تعداد زیادی از ویژگی‌ها هستند و الگوریتم‌های مستخرج از این آنها، بدلیل بکارگیری این حجم بالا از ویژگی‌ها، در شرایط کاربری واقعی، چندان کارآمد نخواهند بود. با توجه به این واقعیات، الگوریتم MGP-Rank در گام نخست و بر اساس مدل مفهومی اطلاعات کلیک از گذر داده، اقدام به تهیه هشت ویژگی ثانویه از روی اطلاعات موجود در مجموعه داده محک اولیه داده شده، موسوم به

[15] Precision
[16] Accuracy
[17] Recall
[18] Specificity

ویژگی‌های کلیک از گذر داده، می‌کند. جزییات مفهومی این ویژگی‌ها در ادامه این بخش، بیان شده است.

در فرآیند استخراج ویژگی‌های کلیک از گذر داده، یک بازنمایی فشرده از مجموعه داده محک اولیه بر اساس مدل مفهومی کلیک از گذر داده ارایه می‌شود [۲۳]. در واقع، مفهوم کلیک از گذر داده، بصورت یک سه‌تایی $\langle Q, R, C \rangle$ شامل پرس‌وجوی کاربر، فهرست نتایج و مجموعه کلیک‌های کاربر، تعریف شده است [۲۱]. با الهام از این مدل، در گام تبدیل، تعداد هشت ویژگی جدید موسوم به ویژگی‌های کلیک از گذر داده، در سه گروه Q، R و C، بشرح زیر تعریف می‌شود:

جدول ۱: ویژگی‌های کلیک از گذر داده معرفی شده مبتنی بر مفهوم اطلاعات کلیک از گذر داده

$Q = \{Repetition, QScore, ResultsAmount\}$
$R = \{AbsoluteRank, StreamLength\}$
$C = \{Specificity, Attractiveness, ClickRate\}$

در این دسته‌بندی، مجموعه Q، شامل ویژگی‌هایی است که تا حدی، مرتبط با ماهیت پرس‌وجوی کاربر هستند. در این دسته، ویژگی $Repetition$، به میزان تکرار واژگان پرس‌وجوی کاربر در بخش‌های مختلف یک سندِ پاسخ، شامل: آدرس (URL)، عنوان و متن مرتبط می‌باشد. ویژگی $QScore$، به میزان مرتبط بودن سند با توجه به پرس‌وجوی کاربر، اشاره می‌کند که در محاسبه آن، ترکیبی از نتایج روش‌های رتبه‌بندی وابسته به پرس‌وجوی کاربر از جمله روش‌های مبتنی بر مدل فضای برداری و روش‌های مدل زبانی، مورد استفاده قرار می‌گیرد. نهایتاً ویژگی $ResultAmount$، معطوف به تعداد پاسخ‌های ارایه شده در قبال پرس‌وجوی مورد نظر می‌باشد.

به همین ترتیب، دسته R، شامل ویژگی‌هایی است که تاکید بیشتری بر مشخصات اسناد، بصورت مستقل از پرس‌وجوی مطرح شده از سوی کاربر دارند. در این دسته، ویژگی $AbsoluteRank$، به رتبه مطلق یک سند صرفنظر از پرس‌وجوی کاربر، می‌پردازد که طبعاً اطلاعاتی نظیر $PageRank$ در آن، نقش مهمی ایفا می‌کنند. ویژگی $StreamLength$ به طول یک سند اشاره دارد و ترکیبی از اطلاعات طول بخش‌های مختلف یک سند از جمله URL، عنوان و متن آن را شامل می‌شود.

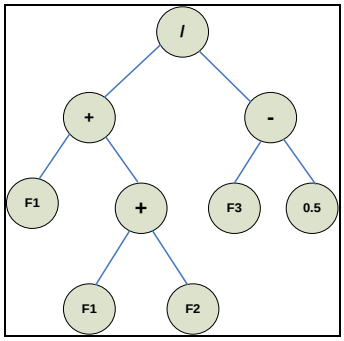
تمرکز عمده ویژگی‌های دسته C، بر اطلاعات قابل حصول از یک زوج سند-پرس‌وجو، در رابطه با کلیک‌های کاربران است. در این دسته، ویژگی $Specificity$، به میزان خاص بودن اطلاعات یک سند، طی تراکنش‌های کاربران در پرس‌وجوهای مختلف، اشاره می‌کند و قصد دارد تا به نوعی، به این نکته بپردازد که یک سند پاسخ، چقدر شامل موضوعات خاص است، بنحوی‌که در تعداد معدودی از پرس‌وجوها، به عنوان یک پاسخ بالقوه، مورد توجه کاربران واقع شده است. ویژگی $Attractiveness$ نیز سعی در نشان دادن میزان جلب توجه کاربران و سامانه‌های بازیابی طی پرس‌وجوهای مختلف به یک سند خاص، دارد. در محاسبه این ویژگی، طبعاً اینکه یک سند، بعنوان اولین یا آخرین پاسخ بالقوه از سوی کاربر، کلیک شده است، حایز اهمیت می‌باشد. نهایتاً ویژگی $ClickRate$، به میزان پتانسیل سند مورد نظر برای دریافت کلیک کاربران طی پرس‌وجوهای مختلف، می‌پردازد.

نکته قابل توجه این است که ویژگی‌های هشت‌گانه مندرج در جدول شماره یک را می‌توان در هر دو صورتِ وجود یا عدم وجودِ اطلاعات صریح کلیک از گذر داده در مجموعه داده محک اولیه، محاسبه نمود. طبعاً روند محاسبه این ویژگی‌ها، وابسته به مفهوم و ساختار ویژگی‌های اصلی موجود در مجموعه‌های داده محک مختلف، متفاوت خواهد بود. بر این اساس، در بخش ۳-۴، سناریوهایی برای محاسبه این ویژگی‌ها روی مجموعه داده فارسی dotIR پیشنهاد می‌شود.

۳-۲-۲- بکارگیری چارچوب برنامه‌نویسی ژنتیک چند لایه

ویژگی‌های کلیک از گذر داده حاصل از اجرای گام نخست الگوریتم MGP-Rank، به عنوان ورودی، جهت بکارگیری در چارچوب مدل برنامه‌نویسی ژنتیک چند لایه و چند جمعیتی موسوم به LAGEP، بکار گرفته می‌شوند [۲۶]. در این ساختار، هر جمعیت، شامل تعدادی افراد است که در اینجا به عنوان توابع رتبه‌بندی بالقوه، مطرح می‌باشند. یک فرد این جمعیت مثلاً I، بصورت سه‌گانه

$< S_v, S_c, S_{op} >$ نمایش داده می‌شود که در آن، S_v، مجموعه متغیرها، S_c، مجموعه مقادیر ثابت و S_{op}، مجموعه عملگرهای ریاضی است. متغیرها، همان ویژگی‌های موجود در داده‌های آموزشی هستند. طی بخش بعدی، جزییات ساختار استفاده شده در آزمایش‌های انجام شده، بیان خواهد شد. بر این اساس، یک جمعیت مانند P که شامل N فرد باشد را می‌توان بصورت زیر تعریف نمود: $P = \{I_1, I_2, \dots, I_N\}$. هر فرد عضو این جمعیت‌ها را می‌توان بصورت یک درخت دودویی نمایش داد که برگ‌های آن، یا متغیرها (ویژگی‌های زوج‌های سند-پرس‌وجو) و یا مقادیر ثابت هستند و گره‌های میانی نیز عملگرهای ریاضی می‌باشند. به عنوان یک نمونه، در شکل شماره سه، بازنمایی فردی از جمعیت را که یک تابع رتبه‌بندی بصورت $\big((F_1 + (F_1 + F_2))/(F_3 - 0.5)\big)$ است، در قالب درخت دودویی، نمایش داده شده است.



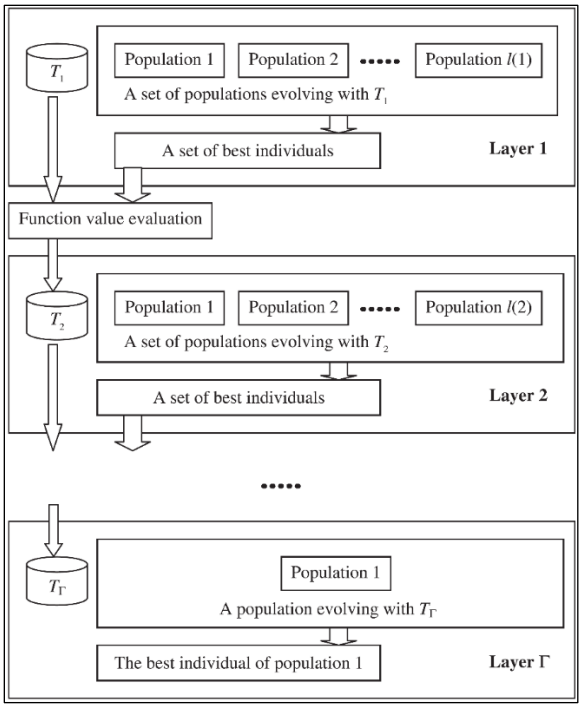
شکل ۳: بازنمایی تابع $\big((F_1 + (F_1 + F_2))/(F_3 - 0.5)\big)$ بصورت درخت دودویی

هر جمعیت با استفاده از سه عملگر ژنتیکی اصلی، یعنی ادغام[19]، جهش[20] و بازتولید[21]، تکامل می‌یابند. نرخ بکارگیری عملگرهای سه‌گانه ادغام، جهش و بازتولید، به ترتیب برابر با R_c، R_m و R_r می‌باشد. در بکارگیری عملگر جهش، نخست بصورت تصادفی، گره‌ای از ساختار دودویی فرد مورد نظر به عنوان نقطه جهش[22] در نظر گرفته می‌شود و با گره دیگری که از مجموعه متناظر یعنی S_v، S_c ویا S_{op}، که بصورت تصادفی انتخاب شده است، جایگزین می‌شود. در واقع، هدف عملگر جهش، اجتناب از بهینه‌های محلی است، زیرا با بکارگیری این عملگر، افراد جدیدی حاصل می‌شوند که ممکن است تا بحال در جمعیت‌های مورد مطالعه، مشاهده نشده باشند. به همین ترتیب، در عملگر ادغام نیز زیردرخت‌هایی از دو فرد که بصورت تصادفی انتخاب شده‌اند، با یکدیگر جایگزین می‌شوند و در نتیجه آن، دو فرد جدید، حاصل می‌شوند. نهایتاً در عملگر بازتولید، قاعده طبیعی بقای قوی‌ترین و مناسب‌ترین افراد، مدل‌سازی می‌شود و بهترین فرد هر نسل از جمعیت، عیناً در نسل بعدی نیز فرصت حضور، خواهد یافت. با توجه به اینکه عملگر جهش، تنوع افراد یک جمعیت را کنترل می‌کند، بدیهی است که نرخ بکارگیری آن (R_m) طی فرآیند تکامل یک جمعیت، بصورت مناسبی تنظیم شود. در LAGEP، این فرآیند از طریق روش تنظیم تطبیقی نرخ جهش (AMRT[23]) صورت می‌پذیرد که جزییات آن در [۲۶] ذکر شده است. بطور خلاصه، در هر نسل از جمعیت، روش AMRT ، اقدام به ارزیابی میزان سازگاری افراد آن جمعیت می‌کند. چنانچه مقادیر میزان سازگاری افراد، یکسان باشد، مقدار R_m افزایش می‌یابد و در غیر این‌صورت، نرخ جهش، برابر با مقدار اولیه آن، خواهد بود. علاوه بر آن، با توجه به اینکه: $R_m + R_c = 1$ می‌باشد، افزایش R_m، به معنی کاهش R_c است. در تحقیق انجام شده، مشخص شده است که استفاده از این تکنیک، منجر به بهبود عملکرد LAGEP خواهد شد [۲۶]. ضمناً بایستی خاطر نشان کرد که برای انتخاب افراد جهت اِعمال عملگرهای ادغام و جهش، می‌توان از روش‌های انتخاب مختلف نظیر انتخاب

[19] Crossover
[20] Mutation
[21] Reproduction
[22] Mutation Point
[23] Adaptive Mutation Rate Tuning

تصادفی[۲۴] یا انتخاب از طریق مسابقه استفاده نمود[۲۵]. در روش انتخاب از طریق مسابقه، ابتدا افرادی از جمعیت هدف به تعداد مشخصی که اصطلاحاً اندازه مسابقه[۲۶] نامیده می‌شود، بصورت تصادفی، انتخاب می‌شوند و سپس بر اساس میزان سازگاری آنها، یک یا دو نفر آنها از به ترتیب، جهت عملگرهای جهش یا ادغام، برگزیده می‌شوند.

در معماری معرفی شده برای LAGEP، جمعیت‌ها در یک ساختار لایه‌ای قرار می‌گیرند که هر لایه، شامل تعدادی جمعیت و مجموعه متغیرها و پارامترهای مختص به خود است. در واقع، هر لایه از مدل LAGEP، خود، یک مدل چند جمعیتی از مدل برنامه‌نویسی ژنتیک است که با استفاده از مجموعه داده آموزشی یکسانی، تکامل می‌یابند و بهترین افراد آنها در یک مجموعه قرار داده می‌شوند و بنا به مدل تشریح شده در [۲۶]، با استفاده از آنها داده‌های آموزشی برای لایه بعدی، تولید می‌شود. در هر لایه و طی تعدادی تکرار، جمعیت‌ها تکامل پیدا می‌کنند و در رقابت افراد آنها با یکدیگر، تنها تعداد معدودی از بهترین آنها با توجه به میزان سازگاری، برای بقا در نسل بعد، باقی می‌مانند. تولید مجموعه متغیرهای لایه‌های بعدی نیز طی روش مشابهی، صورت می‌گیرد [۲۶]. پس از اتمام فرآیند تکامل در همه لایه‌ها، خروجی آخرین لایه، به عنوان خروجی نهایی فرآیند LAGEP، در نظر گرفته می‌شود. شکل شماره چهار، شمای کلی تکنیک برنامه‌نویسی چند لایه و چند جمعیتی را نشان می‌دهد.



شکل ۴: شمای کلی تکنیک برنامه‌نویسی چند لایه و چند جمعیتی (Lin et al. 2008)

۴- آزمایش‌های صورت گرفته

در این بخش، ابتدا به معرفی داده محک dotIR می‌پردازیم و سپس، شاخص‌های مورد استفاده در ارزیابی رتبه‌بندی بیان خواهند شد. ادامه این بخش، به ذکر سناریوهای استخراج ویژگی‌های کلیک از گذر داده روی مجموعه داده محک dotIR، اختصاص یافته است. پس از آن، تنظیمات و پارامترهای آزمایش‌های بعمل آمده، مطرح می‌شود و نهایتاً به طرح نتایج حاصل از آزمایش‌های صورت گرفته و نیز تحلیل نتایج بدست آمده، خواهیم پرداخت.

۴-۱- معرفی داده محک مورد استفاده

در بررسی عملکرد الگوریتم MGP-Rank در محیط وب فارسی، از مجموعه داده dotIR استفاده شده است که مشخصات و ساختار آن در ادامه، بیان شده است. مجموعه داده محک dotIR، تنها داده محک در دسترس برای بازیابی اطلاعات فارسی در محیط وب می‌باشد. این مجموعه داده، توسط دانشگاه تهران و با حمایت مرکز تحقیقات مخابرات ایران، در سال ۱۳۸۸ ایجاد شده است. تولید این مجموعه داده، بر اساس خزش صورت گرفته در دامنه ir. طی سال ۱۳۸۸ و داده‌های جمع‌آوری شده از حدود ۸،۵ میلیون سند

[24] Random Selection
[25] Tournament Selection
[26] Tournament Size

در این دامنه، انجام شده است [۱۲]. این مجموعه داده، شامل ۵۰ پرس‌وجوی مختلف و هزار سند به ازای هر یک از این پرس‌وجوها است که جمعاً پنجاه هزار زوج پرس‌وجو-سند را تشکیل می‌دهند. به ازای هر یک از زوج‌های داده، ۵۶ ویژگی مختلف استخراج شده است که فهرست آنها در جدول پیوست شماره یک، آمده است. این ویژگی‌ها در سه دسته کلی یعنی ویژگی‌های مبتنی بر محتوا، ویژگی‌های مبتنی بر پیوند و نیز ویژگی‌های ترکیبی، ارائه شده‌اند. ضمناً قضاوت انسانی در مورد میزان مرتبط بودن هر زوج پرس‌وجو-سند، در دو سطح مرتبط و نامرتبط، انجام شده است.

به منظور سهولت ارزیابی الگوریتم‌ها، این مجموعه داده به پنج چین مختلف، تقسیم‌بندی شده است. هر دسته، بشرح مندرج در جدول شماره دو، شامل یک بخش آموزش، یک بخش اعتبارسنجی و همچنین یک بخش آزمون می‌باشد.

جدول ۲: دسته‌بندی مجموعه داده dotIR به بخش‌های آموزش، اعتبارسنجی و آزمون

(Darrudi et al. 2009)

چین	داده آموزش	داده اعتبارسنجی	داده آزمون
$Fold_1$	$\{S_1,S_2,S_3\}$	S_4	S_5
$Fold_2$	$\{S_2,S_3,S_4\}$	S_5	S_1
$Fold_3$	$\{S_3,S_4,S_5\}$	S_1	S_2
$Fold_4$	$\{S_4,S_5,S_1\}$	S_2	S_3
$Fold_5$	$\{S_5,S_1,S_2\}$	S_3	S_4

ساختار کلی اطلاعات موجود در این داده محک، در شکل شماره پنج، نشان داده شده است.

Label qid:queryID 1:F_1Value 2:F_2Value 3:F_3Value … 55:F_{55}Value 56:F_{56}Value #docid = docID

شکل ۵: ساختار داده محک dotIR

(Darrudi et al. 2009)

۴-۲- شاخص‌های ارزیابی

به منظور ارزیابی عملکرد الگوریتم MGP-Rank، از شاخص‌های مرجع در ارزیابی رتبه‌بندی یعنی $P@n$، MAP، $NDCG@n$ و $MeanNDCG$، استفاده شده است.

• دقت[۲۷] در نقطه k ($P@k$): اگر قضاوت انسانیِ اسناد، به‌صورت دودویی در اختیار باشد، مثلاً برچسب اسناد مرتبط، یک و برچسب اسناد نامرتبط، صفر باشد، در این صورت، $P@k$ به‌صورت زیر تعریف می‌شود [۳۲]:

$$P@k(\pi,l) = \frac{\sum_{t\le k} I_{\{l_{\pi^{-1}(t)}=1\}}}{k} \quad (۱)$$

که $I_{\{.\}}$ تابع شاخص است و $l_{\pi^{-1}(j)}$ به برچسب سند قرار گرفته در موقعیت j لیست مرتب نتایج، π، اشاره می‌کند.

• متوسط میانگین دقت[۲۸] (MAP): بر اساس تعریف ارائه شده از دقت در موقعیت k، میانگین دقت[۲۹] (AP) عبارتست از:

$$AP(\pi,l) = \frac{\sum_{k=1}^{m} P@k.I_{\{l_{\pi^{-1}(t)}=1\}}}{m_1} \quad (۲)$$

که در آن، m تعداد کل اسناد متناظر پرس‌وجوی q است و m_1 تعداد زیر مجموعه‌ای از این اسناد است که میزان مرتبط بودن آنها برابر با یک باشد. مقدار میانگین AP روی همه پرس‌وجوهای آزمون، متوسط میانگین دقت (MAP) نامیده می‌شود [۳۲].

• بهره انباشته کاهشی هنجار شده[۳۰] (NDCG) در نقطه k ($NDCG@k$): این شاخص بر پایه میانگین‌گیری از شاخص بهره انباشته کاهشی[۳۱] (DCG)، محاسبه می‌شود. شاخص بهره انباشته کاهشی (DCG) می‌تواند قضاوت‌های میزان مرتبط بودن را که به‌صورت طبقه‌بندی‌های مرتب چندگانه ارائه شده باشند، در ارزیابی مورد استفاده قرار دهد؛ ضمن آنکه در تعریف آن، یک ضریب کاهشی بر اساس موقعیت اسناد، در نظر گرفته شده است [۱۸] و [۳۲]۹. برای بیان این موضوع به زبان ریاضی، فرض کنید که لیست

[27] Precision
[28] Mean Average Precision
[29] Average Precision
[30] Normalized Discounted Cumulative Gain
[31] Discounted Cumulative Gain

مرتبی از نتایج نظیر π به ازای پرس‌وجوی q وجود داشته باشد؛ در این صورت، مقدار شاخص DCG در موقعیت k به‌صورت زیر تعریف می‌شود:

$$DCG@k(\pi, l) = \sum_{j=1}^{k} G\left(l_{\pi^{-1}(j)}\right)\eta(j) \quad (۳)$$

که در آن، $G(.)$ رتبه یک سند است که عموماً به‌صورت $G(z) = (2^z - 1)$ می‌باشد و $\eta(j)$ یک ضریب کاهشی بر اساس موقعیت است که غالباً به‌صورت $\eta(j) = 1/log(j+1)$ تعریف می‌شود. با هنجار سازی مقادیر $DCG@k$ با یک مقدار بیشینه (مثلاً Z_k) شاخص بهره انباشته کاهشی هنجار شده ($NDCG$) حاصل می‌گردد که عبارتست از:

$$NDCG@k(\pi, l) = \frac{1}{Z_k}\sum_{j=1}^{k} G\left(l_{\pi^{-1}(j)}\right)\eta(j) \quad (۴)$$

بدیهی است که دامنه مقادیر ممکن برای شاخص $NDCG$، بین صفر و یک قرار می‌گیرد.

- میانگین بهره انباشته کاهشی هنجار شده ($MeanNDCG$): با میانگین‌گیری مقادیر $NDCG@k$ در موقعیت‌های مختلف $k = 1 \dots 10$، شاخص میانگین بهره انباشته کاهشی هنجار شده، محاسبه می‌شود [۳۲].

۴-۳- سناریوهای استخراج ویژگی‌های کلیک از گذر داده

برای تولید ویژگی‌های کلیک از گذر داده در الگوریتم MGP-Rank، بر اساس سناریوهای مندرج در جدول شماره سه، عمل می‌شود و نتایج برترین سناریوها در ادامه، گزارش خواهد شد.

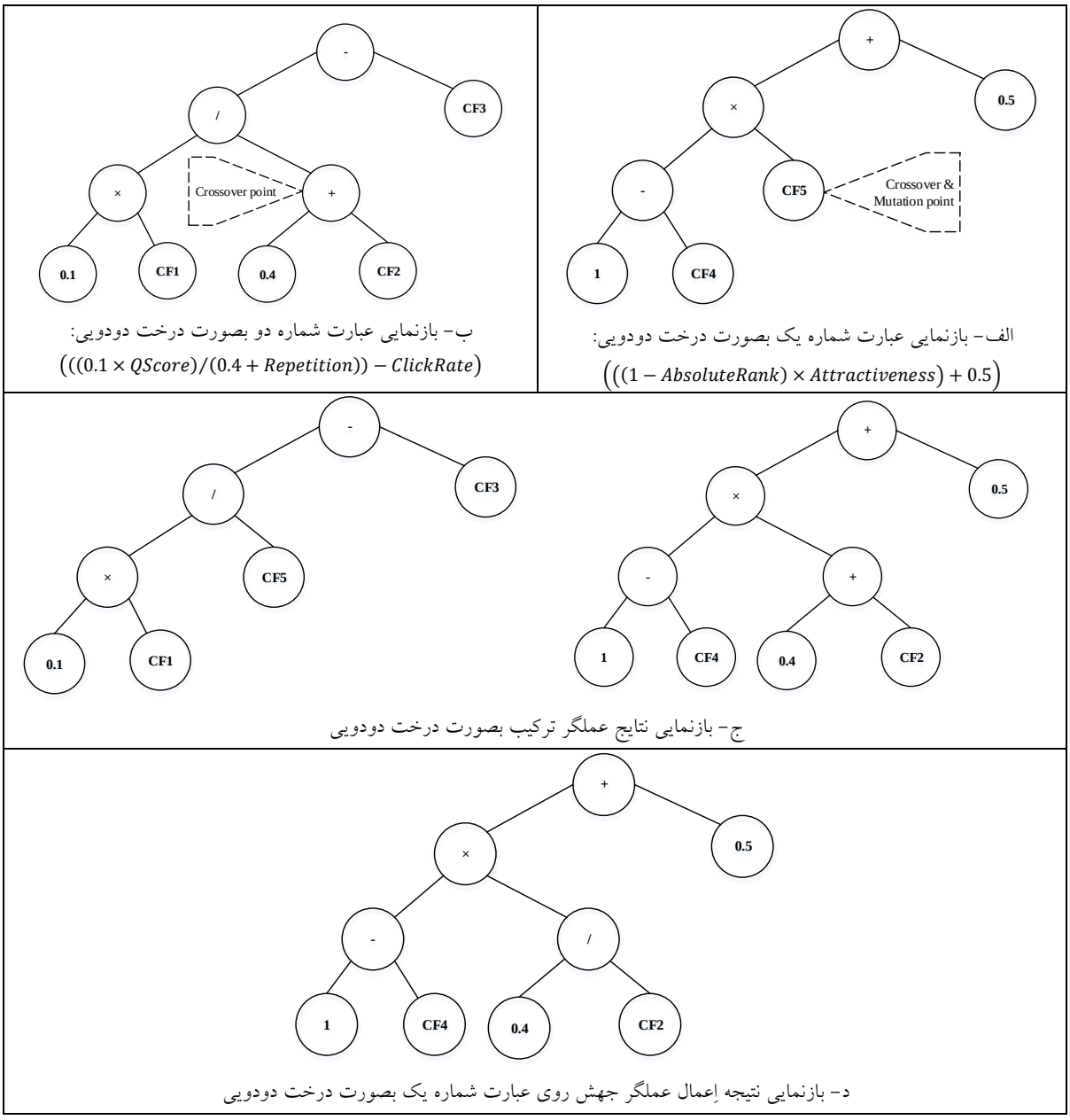
جدول ۳: سناریوهای تولید ویژگی‌های کلیک از گذر داده روی داده محک dotIR

شناسه سناریو	روش محاسبه
IR-DF1	$Q: \begin{cases} Repetition = F_{15}, \\ QScore = \prod_{i=37}^{40} F_i, \\ ResultsAmount = No.\,data\ items\ with: qid = q \end{cases}$ $R: \begin{cases} AbsoluteRank = F_{51}, \\ StreamLength = F_{16} \end{cases}$ $C: \begin{cases} Specificity = F_{54}, \\ Attractiveness = F_{51}, \\ ClickRate = QScore \times AbsoluteRank \end{cases}$
IR-DF2	$Q: \begin{cases} Repetition = F_{15}, \\ QScore = \prod_{i=37}^{40} F_i, \\ ResultsAmount = No.\,data\ items\ with: qid = q \end{cases}$ $R: \begin{cases} AbsoluteRank = F_{51}, \\ StreamLength = F_{20} \end{cases}$ $C: \begin{cases} Specificity = F_{54} \times F_{55}, \\ Attractiveness = F_{51} \times F_{52}, \\ ClickRate = Attractiveness \times QScore \times AbsoluteRank \end{cases}$

$Q:\begin{cases} Repetition = F_{15}, \\ QScore = F_{49} \times \prod_{i=37}^{40} F_i, \\ ResultsAmount = No.\,data\ items\ with: qid = q \end{cases}$ $R:\begin{cases} AbsoluteRank = F_{42} \times F_{51}, \\ StreamLength = F_{20} \end{cases}$ $C:\begin{cases} Specificity = F_{54} \times F_{55}, \\ Attractiveness = F_{43} \times F_{46} \times F_{51} \times F_{52}, \\ ClickRate = Attractiveness \times QScore \times AbsoluteRank \end{cases}$	IR-DF3
$Q:\begin{cases} Repetition = F_{15}, \\ QScore = F_{49} \times \prod_{i=37}^{40} F_i, \\ ResultsAmount = No.\,data\ items\ with: qid = q \end{cases}$ $R:\begin{cases} AbsoluteRank = F_{42} \times F_{51}, \\ StreamLength = F_{20} \end{cases}$ $C:\begin{cases} Specificity = F_{54} \times F_{55}, \\ Attractiveness = \prod_{i=50}^{52} F_i, \\ ClickRate = Attractiveness \times QScore \times AbsoluteRank \end{cases}$	IR-DF4
$Q:\begin{cases} Repetition = F_{15}, \\ QScore = \prod_{i=37}^{40} F_i, \\ ResultsAmount = No.\,data\ items\ with: qid = q \end{cases}$ $R:\begin{cases} AbsoluteRank = F_{51}, \\ StreamLength = F_{20} \end{cases}$ $C:\begin{cases} Specificity = F_{54}, \\ Attractiveness = F_{51} \times F_{52}, \\ ClickRate = Attractiveness \times QScore \times AbsoluteRank \end{cases}$	IR-DF5
$Q:\begin{cases} Repetition = F_{15}, \\ QScore = F_{49} \times \prod_{i=37}^{40} F_i, \\ ResultsAmount = No.\,data\ items\ with: qid = q \end{cases}$ $R:\begin{cases} AbsoluteRank = F_{42} \times F_{51}, \\ StreamLength = F_{20} \end{cases}$ $C:\begin{cases} Specificity = F_{54}, \\ Attractiveness = F_{43} \times F_{46} \times F_{51} \times F_{52}, \\ ClickRate = Attractiveness \times QScore \times AbsoluteRank \end{cases}$	IR-DF6
$Q:\begin{cases} Repetition = F_{15}, \\ QScore = F_{49} \times \prod_{i=37}^{40} F_i, \\ ResultsAmount = No.\,data\ items\ with: qid = q \end{cases}$ $R:\begin{cases} AbsoluteRank = F_{42} \times F_{51}, \\ StreamLength = F_{20} \end{cases}$ $C:\begin{cases} Specificity = F_{54}, \\ Attractiveness = \prod_{i=50}^{52} F_i, \\ ClickRate = Attractiveness \times QScore \times AbsoluteRank \end{cases}$	IR-DF7

شایان توجه است که در همه این سناریوها، ابتدا خصوصیات پایه داده محک dotIR، به روش Min-Max، هنجارسازی شده‌اند و سپس روش هموار سازی موسوم به Dirichlet prior [۳۲]، روی آنها اِعمال شده است.

ویژگی‌های کلیک از گذر داده بدست آمده، در فرآیند تولید جمعیت، به عنوان مجموعه متغیرها، S_v، مورد استفاده قرار می‌گیرند. هر فرد مانند I بصورت یک درخت دودویی نمایش داده می‌شود. در شکل شماره شش، دو فرد نمونه به همراه نتایج حاصل از اِعمال عملگرهای ترکیب و جهش، به تصویر کشیده شده است. در این مثال، متغیرهای CF1، CF2، CF3، CF4 و CF5، به ترتیب نماینده ویژگی‌های کلیک از گذر داده $QScore$، $Repetition$، $ClickRate$، $AbsoluteRank$ و $Attractiveness$ می‌باشند.



شکل ۶: نمونه افراد جمعیت ژنتیکی و نتایج اِعمال عملگرهای ژنتیک روی آنها

۴-۴- تنظیمات و پارامترهای آزمایش‌ها

در این قسمت، نتایج حاصل از بررسی‌های انجام شده مطابق سناریوهای مندرج در بخش ۳-۴، روی مجموعه داده

dotIR ارایه می‌شود. تمامی آزمایش‌ها روی یک رایانه با پردازنده دو هسته‌ای 2.0GHz، شامل 2MB کاشه و 3GB حافظه، صورت پذیرفته است. در ارزیابی عملکرد الگوریتم MGP-Rank از چارچوب LAGEP با تنظیمات مندرج در **Error! Reference source not found.** شماره چهار، استفاده شده است.

جدول ۴: مشخصات چارچوب LAGEP در مطالعه الگوریتم MGP-Rank

پارامتر	مقدار
تعداد لایه‌ها	۳
نرخ بکارگیری عملگرهای ژنتیک	ادغام: ۰,۹۵، جهش: ۰,۰۵
احتمال بکارگیری عملگرهای ریاضی	احتمال مساوی برای $\{+, -, /, \times, Sin(.), Cos(.), Ln(.) \text{ and } Exp(.)\}$
احتمال بکارگیری ضرایب عددی ثابت	احتمال مساوی برای $\{0, 0.1, 0.2, .., 0.9, 1, \pi, e(=2.71828), \pi/2\}$
احتمال بکارگیری انواع گره‌ها	متغیرها: ۰,۰۷۵، مقادیر ثابت: ۰,۰۲۵، عملگر ریاضی: ۰,۹
درصد رشد[۳۲]	۵۰

چارچوب LAGEP مورد استفاده در این تحقیق، از سه لایه مجزا تشکیل یافته است که جزییات آنها در جدول شماره پنج آمده است. هر لایه، شامل تعدادی جمعیت است که در مقداردهی اولیه آنها از روش رشد[۱]، استفاده شده است. در این روش، درخت‌های متناظر با افراد عضو جمعیت‌ها، طی یک فرآیند بازگشتی، ایجاد می‌شوند، به نحوی که عمق درخت حاصله، از مقدار متغیر «عمق درخت»، بیشتر نباشد و حجم درخت نسبت به درخت کامل با همان عمق، از مقدار متغیر «درصد رشد»، تجاوز نکند. جمعیت‌های حاصل از این فرآیند، توسط عملگرهای ژنتیکی با نرخ بکارگیری مشخص شده در جدول شماره چهار، تکامل می‌یابند. سپس از هر جمعیت، فردی با بیشترین میزان تابع برازش، برای بقا در نسل بعد، باقی خواهد ماند. این فرآیند در هر جمعیت تا رسیدن به تعداد مشخصی از نسل‌ها، ادامه پیدا می‌کند.

[32] Grow Percentage

جدول۵: مشخصات لایه‌های چارچوب LAGEP مورد استفاده

مشخصات لایه	لایه شماره یک	لایه شماره دو	لایه شماره سه
تعداد جمعیت‌ها	۱	۲	۳
تعداد افراد هر جمعیت	۱۰۰	۱۰۰	۱۰
تعداد نسل‌ها	۵۰	۱۰	۱۰
اندازه مسابقه	۵	۷	۱
تعداد افراد باقی مانده از یک جمعیت در نسل بعدی	۱	۱	۱
عمق درخت	۴	۶	۴

هر فرد I از یک جمعیت را می‌توان بصورت یک رده‌بندی کننده، در نظر گرفت و کارآیی آن در فرآیند رده‌بندی را توسط یک ماتریس ابهام[۳۳] با چهار مقدار: True Positive (TP)، True Negative (TN)، False Positive (FP) و False Negative (FN)، مندرج در جدول شماره شش، بیان نمود.

جدول ۶: بیان کارآیی فرد I در رده‌بندی بر اساس مقادیر TP، FP، FN و TN

	نمونه مثبت	نمونه منفی
پاسخ I مثبت است	TP	FP
پاسخ I منفی است	FN	TN

در آزمایش‌های انجام شده، از پنج تابع برازش مختلف و مرتبط با حوزه بازیابی اطلاعات، به شرح مندرج در جدول شماره هفت، به منظور ارزیابی کیفیت افراد جمعیت‌ها، استفاده شده است.

جدول۷: توابع سازگاری استفاده شده جهت تعیین سازگاری افراد جمعیت I

شناسه تابع برازش	تابع برازش	روش محاسبه
FF1	Accuracy	(TP + TN) / (TP + TN + FP + FN)
FF2	Precision	TP / (TP + FP)
FF3	Recall	TP / (TP + FN)
FF4	F-measure	2 TP / (TP + FP + FN)
FF5	Specificity	TN / (TN + FP)

همه نتایج گزارش شده در قسمت‌های آتی، بر اساس بازه اطمینان ۹۵٪ حاصل شده‌اند. این نتایج، بر اساس میانگین‌گیری از ۱۵۰ بار تکرار هر نسخه از الگوریتم MGP-Rank روی مجموعه داده dotIR حاصل شده است. دلیل این کار، ماهیت تصادفی مدل برنامه‌نویسی تکاملی است که در نتیجه آن، ممکن است خروجی هر بار اجرای الگوریتم، با دفعات دیگر، تفاوت داشته باشد. تنظیمات مورد استفاده در بهترین نسخه‌های مورد مطالعه عملکرد الگوریتم MGP-Rank در زبان فارسی، در جدول شماره هشت ارایه شده است.

جدول ۸: مشخصات نسخه‌های مختلف الگوریتم MGP-Rank جهت ارزیابی روی مجموعه داده dotIR

شناسه ساختار	شناسه سناریو	شناسه تابع برازش
MGP.IR1	IR-DF7	FF4
MGP.IR2	IR-DF4	FF4
MGP.IR3	IR-DF7	FF1
MGP.IR4	IR-DF6	FF1

[33] Confusion Matrix

۴- نتایج بدست آمده

در جدول شماره نه، نتایج حاصل از بررسی کیفیت بازیابی الگوریتم MGP-Rank بر اساس شاخص دقت روی مجموعه داده dotIR، گزارش شده است. لازم به ذکر است که مقایسه عملکرد الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی بر اساس شاخص دقت، مطابق نتایج درج شده در جدول شماره ده، انجام می‌شود. به منظور تعیین مبنایی جهت مقایسه عملکرد الگوریتم MGP-Rank نسبت به دیگر روش‌های مرجع رتبه‌بندی، در این پژوهش، عملکرد سه الگوریتم پایه رتبه‌بندی، روی مجموعه دادگان dotIR، مورد بررسی قرار گرفته است و نتایج بدست آمده بر اساس شاخص دقت، در جدول شماره ده، گزارش شده است.

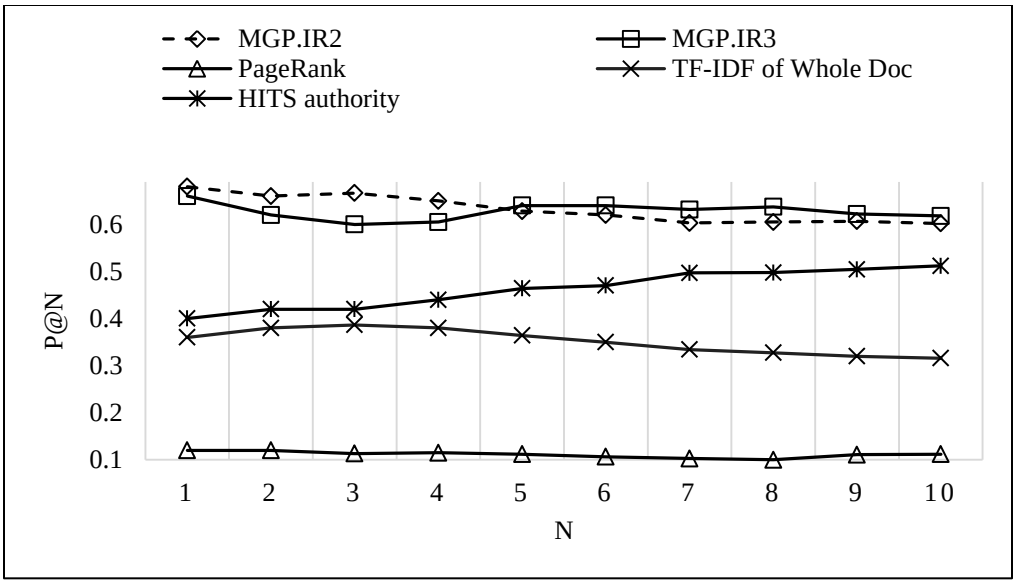
جدول ۹: نتایج حاصل از ارزیابی الگوریتم MGP-Rank بر اساس شاخص دقت روی مجموعه داده dotIR

شناسه ساختار	P@1	P@2	P@3	P@4	P@5	P@6	P@7	P@8	P@9	P@10	MAP
MGP.IR1	0.68	0.61	0.5867	0.57	0.588	0.5867	0.58	0.5775	0.5667	0.56	0.441
MGP.IR2	0.68	0.66	0.6667	0.65	0.628	0.62	0.6029	0.605	0.6067	0.602	0.4408
MGP.IR3	0.66	0.62	0.6	0.605	0.64	0.64	0.6314	0.6375	0.6222	0.618	0.4545
MGP.IR4	0.64	0.61	0.6	0.615	0.64	0.6333	0.6343	0.635	0.62	0.616	0.454

جدول۱۰: نتایج ارزیابی روش‌های استاندارد رتبه‌بندی طبق شاخص دقت روی مجموعه داده dotIR

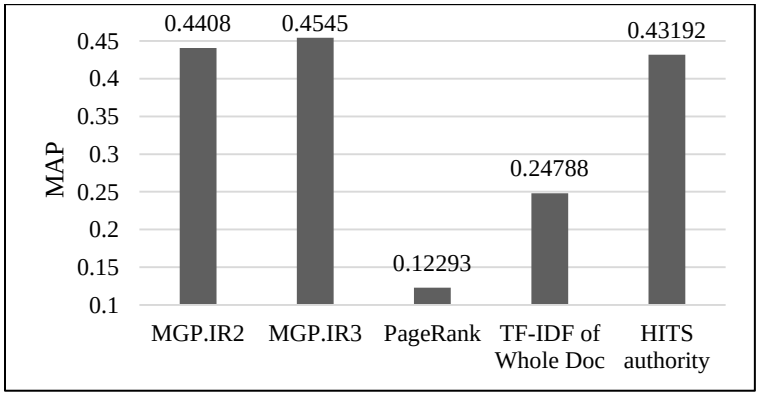
الگوریتم	P@1	P@2			P@3	P@4	P@5	P@6	P@7	P@8	P@9	P@10	MAP
PageRank	0.12	0.12			0.1133	0.115	0.112	0.1067	0.1029	0.1	0.1111	0.112	0.1229
TF-IDF of Whole Doc	0.36	0.38			0.3867	0.38	0.364	0.35	0.3343	0.3275	0.32	0.316	0.2479
HITS authority	0.4	0.42			0.42	0.44	0.464	0.47	0.4971	0.4975	0.5044	0.512	0.4319

مقایسه تصویری اطلاعات جدول شماره نه با نتایج عملکرد روش‌های مرجع مندرج در جدول شماره ده، در شکل شماره هفت، نمایش داده شده است.



شکل۷: مقایسه عملکرد الگوریتم MGP-Rank در زبان فارسی با روش‌های مرجع رتبه‌بندی طبق شاخص دقت

در شکل شماره هشت نیز مقایسه الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی در زبان فارسی طبق شاخص MAP، نمایش داده شده است.



شکل ۸: مقایسه الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی در زبان فارسی طبق شاخص MAP

به همین ترتیب، بررسی عملکرد الگوریتم MGP-Rank روی داده dotIR بر اساس شاخص NDCG نیز انجام شده است که نتایج بدست آمده در جدول شماره یازده، ارایه شده است. ضمن آنکه نتایج بدست آمده به ازای روش‌های رتبه‌بندی مرجع روی دادگان dotIR بر اساس شاخص NDCG، در جدول شماره دوازده، گزارش شده است.

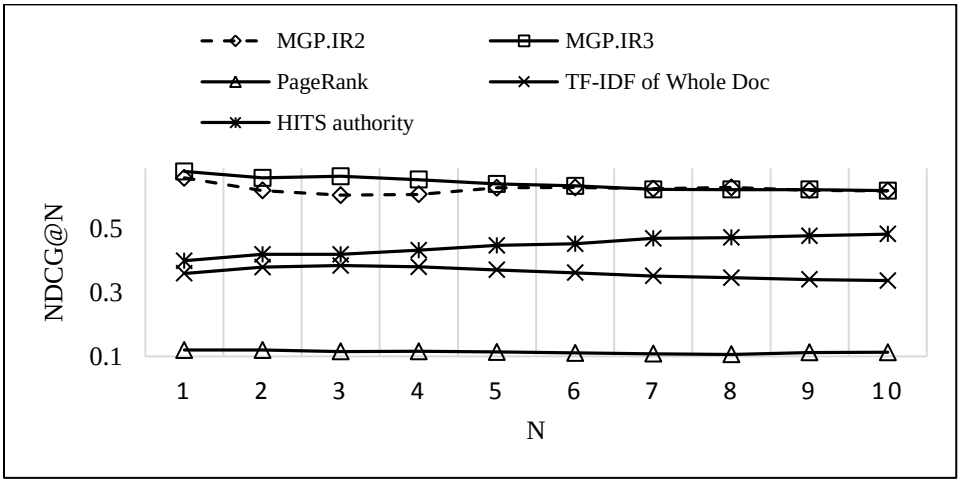
جدول ۱۱: نتایج ارزیابی نسخه‌های الگوریتم MGP-Rank طبق شاخص NDCG روی مجموعه داده dotIR

Mean NDCG	NDCG @10	NDCG @9	NDCG @8	NDCG @7	NDCG @6	NDCG @5	NDCG @4	NDCG @3	NDCG @2	NDCG @1	شناسه ساختار
0.7051	0.5729	0.5773	0.5839	0.5858	0.5899	0.591	0.5815	0.5932	0.61	0.68	MGP.IR1
0.7184	0.619	0.6213	0.6296	0.6257	0.6298	0.6287	0.6079	0.6056	0.62	0.66	MGP.IR2
0.7173	0.6195	0.6231	0.6233	0.6235	0.6347	0.6406	0.6544	0.6648	0.66	0.68	MGP.IR3
0.718	0.6167	0.6189	0.627	0.626	0.6247	0.6274	0.6119	0.6028	0.61	0.64	MGP.IR4

جدول ۱۲: نتایج ارزیابی روش‌های استاندارد رتبه‌بندی بر اساس شاخص NDCG روی مجموعه داده dotIR

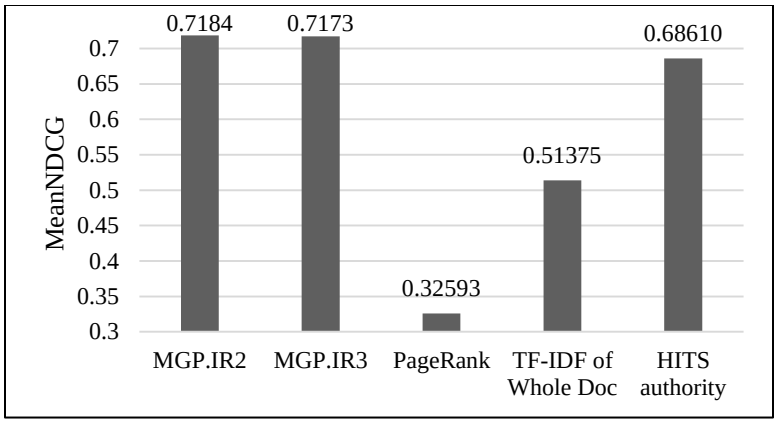
Mean NDCG	NDCG @10	NDCG @9	NDCG @8	NDCG @7	NDCG @6	NDCG @5	NDCG @4	NDCG @3	NDCG @2	NDCG @1	الگوریتم
0.3259	0.1126	0.1121	0.1061	0.1081	0.1107	0.114	0.116	0.1152	0.12	0.12	PageRank
0.5137	0.3378	0.3413	0.3468	0.352	0.3621	0.3711	0.381	0.384	0.38	0.36	TF-IDF of Whole Doc
0.6861	0.4839	0.478	0.4725	0.4703	0.4532	0.4482	0.4328	0.42	0.42	0.4	HITS authority

مقایسه نتایج جدول فوق با عملکرد روش‌های مرجع رتبه‌بندی در شکل شماره نه، انجام شده است که در آن از اطلاعات مربوط به روش‌های مرجع مندرج در جدول شماره دوازده، استفاده شده است.



شکل ۹: مقایسه عملکرد الگوریتم MGP-Rank در زبان فارسی با روش‌های مرجع رتبه‌بندی طبق شاخص NDCG

در شکل شماره ده نیز مقایسه الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی در زبان فارسی طبق شاخص MeanNDCG، نمایش داده شده است.



شکل ۱۰: مقایسه الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی در زبان فارسی طبق شاخص MeanNDCG

تحلیل نتایج

در این بخش، عملکرد الگوریتم یادگیری رتبه‌بندی MGP-Rank، در زبان فارسی، بر اساس شاخص‌های ارزیابی چهارگانه P@n، MAP، NDCG@n و MeanNDCG، روی مجموعه داده محک dotIR مورد بررسی قرار گرفت. یافته‌های حاصل از نتایج از این ارزیابی‌ها، بطور خلاصه، بشرح زیر است:

- در شاخص دقت نقطه‌ای (P@n)، عملکرد الگوریتم MGP-Rank، کاملاً برتر از روش‌های روش‌های مرجع رتبه‌بندی می‌باشد. به عنوان مثال، الگوریتم MGP-Rank، در شاخص P@1، بهبودی معادل ۷۰٪ در مقایسه با برترین روش مرجع رتبه‌بندی، یعنی HITS authority داشته است. به همین ترتیب، در شاخص P@2، نیز افزایش دقتی به ترتیب معادل ۷۰٪ نسبت به بهترین روش مرجع رتبه‌بندی، بدست آمده است.
- در شاخص متوسط میانگین دقت (MAP)، عملکرد الگوریتم MGP-Rank، از روش‌های مرجع رتبه‌بندی بهتر بوده است. در این خصوص، الگوریتم فوق، افزایش دقتی معادل ۲،۰۵٪ نسبت به برترین روش مرجع یعنی HITS authority داشته است.
- در شاخص NDCG@n، نیز عملکرد الگوریتم MGP-Rank، به طرز چشمگیری بهتر از روش‌های مرجع بوده است. به عنوان نمونه، به ازای شاخص NDCG@1، این الگوریتم، بهبود عملکردی معادل ۶۵٪ نسبت به بهترین روش مرجع رتبه‌بندی، یعنی HITS authority، کسب نموده است.
- در شاخص MeanNDCG نیز روش مورد بحث، موفق‌تر از روش‌های مرجع رتبه‌بندی ظاهر شده‌است. در این زمینه، الگوریتم MGP-Rank، حدود ۴،۷۱٪ نسبت به بهترین روش مرجع رتبه‌بندی، یعنی HITS authority، بهبود عملکرد نشان داده است.

یک نکته جالب توجه در بررسی عملکرد الگوریتم MGP-Rank روی مجموعه داده dotIR، این است که استفاده از شاخص متوسط میانگین دقت (MAP) در اولویت‌بندی ویژگی‌ها برای ادغام آنها بکمک عملگرهای میانگین‌گیری مرتب وزن‌دار، مناسب‌تر از دیگر شاخص‌ها بوده است. همچنین بر اساس آمار فوق، می‌توان به این نکته نیز اشاره نمود که الگوریتم یادگیری رتبه‌بندی MGP-Rank در زبان فارسی نیز همانند زبان انگلیسی، موفق بوده است. مساله شایان توجه دیگر این است که در زبان فارسی نیز بهترین عملکرد این الگوریتم، مربوط به بخش آغازین فهرست نتایج بوده است که همانطور که پیش‌تر نیز ذکر شد، بیشترین میزان توجه کاربران جویشگرهای وب را به خود، معطوف می‌سازد. این عملکرد مناسب الگوریتم MGP-Rank در زبان فارسی، با بکارگیری خصوصیاتی مشابهی همچون TF-IDF، BM25، PageRank، HITS، طول سند و نیز خصوصیات مرتبط با مدل‌های زبانی، محقق شده است.

۵- جمع‌بندی و پیشنهاد پژوهش‌های بعدی

با توجه به نقش تعیین کننده جویشگرهای وب در دسترسی کاربران به اطلاعات و خدمات مورد نیاز، فرآیند رتبه‌بندی در این خصوص، بسیار تاثیر گذار است. با عنایت به کاربر-محور بودن فرآیند رتبه‌بندی نتایج جستجو، بکارگیری اطلاعات مربوط به سوابق رفتار کاربران جویشگرهای وب در حین جستجوی اطلاعات مورد نیاز که اصطلاحاً اطلاعات کلیک از گذر داده نامیده می‌شود، در بهبود عملکرد جویشگرها بسیار مفید می‌باشد. با این وجود، ارایه و در نتیجه، استفاده از این قبیل اطلاعات، در اغلب مجموعه‌های داده محک موجود برای یادگیری رتبه‌بندی و به تبع آن، در اکثر روش‌های مطرح شده در این زمینه، مغفول مانده است. همچنین، تعدد ویژگی‌های ارایه شده در این مجموعه‌های داده، ضمن تحمیل هزینه‌های محاسباتی قابل توجه به روش‌های رتبه‌بندی مطرح شده، کاربرد آنها را در شرایط واقعی، دشوار می‌کند.

به منظور پرداختن به این چالش‌ها، در تحقیقات صورت گرفته، رویکرد نوینی برای حل مساله ایجاد یادگیری رتبه‌بندی، ارایه شده است [۲۳] که طی آن، با نگاه مفهومی به اطلاعات کلیک از گذر داده [۲۱]، از ویژگی‌های موجود در مجموعه‌های داده محک مربوط به یادگیری رتبه‌بندی، تعداد معدودی ویژگی‌های ثانویه موسوم به «ویژگی‌های کلیک از گذر داده» تولید می‌شود. این ویژگی‌ها دارای غنای اطلاعاتی مناسب، به منظور ایجاد روش‌های رتبه‌بندی جدید، می‌باشند. در فرآیند تولید ویژگی‌های کلیک از گذر داده، از مدل مفهومی اطلاعات کلیک از گذر داده، الگو برداری شده است. این ویژگی‌ها که تعداد آنها، هشت عدد می‌باشد، در سه دسته ویژگی‌های مرتبط با پرس‌وجوهای کاربران، ویژگی‌های در رابطه با نتایج جستجوها و نیز ویژگی‌های وابسته به الگوهای کلیک کاربران در حین بررسی نتایج، دسته‌بندی می‌شوند. در تحقیقات پیشین، به منظور محاسبه این ویژگی‌ها از اطلاعات موجود در مجموعه‌های داده محک مورد استفاده، تعدادی سناریوی محاسبه، پیشنهاد و بررسی شده است که عمدتاً محدود به اطلاعات و محتوای انگلیسی وب بوده است. به منظور بکارگیری این ایده در زبان فارسی، در این پژوهش، با عنایت به ویژگی‌های زبان فارسی، سناریوهای مناسبی برای تولید ویژگی‌های کلیک از گذر داده، شناسایی شده است. سپس با بکارگیری نیز مدل برنامه‌نویسی ژنتیک چند لایه، روی ویژگی‌های کلیک از گذر داده، از الگوریتم MGP-Rank، استفاده شده است [۲۳]. بطور مشخص، در این الگوریتم، از قابلیت جستجوی فراگیر مدل برنامه‌نویسی ژنتیک چند لایه و چند جمعیتی، برای یافتن توابع رتبه‌بندی مناسب اسناد از طریق ادغام ویژگی‌های کلیک از گذر داده، بهره گرفته می‌شود. با توجه به گزارش موفقیت قابل توجه این الگوریتم در رتبه‌بندی اطلاعات انگلیسی وب، در این مقاله تلاش شده است با پیشنهاد سناریوهای مناسب جهت استخراج ویژگی‌های کلیک از گذر داده روی دادگان فارسی وب، توانمندی آن در زبان فارسی نیز بکار گرفته شود. ارزیابی عملکرد این الگوریتم در بازیابی اطلاعات فارسی وب، نشان‌دهنده افزایش دقت قابل توجه این الگوریتم، نسبت به الگوریتم‌های پایه رتبه‌بندی می‌باشد. این بهبود عملکرد بخصوص در بخش آغازین فهرست نتایج جستجو که بیشتر مورد توجه کاربران است، قابل ملاحظه‌تر بوده است.

این پژوهش از جنبه‌های مختلفی قابل توسعه است. بصورت خاص، به منظور انجام تحقیقات جامع‌تر پیرامون MGP-Rank، ارایه یک مدل استاندارد و خودکار جهت تولید و سنجش عملکرد سناریوهای محاسبه ویژگی‌های کلیک از گذر داده، بر اساس ساختار و مشخصات و نیز ویژگی‌های عرضه شده در مجموعه‌های داده محک پایه مورد استفاده، حائز اهمیت است. در عین حال، بکارگیری روش‌های مختلف ترکیب اطلاعات در سناریوهای محاسبه ویژگی‌های مبتنی بر اطلاعات کلیک از گذر داده، می‌تواند منجر به افزایش توانمندی این الگوریتم گردد. همچنین با عنایت به این نکته که در فرآیند یادگیری رتبه‌بندی، تولید مدل رتبه‌بندی بر مبنای قضاوت کاربران، صورت می‌گیرد و این قضاوت‌های می‌تواند همراه با ابهام و نایقینی باشد، لذا بررسی روش‌های مواجهه با نایقینی و ابهام ذاتی موجود در قضاوت‌های انسانی، در راستای بهبود عملکرد این الگوریتم‌ها، حائز اهمیت خواهد بود. علاوه بر آن، بکارگیری انواع دیگر توابع سازگاری مرتبط با حوزه بازیابی اطلاعات، بخصوص توابع مرتبط با شاخص NDCG، می‌تواند جالب توجه باشد.

پیوست‌ها

پیوست ۱: فهرست ویژگی‌های موجود در مجموعه داده محک dotIR (Darrudi et al. 2009)

شناسه ویژگی	نام ویژگی	نوع ویژگی
F1	Term frequency (TF) of body	مبتنی بر محتوا
F2	TF of anchor	
F3	TF of title	
F4	TF of URL	
F5	TF of whole document	
F6	Inverse document frequency (IDF) of body	
F7	IDF of anchor	
F8	IDF of title	
F9	IDF of URL	
F10	IDF of whole document	
F11	TF×IDF of body	
F12	TF×IDF of anchor	
F13	TF×IDF of title	
F14	TF×IDF of URL	
F15	TF×IDF of whole document	
F16	Document length (DL) of body	
F17	DL of anchor	
F18	DL of title	
F19	DL of URL	
F20	DL of whole document	
F21	BM25 of body	
F22	BM25 of anchor	
F23	BM25 of title	
F24	BM25 of URL	
F25	BM25 of whole document	
F26	LMIR.ABS of body	
F27	LMIR.ABS of anchor	
F28	LMIR.ABS of title	
F29	LMIR.ABS of URL	
F30	LMIR.ABS of whole document	
F31	LMIR.DIR of body	
F32	LMIR.DIR of anchor	
F33	LMIR.DIR of title	
F34	LMIR.DIR of URL	
F35	LMIR.DIR of whole document	
F36	LMIR.JM of body	
F37	LMIR.JM of anchor	
F38	LMIR.JM of title	
F39	LMIR.JM of URL	
F40	LMIR.JM of whole document	
F41	Sitemap based term propagation	ترکیب محتوا و پیوند
F42	Sitemap based score propagation	
F43	Hyperlink base score propagation weighted in-link	
F44	Hyperlink base score propagation weighted out-link	
F45	Hyperlink base score propagation uniform out-link	
F46	Hyperlink base feature propagation weighted in-link	
F47	Hyperlink base feature propagation weighted out-link	
F48	Hyperlink base feature propagation uniform out-link	
F49	HITS authority	مبتنی بر پیوند
F50	HITS hub	
F51	PageRank	
F52	In-link number	
F53	Out-link number	
F54	Number of slash in URL	

	Length of URL	F55
	Number of child page	F56

منابع

1. Q. Ai, K. Bi, J. Guo, W.B. Croft Croft, "*Learning a Deep Listwise Context Model for Ranking Refinement*", In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 135-144, 2018.
2. Q. Ai, X. Wang, S. Bruch, N. Golbandi, M. Bendersky, and M. Najork, "*Learning Groupwise Multivariate Scoring Functions Using Deep Neural Networks*", In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, pp. 85-92, 2019.
3. C.C. Alves, M.A. Gonçalves, D. Sousa, and T. Salles, "*Generalized BROOF-L2R: A General Framework for Learning to Rank Based on Boosting and Random Forests*", In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, pp. 95-104, 2016.
4. Z. Cao, T. Qin, T.Y. Liu, M.F. Tsai, and H. Li, "*Learning to rank: from pairwise approach to listwise approach*", In Proceedings of the 24th International Conference on Machine Learning, pp. 129-136, 2007.
5. S. Chakrabarti, R. Khanna, U. Sawant, and C. Bhattacharyya, "*Structured learning for nonsmooth ranking losses*", In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 88-96, 2008.
6. O. Chapelle & Y. Chang, "*Yahoo! Learning to Rank Challenge Overview*", Journal of Machine Learning Research, pp. 14, 1-24, 2011.
7. O. Chapelle & M. Wu, "*Gradient descent optimization of smoothed information retrieval metrics*", Information Retrieval, Vol. 13, No. 3, pp. 216-235, 2010.
8. W. Chu & Z. Ghahramani, "*Preference learning with Gaussian processes*", In Proceedings of the 22nd International Conference on Machine Learning, pp. 137-144, 2005.
9. W. Chu & S.S. Keerthi, "*New approaches to support vector ordinal regression*", In Proceedings of the 22nd International Conference on Machine Learning, pp. 145-152, 2005.
10. D. Cossock & T. Zhang, "*Subset ranking using regression*", In Proceedings of the 19th annual conference on Learning Theory, pp. 605-619, 2006.
11. F. Dammak, H. Kammoun, and A.B. Hamadou, "*Improving pairwise learning to rank algorithms for document retrieval*", In of IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1-8, 2017.
12. E. Darrudi, H.B. Hashemi, A. Aleahmad, A. Habibian, A. ZarehBidoki, A. Shakery, and M. Rahgozar, "*A standard web test collection for IR domain*", Technical Reoprt, Iran Telecommunication Research Center, 2009.
13. W. Fan, M.D. Gordon, and P. Pathak, "*Ranking function optimization for effective web search by genetic programming: an empirical study*", In Proceedings of the 37th Hawaii International Conference on System Sciences, pp. 1-8, 2004.
14. W. Fan, M.D. Gordon, and P. Pathak, "*On linear mixture of expert approaches to information retrieval*", Decision Support System, Vol. 42, No. 2, pp. 975-987, 2006.
15. N. Fuhr, "*Optimum polynomial retrieval functions based on the probability ranking principle*", ACM Transactions on Information Systems, 183-204, 1989.
16. J. Gao, H. Qi, X. Xia, and J.Y. Nie, "*Linear discriminant model for information retrieval*", In Proceedings of the 28th Annual International ACM SIGIR Conference on

Research and Development in Information Retrieval, pp. 290-297, 2005.
17. Z. Hu, Y. Wang, Q. Peng, and H. Li, "*Unbiased LambdaMART: An Unbiased Pairwise Learning-to-Rank Algorithm*", In Proceedings of the World Wide Web Conference, pp. 2830-2836, 2019.
18. K. Järvelin & J. Kekäläinen, "*IR evaluation methods for retrieving highly relevant documents*", In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 41-48, 2000.
19. K. Järvelin & J. Kekäläinen, "*Cumulated gain-based evaluation of IR techniques*", ACM Transactions on Information Systems, Vol. 20, No. 4, pp. 422-446, 2002.
20. X.B. Jin, G.G. Geng, G.S. Xie, and K. Huang, "*Approximately optimizing NDCG using pair-wise loss*", Information Sciences, Vol. 453, pp. 50-65, 2018.
21. T. Joachims, "*Optimizing search engines using clickthrough data*", In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 133-142, 2002.
22. T. Joachims, L.A. Granka, B. Pan, H.A. Hembrooke, and G .Gay, "*Accurately Interpreting Clickthrough Data as Implicit Feedback*", In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 154-161, 2005.
23. A.H. Keyhanipour, B. Moshiri, F. Oroumchian, M. Rahgozar, and K. Badie, "*Learning to rank: new approach with the layered multi-population genetic programming on click-through features*", Genetic Programming and Evolvable Machines, Vol. 17, pp. 203-230, 2016.
24. M. Köppel, A. Segner, M. Wagener, L. Pensel, A. Karwath, and S. Kramer, "*Pairwise Learning to Rank by Neural Networks Revisited: Reconstruction, Theoretical Analysis and Practical Performance*", arXiv:1909.02768v1, 2019.
25. P. Li, Q. Wu, and C.J. Burges, "*McRank: Learning to rank using multiple classification and gradient boosting*", Advances in Neural Information Processing Systems 20, pp. 845-852, 2008.
26. J.Y. Lin, H.R. Ke, B.C. Chien, and W.P. Yang, "*Classifier design with feature selection and feature extraction using layered genetic programming*", Expert Systems with Applications, Vol. 34, 1384-1393, 2008.
27. Y. Lin, J. Wu, B. Xu, K. Xu, and H. Lin, "*Learning to rank using multiple loss functions*", International Journal of Machine Learning and Cybernetics, Vol. 10, pp. 485-494, 2019.
28. T.Y. Liu, *Learning to Rank for Information Retrieval*, Berlin: Springer-Verlag, 2011.
29. H. Liu, Z. Wu, X. Zhang, "*CPLR: Collaborative pairwise learning to rank for personalized recommendation, Knowledge-Based Systems*", Vol. 148, pp. 31-40, 2018.
30. C. Macdonald, I. Ounis, "*Usefulness of Quality Click-through Data for Training*", In Proceedings of the 2009 workshop on Web Search Click Data, pp. 75-79, 2009.
31. C. Macdonald, R.L. Santos, I. Ounis, "*On the Usefulness of Query Features for Learning to Rank*", In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 2559-2562, 2012.
32. C.D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*, Cambridge University Press, 2008.
33. R. Nallapati, "*Discriminative models for information retrieval*", In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 64-71, 2004.
34. L. Pang, J. Xu, Q. Ai, Y. Lan, X. Cheng, and J. Wen, "*SetRank: Learning a*

ranking gain", In Proceedings of the 26th International Conference on Machine Learning, pp. 1089-1096, 2009.
44. J. Wang, L. Yu, W. Zhang, Y. Gong, Y. Xu, and B. Wang, "*IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models*", In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 515-524, 2017.
45. T. Xia, S. Zhai, and S. Wang, "*Analysis of Regression Tree Fitting Algorithms in Learning to Rank*", arXiv:1909.05965v1, 2019.
46. J. Xu, T.Y. Liu, M. Lu, H. Li, and W.Y. Ma, "*Directly optimizing IR evaluation measures in learning to rank*", In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 107-114, 2008.
47. Q. Xu, M. Li, and M. Yu, "*Learning to rank with relational graph and pointwise constraint for cross-modal retrieval*", Soft Computing, Vol. 23, pp. 9413-9427, 2019.
48. J.Y. Yeh, J.Y. Lin, H.R. Ke, and W.P. Yang, "*Learning to Rank for Information Retrieval Using Genetic Programming*", In Proceedings of the SIGIR 2007 Workshop on Learning to Rank for Information Retrieval, pp. 1-8, 2007.
49. J.Y. Yeh, J.Y. Lin, H.R. Ke, and W.P. Yang, "*Learning to Rank for Information Retrieval Using Layered Multi-Population Genetic Programming*", In Proceedings of the 2012 IEEE International Conference on Computational Intelligence and Cybernetics, pp. 45-49, 2012.
50. Y. Yue, T. Finley, F. Radlinski and T. Joachims, "*A support vector method for optimizing average precision*", In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 271-278, 2007.
Permutation-Invariant Ranking Model for Information Retrieval", arXiv:1912.05891v1, 2019.
35. T. Qin, T.Y. Liu, J. Xu, and H. Li, "*LETOR: Benchmark dataset for research on learning to rank for information retrieval*", In Proceedings of the LR4IR 2007, in conjunction with SIGIR 2007, pp. 3-10, 2007.
36. T. Qin, T.Y. Liu, and H. Li, "*A general approximation framework for direct optimization of information retrieval measures*", Information Retrieval, Vol. 13, No. 4, pp. 375-397, 2009.
37. R. Rahimi, A. Montazeralghaem, and J. Allan, "*Listwise Neural Ranking Models*", In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, pp. 101-104, 2019.
38. E. Renshaw, A. Lazier, C. Burges, T. Shaked, M. Deeds, N. Hamilton, and G. Hullender, "*Learning to rank using gradient descent*", In Proceedings of the 22nd International Conference on Machine Learning, pp. 89-96, 2005.
39. S. Tan, Z. Zhou, and P. Li, "*Fast Item Ranking under Neural Network based Measures*", In Proceedings of the 13th International Conference on Web Search and Data Mining, pp. 591-599, 2020.
40. M. Taylor, J. Guiver, S. Robertson, and T. Minka, "*Softrank: optimising non-smooth rank metrics*", In Proceedings of the 1st International Conference on Web Search and Web Data Mining, pp. 77-86, 2008.
41. A. Trotman, "*Learning to rank*", Information Retrieval, Vol. 8, No. 3, pp. 359-381, 2005.
42. M.F. Tsai, T.Y. Liu, T. Qin, H.H. Chen, and W.Y. Ma, "*Frank: a ranking method with fidelity loss*", In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 383-390, 2007.
43. M.N. Volkovs & R.S. Zemel, "*Boltzrank: learning to maximize expected*

51. Z. Zheng, H. Zha, and G. Sun, "*Query-level learning to rank using isotonic regression*", In Proceedings of the SIGIR 2008 Workshop on Learning to Rank for Information Retrieval, pp. 9-14, 2008.
52. W. Zhou, J. Li, Y. Zhou, M.H. Momen, "*Bayesian pairwise learning to rank via one-class collaborative filtering*", Neurocomputing, Vol. 367, pp. 176-187, 2019.
53. X. Zhu & D. Klabjan, "*Listwise Learning to Rank by Exploring Unique Ratings*", In Proceedings of the 13th International Conference on Web Search and Data Mining, pp. 798-806, 2020.
54. O. Zoeter, M. Taylor, E. Snelson, J. Guiver, N. Craswell, N., and M. Szummer, "*A decision theoretic framework for ranking using implicit feedback*", In Proceedings of the SIGIR 2008 Workshop on Learning to Rank for Information Retrieval, pp. 24-31, 2008.

فصلنامه علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷
صص: ۷۱ -۸۴

زمان‌بندی کارها در محیط‌های ابری با استفاده از چارچوب نگاشت – کاهش و الگوریتم ژنتیک

[*]سید نیما خضر [*]نیما جعفری نویمی‌پور
*کارشناس ارشد، گروه مهندسی کامپیوتر، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران
**دانشیار، گروه مهندسی کامپیوتر، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران
تاریخ دریافت: ۱۳۹۸/۱۰/۰۴ تاریخ پذیرش:۱۳۹۹/۰۳/۱۶

چکیده

زمان‌بندی وظایف یک جزء حیاتی هر سیستم توزیع‌شده همچون گرید، ابر و شبکه‌های نظیر به نظیر می‌باشد که وظایف را برای اجرا به منابع مناسب ارجاع می‌دهد. روش‌های رایج در زمان‌بندی دارای معایبی از قبیل پیچیدگی زمانی بالا، هم‌زمان اجرا نشدن کارهای ورودی و افزایش زمان اجرای برنامه است. الگوریتم‌های زمان‌بندی بر پایه اکتشاف جهت اولویت‌دهی به وظایف از سیاست‌های متفاوتی استفاده می‌کنند که باعث به وجود آمدن زمان‌های اجرای بالا بر روی سیستم‌های رایانش توزیع‌شده ناهمگن می‌شود. بنابراین، روشی مناسب است که اولویت‌دهی آن باعث تولید زمان اجرای کل کمینه گردد. الگوریتم ژنتیک به‌عنوان یکی از روش‌های تکاملی به‌منظور بهینه کردن مسائل NP-کامل[1] به کار گرفته می‌شود. در این مقاله الگوریتم ژنتیک موازی با استفاده از چارچوب نگاشت-کاهش برای زمان‌بندی وظایف بر روی رایانش ابری با استفاده از صف‌های اولویت چندگانه ارائه‌شده است. ایده اصلی این مقاله، استفاده از چارچوب نگاشت-کاهش برای کاهش زمان اجرای کل برنامه می‌باشد. نتایج آزمایش‌ها بر روی مجموعه‌ای از گراف‌های جهت‌دار بدون دور تصادفی حاکی از آن است که روش پیشنهادی زمان اجرای کل دو روش موجود را با سرعت همگرایی بالا بهبود داده است.

واژه‌های کلیدی: رایانش ابری، زمان‌بندی وظایف، کاهش زمان اجرا، الگوریتم ژنتیک، نگاشت-کاهش، هادوپ.

[1]. NP-Completeness

نویسندهٔ عهده‌دار مکاتبات: نیما جعفری نویمی‌پور jafari@iaut.ac.ir

۱- مقدمه

سیستم‌های رایانش توزیع‌شده ناهمگن[۲] نوعی از سیستم‌های رایانش توزیع‌شده است که وظایف را روی چندین پردازنده‌های اجرا میکند [۱]. زمان‌بندی وظایف در محیط‌های ابری به مجموعه‌ای از زیر وظایف کوچک‌تر اولویت‌دار، جهت پردازش تجزیه می‌شود [۲]. این زیروظایف نشان‌دهنده محدودیت تقدم[۳] می‌باشند به این معنی که نتیجه دیگر وظایف قبل از اجرای زیر وظیفه فعلی نیاز است [۲]. با تجزیه یک کار رایانش به زیر وظایف و اجرای آن‌ها بر روی چندین پردازنده، زمان اتمام وظیفه می‌تواند کاهش پیدا کند. بنابراین هدف از این کار، زمان‌بندی زیروظایف بر روی تعدادی پردازنده در دسترس جهت کاهش زمان اتمام وظایف بدون نقض محدودیت‌های تقدم می‌باشد [۳]. توسعه[۴] الگوریتم‌های زمان‌بندی وظایف که زیروظایف یک برنامه را به پردازنده‌ها تخصیص می‌دهند، در سیستم‌های ابری یک چالش می‌باشد. باوجود تلاش‌های بسیاری که اخیراً در زمینهٔ زمان‌بندی وظایف صورت گرفته است، همچنان این مسئله در محیط‌های رایانش ناهمگن به‌عنوان یکی چالش مهم باقی‌مانده است [۴]. تعدادی از چالش‌های مهم زمان‌بندی وظایف در سیستم‌های ابری در زیر آورده شده است:

- ناهمگن بودن منابع
- زمان اجرای الگوریتم[۵]
- زمان اجرای کل[۶]
- سرعت همگرایی[۷] راه‌حل‌ها در روش‌های تکاملی
- بهره‌وری[۸] الگوریتم زمان‌بندی

با توجه به افزایش حجم داده‌ها در محیط ابری کاهش زمان اجرای الگوریتم زمان‌بندی همچنین به‌عنوان یک چالش مهم باقی‌مانده است. بنابراین الگوریتم‌های گوناگونی جهت کاهش زمان اتمام وظایف با موازی‌سازی زیروظایف و رعایت رابطه‌های تقدم آن‌ها ارائه‌شده‌اند. معمولاً رابطه‌های تقدم با گراف جهت‌دار بدون دور[۹] نشان داده می‌شوند که از رأس‌ها که نشان‌دهنده کارها و یال‌های جهت‌دار نشانگر وابستگی بین کارها می‌باشد. گراف‌های جهت‌دار بدون دور برای نشان دادن برنامه‌ها با حجم زیاد و مختلف مناسب و گویا است [۵]. از طرف دیگر سیستم هادوپ، اجرا وظایف را در یک مرکز داده مشترک را فراهم می کند[۱۸]. نگاشت-کاهش یک ابزار مؤثر در هادوپ برای پردازش داده‌های بزرگ در رایانش ابری می‌باشد که با استفاده از پردازنده‌ها و یا کامپیوترها به‌صورت موازی دستورات و برنامه‌ها را اجرا می-نماید [۶]. این چهارچوب برنامه نویسی اجازه اجرای پردازش موازی و توزیع شده بروی مجموعه بزرگی از داده‌ها در یک محیط توزیع‌یافته را می‌دهد [۱۶]. این چارچوب با رویکردها مختلفی ارائه‌شده است و در زمینه‌های مختلف کاربردهای مختلفی دارد ازجمله الگوریتم‌هایی که در داده‌های بزرگ دارای پیچیدگی زمانی بالایی است که با بهبود پردازش به‌صورت موازی زمان اجرای الگوریتم را کاهش می‌دهد [۷]. لذا این مقاله به چگونگی، زمان‌بندی کارهای اولویت‌دار بر روی گراف‌های مستقیم بدون دور بزرگ به کمک الگوریتم ژنتیک و چارچوب نگاشت-کاهش می‌پردازد.

به‌صورت خلاصه اهداف مقاله به‌صورت زیر است:

- استفاده از الگوریتم ژنتیک برای زمان‌بندی در محیط‌های ابری ناهمگن با استفاده چارچوب نگاشت-کاهش
- بهبود زمان اجرای کل برنامه
- افزایش سرعت همگرایی راه‌حل‌ها و جلوگیری از همگرایی زودرس

ادامه مقاله به صورت زیر سازماندهی خواهد شد: در بخش ۲ برخی از روش‌های مهم در زمان‌بندی کارها در محیط‌های ابری مورد مطالعه و بررسی قرار خواهد گرفت. در بخش ۳ ساختار کلی و روش پیشنهادی با استفاده از چارچوب نگاشت-کاهش و الگوریتم ژنتیک نشان داده خواهد شد. در بخش ۴ هم به بررسی نتایج بدست آمده و مقایسه روش

---

[2] .Heterogeneous distributed computing systems
[3] .Precedence constraint
[4] .Development
[5] .Execution time
[6] .Makespan
[7] .Convergence speed
[8] .Efficiency
[9] .Directed acyclic graph

پیشنهادی با روش‌های دیگر در زمان اجرای کل برنامه پرداخته خواهد شد. و در نهایت نتیجه گیری و پیشنهاد کارهای آتی نشان داده خواهد شد.

۲-پیشینهٔ تحقیق

در این بخش روش‌های موجود در زمینه زمان‌بندی وظایف و بررسی مزایا و معایب این روش‌ها دارد. همچنین بعد از بررسی روش‌های موجود ایده اصلی روش زمان‌بندی پیشنهادی این مقاله شرح داده می‌شود. طبیعت چندهدفه در مسئله زمان‌بندی در ابر، حل آن مخصوصاً درزمانی که کارها زیاد و پیچیده می‌باشند را سخت می‌کند. در میان مسائل مختلف زمان‌بندی، زمان‌بندی ماشین با پردازنده‌های موازی یکی از مسائل بهینه‌سازی NP-سخت است. در مسئله زمان‌بندی پردازنده موازی، مجموعه‌ای از n کار وجود دارد ($T_1, T_2, \ldots, T_n$)؛ هرکدام می‌توانند بر روی m پردازنده ($P_1, P_2, \ldots, P_m$) پردازش شوند، هر کار باید با رعایت تقدم بر روی پردازنده پردازش شود. هر پردازنده فقط می‌تواند یک کار را در یک‌زمان معین پردازش کند. در ادامه چندین روش برای زمان‌بندی کارها در محیط‌های ابری بررسی خواهد شد.

پژوهشگران در مقاله [۸] جهت رسیدن همزمان به دو معیار کارآیی بالا و زمان‌بندی سریع دو الگوریتم زمان‌بندی بر پایه لیست[۱۰] HEFT وCPOP ارائه داده‌اند. الگوریتم HEFT در هر مرحله یک وظیفه با بیشترین ارزش رتبه رو به بالا[۱۱] را انتخاب و با استفاده از روش درج[۱۲] به پردازنده‌ای تخصیص می‌دهد که زودترین زمان شروع را کاهش دهد. درحالی‌که الگوریتم CPOP برای اولویت‌دهی از تجمیع مقدارهای رو به بالا و رو به پائین[۱۳] به وظایف استفاده می‌کند. تفاوت دیگر این دو الگوریتم در قسمت انتخاب پردازنده به وظایف می‌باشد به‌طوری که الگوریتم CPOP وظایف موجود بر روی مسیر بحرانی را بر روی پردازنده‌ای اجرا می‌کند که زمان اجرای کل وظایف مسیر بحرانی را کمینه کند. در مقاله [۹] یک الگوریتم زمان‌بندی مبتنی بر الگوریتم کلونی مورچه ارائه‌شده است. هدف این زمان‌بند کمینه کردن زمان گردش مجموعه‌ی کارها همزمان با کمینه کردن حداکثر زمان اتمام انجام کل کارها می‌باشد. این روش ارائه‌شده اجازه پرداختن به کارهای چابک‌تر درحالی‌که زمان تکمیل کاهش می‌یابد را می‌دهد. همچنین نتایج ارزیابی نشان می‌دهد که الگوریتم کلونی مورچه بهتر از الگوریتم‌های تصادفی و بهترین تلاش عمل می‌کند. در مقاله [۱۰] روشی در زمینه زمان‌بندی کارها در ابر خبره ارائه‌شده است. این روش دارای عملکرد سریع‌تر با دقت بالاتر و نرخ شکست پایین‌تر در مقایسه با روش‌های بررسی‌شده می‌باشد. نتایج پیاده‌سازی نشانگر بهبود حداکثر ۵۶ درصدی، ۱۴ درصدی و ۷٫۵ درصدی به ترتیب بر اساس معیارهای زمان کلی اجرا، نرخ شکست و کارایی منابع انسانی در مقایسه با روش‌های دیگر می‌باشد. در مقاله [۱۱] یک الگوریتم زمان‌بندی سطوح پویا ارائه‌شده است. نتایج شبیه‌سازی‌ها اثبات کرده‌اند که این الگوریتم می‌تواند به‌صورت کارا نیازهای جریان کار رایانش ابری را ازنظر اعتماد، کاهش هزینه‌های زمانی و اطمینان از اجرای کارها در یک حالت امن را تأمین کند. پژوهشگران در مقاله [۱۲] یک روش زمان‌بندی وظایف در سیستم‌های رایانش ناهمگن با استفاده از الگوریتم ژنتیک و صف‌های اولویت‌دار چندگانه به نام MPQGA را ارائه کرده‌اند.

ایده اصلی روش ارائه‌شده به‌کارگیری مزیت‌های هر دو نوع الگوریتم اکتشافی و تکاملی و اجتناب از ایرادات آن‌ها می‌باشد. روش ارائه‌شده الگوریتم ژنتیک را جهت تخصیص اولویت به هر وظیفه و روش اکتشافی EFT را جهت نگاشت[۱۴] وظایف به پردازنده‌ها به کار گرفته است. همچنین روش MPQGA عملگرهای ترکیب، جهش و تابع برازندگی مناسبی را جهت زمان‌بندی گراف جهت‌دار بدون دور طراحی کرده است. نتایج شبیه‌سازی انجام‌شده بر روی گراف‌های دنیای واقعی[۱۵] و تصادفی نشان می‌دهد که

[10] .List-based scheduling
[11] .Upward rank
[12] .Insertion-based approach
[13] .Downward rank
[14]. Mapping
[15] .Real world graphs

الگوریتم MPQGA دو روش غیرتکاملی و یک روش جستجوی تصادفی به نام BGA [16]را بهبود بخشیده است. در این ادامه به مقایسه روش‌های موجود پرداخته‌شده است. در الگوریتم HEFT وظایف با استفاده از ارزش رو بالا انتخاب می‌شود و دارای همگرای متوسط می‌باشد، در روش CPOP وظایف با استفاده از ارزش رو بالا و پایین انتخاب می‌شود و مانند HEFT دارای سرعت همگرایی متوسط می‌باشد. در الگوریتم MPQGA که بهبودیافته الگوریتم HEFT است برای بهبود زمان‌بندی از الگوریتم ژنتیک استفاده‌شده است. جدول ۱ الگوریتم‌های موجود را از جنبه روش مورداستفاده، مزایا و معایب آن‌ها مورد مقایسه قرار می‌دهد.

جدول۱: مقایسه روش‌های بررسی‌شده برای زمان‌بندی کارها در محیط‌های ابری

الگوریتم‌های زمان‌بندی	معایب	مزایا	زمان اجرا
[۸]	همگرایی کند، عدم تولید نتایج پایدار در مسائل با طیف گسترده	پیچیدگی کمتر	متوسط
[۹]	نتایج نامناسب در گراف‌های بر پایه محاسبات	پیچیدگی و هزینه ارتباطی کمتر	بالا
[۱۰]	همگرایی کند	هزینه ارتباطی کمتر	متوسط
[۱۱]	مناسب برای محیط‌های همگن	هزینه ارتباطی و اجرای کمتر	متوسط
[۱۲]	همگرایی کند و پیچیدگی بیشتر	کارایی بهتر	متوسط

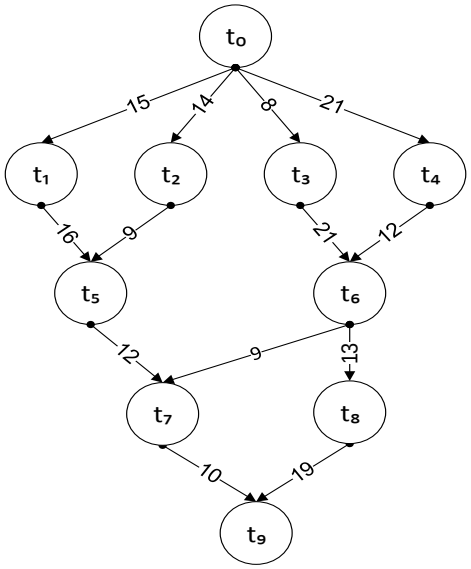
در روش‌های بررسی‌شده، زمان اجرای کل برنامه‌ها بالا است. در صورتی که تعداد زیر وظایف افزایش پیدا کند زمان اجرای کل برنامه افزایش می‌یابد. مطالعات اخیر نشان می‌دهد که الگوریتم‌های ژنتیک بر روی انواع مسائل طبیعی عملکرد موفقی داشته‌اند. الگوریتم ژنتیک یک روش جستجوی مؤثر برای پیدا کردن یک جواب بهینه یا نزدیک به بهینه هست. با توجه به‌سرعت پایین الگوریتم ژنتیک در مسائل بزرگ، دراین مقاله برای افزایش سرعت همگرایی و دوری از همگرایی زودرس و کاهش زمان اجرای برنامه از چارچوب نگاشت-کاهش استفاده شده است.

۳-روش پیشنهادی

در این بخش به روش پیشنهادی پرداخته شده است. در ابتدا تعاریف متغییرهای مسئله بیان شده است در ادامه به معرفی کردن الگوریتم ژنتیک برای ترکیب کردن خدمات ابری پرداخته شده است. در ادامه یک الگوریتم ژنتیک موازی جدید زمان‌بندی وظیفه با استفاده از چارچوب نگاشت-کاهش بر روی رایانش ابری با استفاده از صف‌های اولویت چندگانه ارائه‌شده است.

۳-۱-تعاریف اولیه

روش پیشنهادی دارای یک مجموعه از پردازنده‌های ناهمگن P می‌باشد که توسط شبکه پرسرعت به همدیگر متصل شده‌اند. در این روش مسئله زمان‌بندی وظایف توسط گراف جهت‌دار بدون دور G(V,E) نمایش داده‌شده است که شکل ۱ نمونه ارائه‌شده در این مقاله را نشان می‌دهد.



شکل ۱: گراف جهت‌دار بدون دور با ۱۰ زیروظیفه [۵]

[16] Basic genetic algorithm

به‌طوری که V نشان‌دهنده زیروظایف برنامه و E نشان‌دهنده لبه‌های گراف است که وابستگی‌های بین زیروظایف را مشخص می‌کنند. هر زیر وظیفه در گراف فقط بر روی یک پردازنده می‌تواند اجرا شود. مقدارهای لبه‌ها نشانگر مقدار هزینه ارتباطی بین دو زیر وظیفه است و میانگین هزینه محاسباتی هر زیر وظیفه در پردازنده‌ها بر روی هر گره مقدارگذاری شده است. جدول ۲ نمادهای استفاده‌شده در رابطه‌ها و الگوریتم‌های این مقاله را توضیح می‌دهد.

جدول ۲ :نمادهای استفاده‌شده در روش پیشنهادی

نماد	شرح
ti	i امین زیر وظیفه در گراف
pk	k امین پردازنده در سیستم
E	مجموعه لبه‌ها در گراف
T	مجموعه زیروظایف در گراف
P	مجموعه پردازنده‌ها در سیستم
tentry	زیر وظیفه ورودی در گراف
texit	زیر وظیفه خروجی در گراف
Succ(ti)	مجموعه تأخرهای زیر وظیفه ti
Pred(ti)	مجموعه تقدم‌های زیر وظیفه ti
$\overline{W(t_i)}$	میانگین هزینه محاسباتی زیر وظیفه ti
$W(t_i, p_k)$	هزینه محاسباتی زیر وظیفه ti بر روی پردازنده
$C(t_i, t_j)$	مقدار هزینه ارتباطی بین وظایف ti و t_j
$rank_b(t_i)$	رتبه رو به بالای زیروظیفه ti
$rank_t(t_i)$	رتبه رو به پایین زیروظیفه ti
Rankb+t(ti)	جمع رتبه روبه بالا و پایین زیروظیفه ti
$EST(t_i, p_k)$	زودترین زمان شروع زیروظیفه ti بر روی
$EFT(t_i, p_k)$	زودترین زمان پایان زیروظیفه ti بر روی
$AST(ti, pk)$	زمان شروع واقعی زیروظیفه ti بر روی
$AFT(ti, pk)$	زمان پایان واقعی زیروظیفه ti بر روی پردازنده
$avail\{k\}$	زمان بیکار و در دسترس بودن پردازنده pk
pop	مخفف کلمه انگلیسی جمعیت
PopSize	مخفف کلمه انگلیسی اندازه جمعیت
ChSize	مخفف کلمه انگلیسی اندازه کروموزوم

درصورتی‌که دو زیر وظیفه متصل به یکدیگر بر روی پردازنده یکسان زمان‌بندی شوند هزینه ارتباطی بین آن‌ها صفر در نظر گرفته می‌شود. گره‌های شروع و خروجی گراف به ترتیب نشان‌دهنده شروع و پایان برنامه هستند. مدل استفاده برای الگوریتم پیشنهادی نامرتبط[17] می‌باشد، بدین معنی که یک پردازنده ممکن است بعضی از وظایف را در زمان کمتر و بعضی دیگر را در زمان بیشتر اجرا کند که در جدول ۳ نشان داده شده است. همچنین برای شروع اجرای یک زیروظیفه می‌بایست تمام زیر وظایف تقدم[18] آن زمان‌بندی و اجراشده باشند

جدول ۳: هزینه‌های محاسباتی زیروظایف بر روی پردازنده‌ها [۱۷]

Task	p_0	p_1	p_2	$\bar{\omega}$
t_0	۸	۹	۱۰	۹
t_1	۹	۱۰	۱۱	۱۰
t_2	۱۰	٦	۱۱	۹
t_3	۱۲	۸	۱٦	۱۲
t_4	۲۵	۱۸	۱۷	۲۰
t_5	۱۳	۱۰	۱٦	۱۳
t_6	۷	۱۲	۱۷	۱۲
t_7	۱۷	۱۰	۱۲	۱۳
t_8	۱۲	۸	۱۳	۱۱
t_9	۱۳	۱۰	۱۳	۱۲

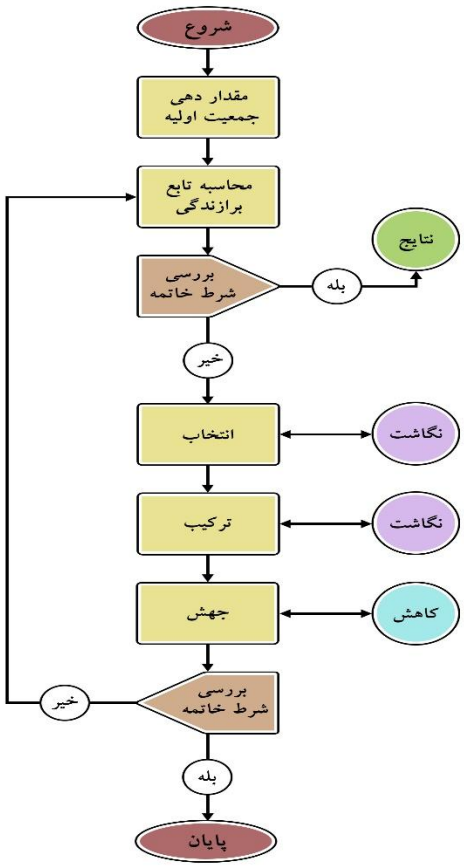
۲-۳-روش پیشنهادی

در این بخش ابتدا ساختار الگوریتم ژنتیک ارائه می‌شود. سپس به تشریح جزئیات روش پیشنهادی پرداخته خواهد شد. همچنین برای بهبود الگوریتم ژنتیک پیشنهادی از چارچوب برنامه‌نویسی نگاشت-کاهش استفاده‌شده است. الگوریتم تکاملی ژنتیک از عملگرهای مختلفی استفاده می‌کند که عملگرهای ژنتیک نامیده می‌شوند. این عملگرها باعث تولید کروموزوم‌هایی بهتر از جمعیت اولیه می‌شوند. این عملگرها جهت همگرایی راه‌حل‌های مسئله لازم و ضروری هستند [۱۳]. یک الگوریتم ساده ژنتیک دارای سه عملگر اساسی ژنتیک است. این عملگرها شامل عملگرهای انتخاب، ترکیب و جهش هستند. عملگر انتخاب بعضی از راه حل ها را به عنوان والد از میان جمعیت انتخاب می کند. عملگر ترکیب والد های انتخاب شده را جهت تولید فرزندان

[17] .Unrelated
[18] .Precedence

جدید ترکیب می کند و عملگر جهش مطابق قوانین مشخص شده فرزندان را تغییر می دهد.

برای توسعه و پیاده‌سازی یک الگوریتم جدید زمان‌بندی وظایف، از ترکیب الگوریتم ژنتیک و چارچوب نگاشت-کاهش استفاده‌شده است. با تکنیک نگاشت-کاهش می‌توان الگوریتم‌ها را در فضای موازی پیاده‌سازی نمود. هادوپ ابزاری است که این تکنیک را پیاده‌سازی نموده است و برای هر الگوریتم تنها کافی است دو قسمت نگاشت و کاهش آن را پیاده‌سازی کنیم [۲۲]. مراحل الگوریتم پیشنهادی در شکل ۲ نمایش داده‌شده‌اند. الگوریتم پیشنهادی با مقداردهی جمعیت اولیه شروع می‌شود.



شکل ۲: مراحل روش پیشنهادی

بعد از تولید جمعیت اولیه هر کروموزوم ارزیابی‌شده و با مقدار زمان اجرای کل رتبه‌بندی می‌شوند و این رتبه‌بندی به‌عنوان مقدار برازندگی هر کروموزوم شناخته می‌شود. رتبه بندی انجام‌شده از روش تخصیص پردازنده HEFT استفاده می‌کند. سپس تعدادی از کروموزوم‌ها جهت انجام عمل ترکیب انتخاب و ترکیب می‌شوند. بعد از مرحله ترکیب، عمل موازی‌سازی روی کروموزم‌ها اعمال می‌شود. الگوریتم پیشنهادی تا زمان وقوع شرط خاتمه تکرار می‌شود. در ادامه این مقاله به جزئیات تک‌تک مراحل روش پیشنهادی و شبه کدهای آن پرداخته می‌شود.

روش پیشنهادی زمان‌بندی وظایف را در دو مرحله انجام می دهد: در مرحله یکم به‌منظور تخصیص وظایف به پردازنده‌ها از روش ژنتیک به همراه روش‌های اکتشافی استفاده‌شده است و در مرحله دوم تخصیص وظایف به پردازنده‌ها از روش ژنتیک به همراه چارچوب نگاشت-کاهش استفاده‌شده است. تابع اصلی الگوریتم پیشنهادی جهت زمان‌بندی کردن وظایف بر روی منابع ناهمگن در ابر است.

۳-۳-مقداردهی جمعیت اولیه

ساختن یک کروموزوم توسط روش رمزنگاری[19] مرحله ابتدایی در مقداردهی جمعیت اولیه می‌باشد [۱۹]. ساختمان کروموزوم الگوریتم پیشنهادی دارای ژن[20] می باشد که نشان‌دهنده یک راه‌حل از مسئله زمان‌بندی وظایف می‌باشد [۲۰]. ساختمان یک کروموزوم شامل یک جایگشت از اعداد طبیعی ۰ تا n-1 می‌باشد که نشانگر ترتیب اولویت‌های وظایف در گراف جهت‌دار بدون دور با تعداد n می‌باشد. علاوه براین ترتیب اولویت وظایف در کروموزوم می‌بایست یک ترتیب توپولوژیکی معتبر باشد بطوریکه گره شروع بایستی در اول کروموزوم و گره پایانی در آخر کروموزوم قرار گیرد و همچنین سایر وظایف در مکان‌هایی از کروموزوم قرار گیرند که محدودیت‌های تقدم وظایف را نقض نکنند. به‌منظور عملی بودن زمان‌بندی باید تمام زیر وظایف موجود در گراف زمان‌بندی شوند.

---

[19] Encoding.

[20] Genes.

جدول ۳: اولویت‌های وظایف

Subtask (ti)	Rankb	Rankt	Rankb+t	level
t0	۱۲۹	۰	۱۲۳	۰
t1	۸۱	۲۴	۱۰۵	۱
t2	۷۸	۲۳	۱۰۱	۱
t3	۱۰۰	۱۷	۱۱۷	۱
t4	۹۹	۳۰	۱۲۹	۱
t5	۶۰	۵۰	۱۱۰	۲
t6	۶۷	۵۲	۱۱۹	۲
t7	۳۵	۷۵	۱۱۰	۳
t8	۴۲	۷۷	۱۱۹	۳
t9	۱۲	۱۰۷	۱۱۹	۴

اندازه جمعیت اولیه الگوریتم پیشنهادی با نماد PopSize نمایش داده می‌شود و مقدار آن ۴ برابر تعداد وظایف موجود در گراف است (n×۴). اندازه جمعیت الگوریتم تا پایان یافتن الگوریتم ثابت است یعنی در طول تولید نسل‌های جدید این مقدار تغییری نمی‌کند. در الگوریتم پیشنهادی به‌منظور داشتن تخم‌ریزی[21] خوب در مقداردهی جمعیت اولیه از سه سیاست رتبه‌بندی اکتشافی بنام‌های رتبه رو به بالا(رابطه (۱))، رتبه رو به پایین (رابطه (۲)) و ترکیبی از این دو روش با رتبه سطح (رابطه (۳)) برای مقداردهی سه کروموزوم اول جمعیت استفاده‌شده است [۱۲]. در جدول ۳ با استفاده از سه سیاست رتبه‌بندی شده است. اولویت‌های کروموزوم‌های باقیمانده با جایگشت‌های تصادفی تولید می‌شوند. کروموزوم‌های تولیدشده تصادفی به دلیل اینکه اولویت‌هایشان معتبر نیست از سمت چپ به راست به‌طوری‌که زیروظایف محدودیت‌های اولویت را نقض نکنند مرتب‌سازی می‌شوند. فرآیند تولید جمعیت اولیه درالگوریتم ۱ نشان داده شده است.

$$rank_b(t_i) = \overline{W(t_i)} + \max_{t_j \in succ(t_i)} \left(C(t_i, t_j) + rank_b(t_j) \right) \quad (۱)$$

در این رابطه) $\overline{W(t_i)}$ میانگین هزینه محاسباتی وظیفه ti، $C(t_i, t_j)$ مقدار هزینه ارتباطی بین وظایف ti وtj و $rank_b(t_j)$ رتبه رو به بالای تاخر زیر وظیفه ti است.

$$rank_t(t_i) = \max_{t_j \in pred(t_i)} \left(rank_t(t_j) + \overline{W(t_j)} + C(t_i, t_j) \right) \quad (۲)$$

در این رابطه) $rank_t(t_j)$ رتبه رو به پایین زیر وظیفه تقدم ti است.

$$Level(t_i) = \begin{cases} 0, & \text{if } t_i = t_{entry}; \\ \max_{t_j \in pred(t_i)} \left(Level(t_j) \right) + 1, & \text{otherwise} \end{cases} \quad (۳)$$

در این رابطه) $Level(t_j)$ رتبه سطح تقدم زیر وظیفه t_i است.

Algorithm 1. Initial population

1: Initial three of the chromosomes by using three heuristic rank policies
2: for i=3 to PopSize-1 do
3: for j=0 to ChSize-1 do
4: generate a genes j ∈ $(0, ChSize - 1)$ randomly that not generated in pervious genes
5: move chromosome i from left to right in first valid place in the queue in order to has a valid topological order
6: end for
7: end for

الگوریتم ۱: مقداردهی جمعیت اولیه

۳-۴-تخصیص زیروظایف به پردازنده‌ها

در جمعیت اولیه تولیدشده هر کروموزوم باید دارای ترتیب اولویت معتبر باشد یعنی محدودیت‌های تقدم در کروموزوم نقض نشود. در روش پیشنهادی به‌منظور تخصیص وظایف به پردازنده‌ها از سیاست تخصیص پردازنده HEFT استفاده‌شده است [۸]. در این روش زیروظیفه با بالاترین اولویت از بین زیروظایف کروموزوم انتخاب و به پردازنده‌ای تخصیص داده می‌شود که زمان اجرای آن از سایر پردازنده‌ها کمینه باشد. این روش تخصیص پردازنده از سیاست زمان‌بندی بر اساس درج استفاده می‌کند. در ادامه این بخش به تشریح این روش پرداخته‌شده است. زودترین زمان شروع زیر وظیفه ti بر روی پردازنده pk به‌صورت $EST(t_i, p_k)$ نمادگذاری می‌شود و از رابطه (۴) به دست می‌آید.

[21] .Seeding

$$EST(t_i,p_k) = max\left\{\max_{t_j \in pred(t_i)}\left(AFT(t_j) + C(t_j,t_i)\right)\right\} \quad (۴)$$

زمان شروع واقعی زیر وظیفه ti بر روی پردازنده pk با $AST(t_i,p_k)$ نمادگذاری می‌شود و از رابطه (۵) به دست می‌آید.

$$AST(t_i,p_k) = max(EST(t_i,p_k), Avail\ (p_k)) \quad (۵)$$

که Avail (pk) زمانی است که پردازنده pk بیکار و آماده اجرای وظایف باشد. زودترین زمان پایان زیر وظیفه ti بر روی پردازنده pk با $EFT(t_i,p_k)$ نمادگذاری شده است و از رابطه (۶) محاسبه می‌شود.

$$EFT(t_i,p_k) = W(t_i,p_k) + EST(t_i,p_k) \quad (۶)$$

که $W(t_i,p_k)$ هزینه محاسباتی زیر وظیفه ti بر روی پردازنده pk است. زمان واقعی پایان زیر وظیفه ti بر روی پردازنده pk با $AFT(t_i,p_l)$ نشان داده می‌شود و از رابطه (۷) به دست می‌آید.

$$AFT(t_i,p_k) = \min_{1\le l\le P} EFT(t_i,p_l) \quad (۷)$$

که pk در رابطه بالا مناسب‌ترین[۲۲] پردازنده برای زیر وظیفه ti می‌باشد. شبه کد تخصیص وظایف به پردازنده‌ها در الگوریتم ۲ در تشریح شده است.

Algorithm 2. Assigning subtasks to processors functin
1: Fill the priority queue with subtasks 2: while the priority queue is not empty do 3: Select the first subtask ti from the priority queue 4: for each processor pk in the processor set do 5: Compute $EFT(t_i,p_k)$ value using the insertion-based HEFT scheduling policy 6: Assign subtask ti to the processor pk that minimizes $EFT(t_i,p_k)$ 7: end for 8: Remove ti from the priority queue 9: end while 10: Return makespan= $AFT(t_{exit})$.

الگوریتم ۲) تخصیص وظایف به پردازنده‌ها

۳-۵-تابع برازندگی

مقدار برازندگی کروموزوم‌ها در تصمیم‌گیری اینکه کدام کروموزوم‌ها به نسل آینده انتقال یابند بسیار مهم است. زمان پایان واقعی آخرین زیر وظیفه در گراف جهت‌دار بدون دور یا به عبارتی گره خروجی گراف زمان اجرای کل برنامه می‌باشد. زمان اجرای کل برنامه از رابطه (۸) به دست می‌آید. که همان مقدار برازندگی کروموزوم i است.

$$makespan\ i = max\{AFT(t_{exit})\} \quad (۸)$$

مقدار برازندگی کوچک بیانگر کوتاه بودن زمان اجرای کل است. بنابراین بیشترین و کمترین مقدار برازندگی در میان جمعیت به ترتیب بدترین و بهترین کروموزوم‌ها هستند.

۳-۶-ترکیب

ترکیب، یک عملگر مهم در الگوریتم‌های ژنتیک است که به‌منظور ایجاد تغییر در کروموزوم‌های جمعیت مورداستفاده قرار می‌گیرد [۲۱]. همچنین عملگر ترکیب به تکامل جمعیت در الگوریتم‌های ژنتیک کمک می‌کند.

عملگر ادغام بیش از یک کروموزوم را برای تولید نسل جدید کروموزوم‌ها ترکیب می‌کند. کروموزوم جدید برخی ویژگی‌ها را از والد نخست و بقیه ویژگی‌ها را از والد دوم به ارث می‌برد. الگوریتم ۳ جزئیات عملگر ترکیب الگوریتم پیشنهادی را تشریح می‌کند.

Algorithm 3. Crossover operation
Input: Two parents from the current population. Output: Two new offsprings 1: Choose randomly a suitable crossover point i 2: Cut the father's chromosome and the mother's chromosome into left and right segments 3: Generate a new offspring, namely the son 4: Inherit the left segment of the father's chromosome to the left segment of the son's chromosome 5: Copy genes in mother's chromosome that do not appear in the left segment of father's chromosome to the right segment of son's chromosome 6: Generate a new offspring, namely the daughter 7: Inherit the left segment of the mother's chromosome to the left segment of the daughter's chromosome 8: Copy genes in father's chromosome that do not appear in the left segment of mother's chromosome to the right segment of daughter's chromosome 9: Return the two new offsprings

الگوریتم ۳: ترکیب تک نقطه‌ای

[22]. Fittest

در الگوریتم ۴ عملگر تک نقطه‌ای به‌صورت چارچوب نگاشت-کاهش در آمده است به صورتی که هر عملکر ترکیب به یک نگاشت تخصیص داده می شود.

Algorithm 4. Crossover Mapper
1: Run HDFS engine 2: Assign algorithm 5 to each mapper 3: Run the MapReduce job tracker 4: Start the calculation 5: Write the result into M blocks 6: Send the result to the shuffle phase 7: if all individuals have been processed successfully then Write best individual to global file in DFS 8: Copy and send the result to the Reducer

الگوریتم ۴: ترکیب تک نقطه‌ای با استفاده از چارچوب نگاشت-کاهش

۳-۷-جهش

عملگر جهش در الگوریتم ژنتیک برای نگه داشتن تنوع جمعیت بوسیله تغییر کروموزوم بکار می‌رود. بعد از اعمال عملگر ترکیب به منظور اجتناب از همگرایی به بهینه محلی و ایجاد تنوع و گوناگونی درجمعیت با استفاده از عملگر جهش یک تعداد از کروموزومهای به دست آمده تغییر داده می‌شوند. همچنین دو ژن برای انجام عملیات جهش تعریف می‌شوند. الگوریتم ۵ جزئیات عملگر جهش الگوریتم پیشنهادی را تشریح می‌کند.

Algorithm 5. Mutation operation
Input: A randomly chosen chromosome. Output: A new chromosome. 1: Choose randomly a gene Ti 2: Find the first successor Tj member Succ(i) 3: Choose randomly a gene Tk in the interval [i+1, j-1] 4: if l < i for all Tl member Pred(k) then 5: Generate a new offspring by interchanging gene Ti and gene Tk 6: return the new offspring 7: else 8: Go to Step 1 9: end if

الگوریتم ۵) عملگر جهش

در الگوریتم ۶ عملگر جهش به‌صورت چارچوب نگاشت-کاهش درآمده است به صورتی که هر عملگر ترکیب به یک نگاشت تخصیص داده می‌شود.

Algorithm 6. Mutation Reduce
1: Check the HDFS engine 2: Assign the result of algorithm 6 to each Reducer 3: Run the MapReduce job tracker 4: Start the calculation 5: Write the result into M blocks 6: Check the termination 7: if all individuals have been processed successfully then Write best individual to global file in DFS 8: Send the result to the cluster

الگوریتم ۶) جهش با استفاده از چارچوب نگاشت-کاهش

۳-۸-انتخاب

در الگوریتم پیشنهادی به‌منظور انتخاب بعضی از کروموزوم-ها برای الگوریتم جستجوی محلی از روش انتخاب چرخ رولت با پذیرش اتفاقی[۲۳] استفاده‌شده است [۱۴]. این روش دارای پیچیدگی زمانی o(۱) می‌باشد. این الگوریتم انتخاب دارای دو مرحله اصلی است: در مرحله اول یک کروموزوم به‌طور تصادفی از میان جمعیت انتخاب می‌شود که این انتخاب با احتمال (۱/n) صورت می‌گیرد. در مرحله بعدی مقدار برازندگی کروموزوم ارزیابی‌شده و اگر کروموزوم انتخاب‌شده یکی از نخبه‌های جمعیت باشد برای جستجوی محلی انتخاب می‌شود در غیر این صورت کروموزوم انتخاب‌شده پذیرفته نمی‌شود و تابع انتخاب دوباره از مرحله ۱ تکرار می‌شود. شبه کد روش انتخاب چرخ رولت با پذیرش اتفاقی در الگوریتم ۷ آورده شده است.

Algorithm 7. Roulette-wheel selection via stochastic acceptance
Input: A current population chromosome. Output: A selected chromosome 1:do Generate a random number R∈ [0,PopSize-1]; 2: if $\frac{makespan_R}{\text{makespan}_{min}} == 1$ then 3: return chromosome$_R$; 4: end if 6: while (founding a one of the fittest solution).

الگوریتم ۷: انتخاب چرخ رولت با پذیرش اتفاقی

[23] .Roulette-wheel selection via stochastic acceptance

۹-۳- شرط خاتمه

تفاوت اصلی بین تکامل طبیعت و مسئله در تکامل طبیعی می‌باشد که گونه‌ها در طبیعت تمایل به خاتمه ندارند ولی در حل کردن یک مسئله با توجه به بودجه مشخص‌شده فرآیند را بایستی متوقف کرد [۱۵]. روش‌های معمول مثل تنظیم یک محدودیت بر روی تعداد ارزیابی برازندگی‌ها یا زمان ساعت کامپیوتر و یا دنبال کردن تنوع و توقف آن موقع کم شدن تنوع در جمعیت برای بررسی شرط خاتمه الگوریتم‌ها وجود دارد [۱۲]. در الگوریتم پیشنهادی شرط خاتمه موقعی اتفاق می‌افتد که تمام کروموزوم‌ها یا به عبارتی راه‌حل‌ها به یک مقدار برازندگی یکسان همگرا شوند.

۴-نتایج شبیه سازی

در این قسمت روش پیشنهادی با دو الگوریتم MPQGA و جعفری نویمی پور و همکاران از جنبه‌های زمان کل اتمام کار و مصرف انرژی مقایسه شده است. نتایج آزمایش‌ها مقایسه الگوریتم‌ها با استفاده از گراف جهت‌دار بدون دور و همچنین گراف‌های جهت‌دار بدون دور تولیدشده تصادفی به‌دست‌آمده‌اند.

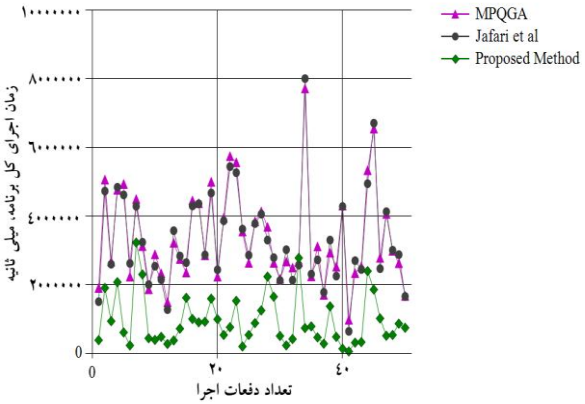
جدول ۴: مشخصات خوشه‌های سخت‌افزاری و نرم‌افزاری در هادوپ

نوع	سیستم‌عامل	پردازنده	حافظه
خوشه ۱	ubuntu-14.04.2	4-core, 3.07 GHZ	4G
خوشه ۲	ubuntu-14.04.2	4-core, 2.07 GHZ	4G
خوشه ۳	ubuntu-14.04.2	4-core, 2.07 GHZ	4G
خوشه ۴	ubuntu-14.04.2	4-core, 2.07 GHZ	4G

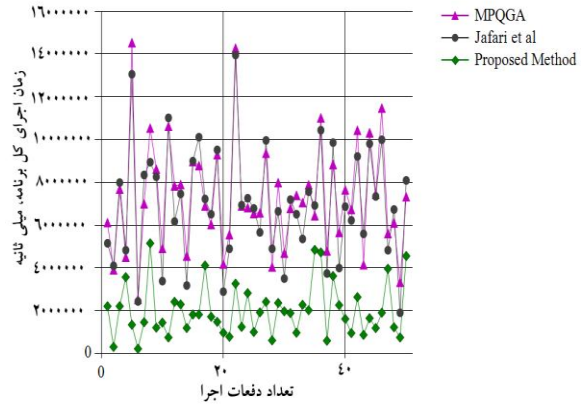
روش پیشنهادی و سایر الگوریتم‌ها در زبان برنامه‌نویسی جاوا در محیط MapR پیاده‌سازی شده‌اند و شبیه‌سازی و اجرای الگوریتم‌ها بر روی چهار خوشه در جدول ۴ با مشخصات زیر اجراشده است.

تولید گراف‌های جهت‌دار بدون دور تصادفی این امکان را می‌دهد که الگوریتم‌های مقایسه‌ای تحت شرایط مختلف بررسی و ارزیابی شوند. الگوریتم پیشنهادی با سایر الگوریتم‌ها ازلحاظ زمان اجرای کل برنامه و مصرف انرژی مقایسه شده است. زمان اجرای کل وظایف از لحظه ورود به سیستم تا زمان پاسخ دهی، چه مدت در سیستم میماند تا پاسخ خود را دریافت کند. مقدار زمان اجرای کل کارها بر حسب میلی ثانیه در نظر گرفته شده است.

هزینه‌های محاسباتی و ارتباطی گراف‌ها به‌صورت تصادفی از میان یک بازه انتخاب می‌شوند. شکل ۳ زمان‌های اجرای کل گراف تصادفی را با ۵۰ زیر وظیفه و شکل ۴ زمان‌های اجرای کل گراف تصادفی را با ۱۰۰ زیر وظیفه را نشان می‌دهند. با توجه به نتایج نمودارها در شکل‌ها، مشهود است که روش پیشنهادی زمان‌های اجرای دو الگوریتم MPQGA و جعفری نویمی پور و همکاران را همواره بهبود می‌بخشد. بهبود زمان‌های اجرای کل گراف‌ها با افزایش تعداد زیر وظایف مشهود است.



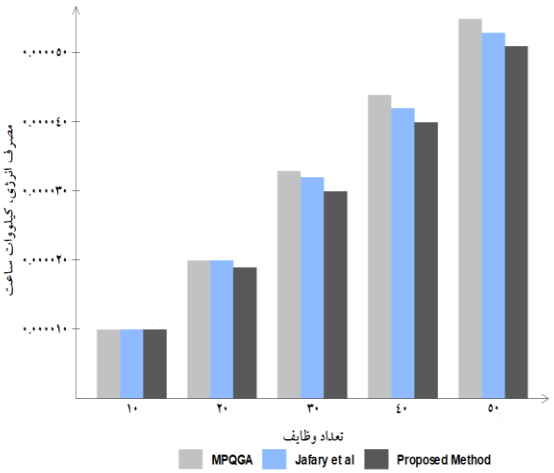
شکل ۳: زمان‌های اجرای کل برنامه با ۵۰ کار



شکل ۴: زمان‌های اجرای کل برنامه با ۱۰۰ کار

سدر روش پیشنهادی با استفاده از هادوپ و مدل نگاشت-کاهش کاهش انرژی مصرفی کل را در پی دارد. مصرف انرژی با دو الگوریتم MPQGA و جعفری نویمی پور و

همکاران مقایسه شده است. شکل ۵ مصرف انرژی با تعداد وظایف مختلف وظیفه را نشان می‌دهند. محور افقی برابر با تعداد وظایف و محور عمودی برابر مصرف انرژی وظایف در نظر گرفته می شود. در هر باراجرا، تعداد وظایف از ۱۰ تا ۵۰ متغیر فرض شده است که در هر بازه که سیستم اجرا میشود، مقادیر متفاوتی به دست میآید که در شکل ۵ قابل مشاهده است.



شکل ۵: مصرف انرژی با تعداد وظایف مختلف

۵-نتیجه گیری و پیشنهاد کارهای آتی

زمان‌بندی در سیستم‌های رایانش توزیع‌شده یک مسئله مهم می‌باشد و الگوریتم‌های زمان‌بندی گوناگونی در این زمینه ارائه‌شده است. در این مقاله یک الگوریتم ژنتیک موازی با استفاده از چارچوب نگاشت-کاهش برای وظایف ایستا در سیستم‌های رایانش ابری ارائه گردید. این الگوریتم با ترکیب الگوریتم‌های ژنتیک استفاده از الگوریتم HEFT برای تخصیص زیروظایف به پردازنده‌ها توسعه داده شد. نتایج بدست آمده بر روی یک گراف ساده و همچنین گراف‌های تولیدشده تصادفی حاکی از آن است که الگوریتم ارائه‌شده از دو الگوریتم MPQGA و جعفری نویمی پور و همکاران در زمینه زمان اجرای کل بهینه عمل کرده است. لازم به ذکر است که هزینه‌های محاسباتی و ارتباطی گراف‌ها به‌صورت تصادفی از بین یک بازه انتخاب می‌شوند از معایب این روش محسوب می‌شود. از طرف دیگر، نگاشت-کاهش برنامه نویسی سطح پایین است و برای بدست آوردن تابع نگاشت و کاهش باید کدهای زیادی استفاده کرد. در این مقاله مقایسه‌های الگوریتم پیشنهادی با سایر الگوریتم‌ها بر روی گراف‌های جهت‌دار بدون دور تصادفی و یک گراف ساده صورت گرفته است. به همین دلیل مطالعات بعدی در مورد الگوریتم پیشنهادی می‌تواند بر روی گراف‌های جهان واقعی مانند FFT[۲۴]، Molecular dynamics code و غیره صورت گیرد. در شیوه نگاشت-کاهش که از سوی شرکت گوگل در سال ۲۰۰۴ معرفی‌شده است به‌طور مختصر همپوشانی در قسمت فازهای نگاشت و ارتباط دیده می‌شود. متأسفانه این همپوشانی[۲۵] در نرم‌افزارهای نوشته‌شده به شیوه نگاشت-کاهش که دارای فاز ارتباط حجیم‌تری نسبت به فاز نگاشت می‌باشند تأثیری چندانی روی کاهش زمان پاسخ این نرم‌افزارها ندارد. اگر بتوان راهکارهایی ارائه دهیم که توانایی همپوشانی فازهای حجیم ارتباط و کاهش را در این‌گونه نرم‌افزارها را داشته باشد می‌توان زمان بحرانی اجرای برنامه را کاهش داد و همچنین از منابع در دسترس حداکثر استفاده را نمود.

[24] .Fast Fourier transformation
[25] .Overlapping

منابع

1. Khemka, B., et al., Utility maximizing dynamic resource management in an oversubscribed energy-constrained heterogeneous computing system. Sustainable Computing: Informatics and Systems, 2015. 5: p. 14-30.
2. Arunarani, A.R., Manjula, D. and Sugumaran, V., 2019. Task scheduling techniques in cloud computing: A literature survey. Future Generation Computer Systems, 91, pp.407-415.
3. Zang, X., Sun, J., Zhao, J. and Zeng, W., 2018, May. Research Progress of Cloud Computing Task Scheduling Technology. In 2018 3rd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2018). Atlantis Press.
4. Young Choon, L. and A.Y. Zomaya, A Novel State Transition Method for Metaheuristic-Based Scheduling in Heterogeneous Computing Systems. Parallel and Distributed Systems, IEEE Transactions on, 2008. 19(9): p. 1215-1223.
5. Keshanchi, B. and Navimipour, N.J., 2016. Priority-based task scheduling in the cloud systems using a memetic algorithm. Journal of Circuits, Systems and Computers, 25(10), p.1650119.
6. Dean, J. and S. Ghemawat, MapReduce: a flexible data processing tool. Communications of the ACM, 2010. 53(1): p. 72-77.
7. Khezr, S.N. and Navimipour, N.J., 2017. MapReduce and its applications, challenges, and architecture: a comprehensive review and directions for future research. Journal of Grid Computing, 15(3), pp.295-321.
8. Topcuoglu, H., S. Hariri, and W. Min-You, Performance-effective and low-complexity task scheduling for heterogeneous computing. Parallel and Distributed Systems, IEEE Transactions on, 2002. 13(3): p. 260-274.
9. Mateos, C., E. Pacini, and C.G. Garino, An ACO-inspired algorithm for minimizing weighted flowtime in cloud-based parameter sweep experiments. Advances in Engineering Software, 2013. 56: p. 38-50.
10. Ashouraie, M. and N. Jafari Navimipour, Priority-based task scheduling on heterogeneous resources in the Expert Cloud. Kybernetes, 2015. 44(10): p. 1455-1471.
11. Wang, W., et al., Cloud-DLS: Dynamic trusted scheduling for Cloud computing. Expert Systems with Applications, 2012. 39(3): p. 2321-2329.
12. Xu, Y., et al., A genetic algorithm for task scheduling on heterogeneous computing systems using multiple priority queues. Information Sciences, 2014. 270(0): p. 255-287.
13. Abed, I., et al., Optimization of the Time of Task Scheduling for Dual Manipulators using a Modified Electromagnetism-Like Algorithm and Genetic Algorithm. Arabian Journal for Science and Engineering, 2014. 39(8): p. 6269-6285.
14. Lipowski, A. and D. Lipowska, Roulette-wheel selection via stochastic acceptance. Physica A: Statistical Mechanics and its Applications, 2012. 391(6): p. 2193-2196.
15. Ong, B. and M. Fukushima, Genetic algorithm with automatic termination and search space rotation. Memetic Computing, 2011. 3(2): p. 111-127.
16. Hashem, Ibrahim Abaker Targio, Nor Badrul Anuar, Mohsen Marjani, Ejaz Ahmed, Haruna Chiroma, Ahmad Firdaus, Muhamad Taufik Abdullah et al. "MapReduce scheduling algorithms: a review." The Journal of Supercomputing (2018): 1-31.
17. Keshanchi B, Souri A, Navimipour NJ. An improved genetic algorithm for task scheduling in the cloud environments using the priority queues: formal verification,

simulation, and statistical testing. Journal of Systems and Software. 2017 Feb 1;124:1-21.

18. Shabestari, F., Rahmani, A. M., Navimipour, N. J., & Jabbehdari, S. (2019). A taxonomy of software-based and hardware-based approaches for energy efficiency management in the Hadoop. Journal of Network and Computer Applications, 126, 162-177.

19. Ashouraei, M., Khezr, S. N., Benlamri, R., & Navimipour, N. J. (2018, August). A new SLA-aware load balancing method in the cloud using an improved parallel task scheduling algorithm. In 2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud) (pp. 71-76). IEEE.

20. Panahi, V., & Navimipour, N. J. (2019). Join query optimization in the distributed database system using an artificial bee colony algorithm and genetic operators. Concurrency and Computation: Practice and Experience, e5218.

21. Mirjalili, S. (2019). Genetic algorithm. In Evolutionary Algorithms and Neural Networks (pp. 43-55). Springer, Cham.

22. Malik, M., Neshatpour, K., Rafatirad, S., Joshi, R. V., Mohsenin, T., Ghasemzadeh, H., & Homayoun, H. (2019). Big vs little core for energy-efficient Hadoop computing. Journal of Parallel and Distributed Computing, 129, 110-124.

فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷
صص: ۸۵- ۱۰۰

یک روش جدید حریصانه مبتنی بر مدل آبشاری برای محاسبه‌ی حداکثر سازی نفوذ در شبکه‌های اجتماعی

* عسگرعلی بویر ** حمید احمدی بنی

* دانشیار، گروه مهندسی کامپیوتر، دانشکده فناوری اطلاعات و مهندسی کامپیوتر، دانشگاه شهید مدنی آذربایجان، تبریز

** کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشکده فناوری اطلاعات و مهندسی کامپیوتر، دانشگاه شهید مدنی آذربایجان، تبریز

تاریخ دریافت:۱۳۹۸/۰۴/۲۶ تاریخ پذیرش: ۱۳۹۹/۰۲/۱۸

چکیده

در مسئله حداکثر سازی نفوذ، هدف یافتن حداقل تعدادی گره هست که بیشترین انتشار و نفوذ را در شبکه داشته باشند. مطالعات راجع به حداکثر سازی نفوذ و انتشار به‌صورت گسترده‌ای در حال گسترش است. در سال‌های اخیر الگوریتم‌های زیادی درزمینهٔ مسئله حداکثر سازی نفوذ در شبکه‌های اجتماعی ارائه‌شده است. این مطالعات شامل بازاریابی ویروسی، گسترش شایعات، اتخاذ نوآوری و شیوع بیماری‌های همه‌گیر و ... است. هر یک از مطالعات پیشین دارای کاستی‌هایی دریافتن گره‌های مناسب و یا پیچیدگی زمانی بالا هستند. در این مقاله، روشی جدید با عنوان ICIM-GREEDY برای حل مسئله حداکثر سازی نفوذ ارائه کرده‌ایم. در الگوریتم ICIM-GREEDY دو معیار مهم که در کارهای انجام‌شده قبلی در نظر گرفته نشده‌اند را در نظر می-گیریم، یکی قدرت نفوذ و دیگری حساسیت به نفوذ. این دو معیار همیشه در زندگی اجتماعی انسان‌ها وجود دارد. روش پیشنهادی روی دیتاست‌های استاندارد مورد ارزیابی قرارگرفته‌شده است. نتایج به‌دست‌آمده نشان می‌دهد که روش مذکور نسبت به دیگر الگوریتم‌های مقایسه شده از کیفیت بهتری در پیدا کردن نودهای بانفوذ در ۳۰ گره Seed برخوردار است. همچنین این روش از لحاظ زمانی نیز نسبت به الگوریتم‌های مقایسه شده به لحاظ همگرایی نسبتاً سریع، بهتر عمل می‌کند.

واژه‌های کلیدی: مدل آبشاری مستقل، حداکثر سازی نفوذ، انتشار، شبکه اجتماعی

۱- مقدمه

مطالعه گسترش نفوذ در شبکه‌های اجتماعی و شبکه‌های پیچیده به‌صورت قابل‌توجهی افزایش پیداکرده است. حداکثر سازی نفوذ برای اولین بار توسط دومنیکس و ریچاردسون در سال ۲۰۰۱ مطرح شد [۱].کمپ و همکارانش برای اولین بار به مسئله حداکثر سازی نفوذ پرداختند [۲]. در شبکه‌های اجتماعی و البته مسئله حداکثر سازی نفوذ به دنبال نودهایی با ضریب نفوذ بالا هستیم که بیشترین تأثیر را بر روی شبکه داشته باشند. برای حداکثر سازی نفوذ دو مدل پایه‌ای آبشاری مستقل و آستانه‌ی خطی معرفی‌شده است. انتشار اطلاعات و نفوذ در زمان و مراحل گسسته انجام می‌شود. هر نود V در گراف G در دو حالت فعال یا غیرفعال است. در حالت فعال یک نود تحت نفوذ قرارگرفته است یعنی یک شایعه یا یک ایده یا یک محصول جدید را پذیرفته است. در حالت غیرفعال ایده یا محصول جدید را نپذیرفته است که ناشی از نپذیرفتن یا نرسیدن ایده یا محصول جدید یا یک شایعه باشد. برای درک بهتر مدل‌های انتشار اغلب از مدل‌های هم‌ارزی استفاده می‌شود. مدل‌های تصادفی را هم-

نویسندهٔ عهده‌دار مکاتبات: عسگرعلی بویر a.bouyer@azaruniv.ac.ir

ارز می‌گویند که دارای توزیع احتمال برای نودهای فعال‌شده توسط نودهای S_0 یکسان باشد. مدل‌های LT و IC دل‌های پایه‌ای درحداکثرسازی نفوذ است [۲]. هرچند در سال‌های اخیر مدل‌های دیگر نیز در این زمینه ارائه‌شده است، ولی پایه و اساس مدل‌های ارائه‌شده، مدل LT و IC است [۳-۹].

مدل آبشاری مستقل برای اولین بار توسط کمپ و همکارانش مطرح شد [۲]. در مدل آبشاری مستقل هرگاه نودی فعال می‌شود با احتمالی می‌تواند نودهای همسایه خود را فعال کند این احتمال برابر با وزن یال ارتباطی u و v یعنی p_{uv} ا . فعال‌سازی تنها یک‌بار صورت می‌گیرد به این معنی که اگر نود v ن انست نود u را فعال کند، دیگر شانسی برای فعال‌سازی نود u ن رد. در مدل آستانه خطی احتمال اثرگذاری نود v ب نود u با وزن بین یال‌ها، که با $b_{v,w}$ نشان می‌دهیم، بیان می‌کنیم. هر نود v دارای یک حد آستانه θ_vاست که به‌صورت تصادفی در بازه [0,1] به‌صورت تصادفی انتخاب می‌شود. حداکثر سازی نفوذ برای هر دو مدل آبشاری مستقل و حد آستانه خطی یک مسئله NP-hardاست [۱۰]. در مدل بشاری مستقل و آستانه خطی (LT) تابع خروجی σ(۰)، خواص زیر بخشی و یکنوایی دارد که σ(۰) تابع گسترش نفوذ است. الگوریتم حریصانه یک رویکرد طبیعی برای انتخاب مجموعه‌ای از گره‌ها است که حداکثر نفوذ را داشته باشند. الگوریتم حریصانه می‌تواند ضریب تقریبی $1-\frac{1}{n}-\varepsilon$ ارائه دهد که n تعداد نودهای گراف است و ε عددی بسیار ناچیز است [۲].

در این مقاله یک الگوریتم جدید به نام ICIM-GREEDY را ارائه می‌دهیم. در مقالات ارائه‌شده در سال‌های اخیر دو نکته بسیار مهم را در نظر گرفته نشده است[۱۱-۱۷]، درصورتی‌که در دنیای واقعی این دو معیار وجود دارد. در دنیای واقعی نفوذ افراد با یکدیگر متفاوت است، افراد در جامعه با توجه به خصوصیات اجتماعی قدرت نفوذ متفاوتی دارند. در این مقاله نفوذ گره‌ها با توجه به دو معیار دنیای واقعی سنجیده می‌شود: قدرت نفوذ و حساسیت به نفوذ. قدرت نفوذ برای نودها متفاوت است به‌عنوان‌مثال فرض کنید یک دانشجو مسلماً دارای درجه بالایی است و استاد دانشگاه نسبت به دانشجو دارای درجه پایین‌تری به دلیل برقراری رابطه با افرادی خاص است. ولی نودهایی که با استاد دانشگاه در ارتباط هستند دارای قدرت نفوذ بیشتری هستند. فرض کنید هر دو نود استاد دانشگاه و دانشجو ۵۰ نود را فعال می‌کنند ولی آیا قدرت نفوذ هر دو یکسان است، مسلماً این‌طور نیست. حساسیت به نفوذ، در نودها می‌تواند متفاوت باشد که یک نود می‌تواند در مقابل نفوذ (اتخاذ محصول جدید یا ایده) مقابله کند یا به‌راحتی آن را بپذیرد. در این مقاله برای محاسبه قدرت نفوذ می‌توانیم از دو فاکتور مهم نودهای حاشیه و نودهای هسته استفاده کنیم. به این صورت نودهایی که در هسته قرار دارند مسلماً قدرت نفوذ بیشتری دارند و نودهایی که در حاشیه قرار دارند قدرت نفوذ کمتری دارند.

ادامه مقاله به این صورت سازمان‌دهی شده است. در بخش ۲ یک دسته‌بندی کلی از الگوریتم‌های ارائه‌شده درزمینهٔ حداکثر سازی نفوذ موردبررسی و ارزیابی قرارگرفته است. در بخش ۳ الگوریتم ICIM-GREEDY با استفاده از فاکتورهای جدید حساسیت به نفوذ و قدرت نفوذ ارائه‌شده است. تحلیل نتایج در بخش ۴ صورت گرفته است و در بخش ۵ به نتیجه‌گیری پرداخته‌ایم.

۲- کارهای انجام شده

حداکثر سازی نفوذ، به‌عنوان یک روش الگوریتمی برای بازاریابی ویروسی برای اولین بار توسط دومینگوس و ریچاردسون در سال ۲۰۰۱ پیشنهاد شده است [۱]. کمپ و همکارانش در سال ۲۰۰۳ برای اولین بار به تدوین و فرموله سازی مسئله‌ی حداکثر سازی نفوذ به‌عنوان یک مسئله بهینه‌سازی تصادفی گسسته پرداختند [۲]. مسئله حداکثر سازی نفوذ را می‌توان طبق رابطه‌ی (۱) مطرح کرد که در یک گراف G=(V,E)، یک مدل تصادفی بر روی G، مجموعه‌ی نودهای اولیه فعال را با $|S| < k$ نشان می‌دهد که هدف آن بیشینه‌سازی $\sigma_G(S^*)$ است که به‌صورت زیر محاسبه می‌شود:

$$\sigma_G(S^*) = argmax_{S \subseteq V \wedge |S|<k} \sigma_G(S) \qquad (۱)$$

وانگ و همکارانش در سال ۲۰۱۰ اثبات کردند که محاسبه گسترش نفوذ$\sigma_G(S)$ در یک گراف اجتماعی

G=(V,E) با مجموعه S_0، برای هر دو مدل LT و IC، NP-hard است [۱۰]. مسئله حداکثر سازی نفوذ یک مسئله سخت است. روش اصلی توسط کمپ و همکارانش ارائه‌شده است که در این روش از فرایند شبیه‌سازی مونت‌کارلو در فرایند انتشار برای محاسبه گسترش نفوذ استفاده می‌شود [۲]. مجموعه تأثیرگذار S برای R تکرار شبیه‌سازی می‌شود. بعد از پایان فرایند انتشار تعداد نودهای فعال شمرده می‌شود و سپس متوسط آن‌ها را در R تکرار محاسبه می‌شود. الگوریتم حریصانه با تعداد زیاد تکرار شبیه‌سازی مونت‌کارلو باعث دستیابی به بهترین نود نفوذپذیر در میان الگوریتم‌های مختلف شود. یک مشکل جدی برای الگوریتم حریصانه مونت‌کارلو MC-Greedy(G,K) ناکارآمدی وقت‌گیر بودن آن است. به همین دلیل لسکوک و همکارانش الگوریتم CELF را مطرح کردند که با استفاده از ارزیابی تنبل، محاسبات گسترش نفوذ را کاهش می‌دهد [۱۸]. ولی همچنان الگوریتم‌های CELF به دلیل استفاده از شبیه‌سازی مونت‌کارلو در تعداد تکرار بالا، ناکارآمد بودند. بدین ترتیب چن و همکارانش الگوریتم NewGreedyIC را مطرح کردند که با استفاده از مجموعه نودهای قابل‌دسترس، گسترش نفوذ را برای هر نود محاسبه می‌کند. همچنین در ادامه، چن و همکاران برای بهبود زمان اجرای NewGreedyIC، الگوریتم DegreeDiscount را مطرح کردند که بعد از انتخاب نودی با بالاترین درجه به‌عنوان اولین نود Seed، برای انتخاب نودهای seed بعدی، درجه‌ی نودهایی که در همسایگی آن‌ها، seed وجود دارد، را کاهش می‌دهد [۱۹]. این الگوریتم تقریب بهینه را برای مسئله حداکثر سازی نفوذ تضمین نمی‌کند ولی زمان اجرای بسیار مناسبی به نسبت الگوریتم‌های Greedy، CELF و NewGreedyIC دارد. سپس، نارایانام و همکارانش الگوریتم SPIN را ارائه دادند که با استفاده از مقدار شیپلی، گسترش نفوذ را برای هر نود محاسبه می‌کند [۲۰]. این الگوریتم زمان اجرای مناسبی دارد ولی تقریب بهینه را در مقایسه با الگوریتم Greedy تضمین نمی‌کند. سپس، کیتساک و همکاران، الگوریتم K-core را مطرح کردند که نودهای هسته و حاشیه را در گراف مشخص می‌کند و نودهای هسته را به‌عنوان نود seed در نظر می‌گیرد [۲۱]. در ادامه، چانگ و همکارانش الگوریتم StaticGreedyDU را برای بهبود زمان اجرای الگوریتم NewGreedyIC مطرح کردند [۲۲]. اما الگوریتم StaticGreedyDU تقریب بهینه را تضمین نمی‌کند. در ادامه برای بالا بردن سرعت محاسبات گسترش نفوذ، گویال و همکارانش الگوریتم SIMPATH را مطرح کردند که با استفاده از محدود کردن شمارش مسیرهای ساده، نودهای seed را انتخاب می‌کند [۲۳]. الگوریتم Simpath زمان اجرای مناسبی برای انتخاب k نود seed دارد و همچنین سربار حافظه مصرفی کمی دارد ولی ضریب تقریب بهینه را تضمین نمی‌کند به همین دلیل الگوریتم MIA توسط وانگ و همکارانش مطرح شد که الگوریتمی مقیاس‌پذیر است و با استفاده از درخت arborescence گسترش نفوذ به‌صورت محلی محاسبه می‌شود [۱۰]. الگوریتم MIA به دلیل اینکه محاسبات گسترش نفوذ را بدون استفاده از شبیه‌سازی مونت‌کارلو انجام می‌دهد، زمان اجرای مناسبی دارد. ولی این الگوریتم سربار حافظه مصرفی بالایی دارد. سپس، چن و همکارانش الگوریتم CIM که مبتنی بر تشخیص جامعه می باشد را مطرح کردند که الگوریتم CIM ابتدا تشخیص جامعه انجام می‌شود و پس‌ازآن نودهای کاندید تولید می‌شوند و از نودهای کاندید نودهای seed انتخاب می‌شوند [۲۴]. هرچند الگوریتم CIM زمان اجرای مناسبی دارد ولی تقریب بهینه را تضمین نمی‌کند. چن و همکارانش الگوریتم LUGreedy را مطرح کردند که در این الگوریتم با توجه به تصادفی بودن احتمال فعال‌سازی، عدم قطعیت را برای مسئله حداکثر سازی نفوذ در نظر می‌گیرد [۲۵]. ژیانگ و همکارانش، الگوریتم VoteRank را که مبتنی بر رأی‌دهی را مطرح کردند [۲۶]. این الگوریتم زمان اجرای مناسبی دارد. همچنین به دلیل اینکه این الگوریتم خاصیت سابماژولاریتی ندارد، تقریب بهینه را تضمین نمی‌کند. سپس، شانگ و همکارانش الگوریتم CoFIM را مطرح کردند که مبتنی بر تشخیص جامعه است [۲۷]. زمان اجرای الگوریتم CoFIM به تعداد نودهای seed وابسته است. سپس، مورنه و همکاران،

الگوریتم CI که مبتنی بر محلی سازی محاسبات گسترش نفوذ بود را مطرح کردند [۲۸]. در این الگوریتم محاسبات گسترش نفوذ در دایره‌ای به شعاع L محدود می‌شود.زمان اجرای الگوریتم CI به L و تعداد نودهای seed، وابسته است. همچنین، لئو همکارانش الگوریتم LIR که مبتنی بر روش‌های اکتشافی بود را مطرح کردند [۲۹]. در الگوریتم LIRمقدار LI برای هر نود بر اساس درجه همسایگان محاسبه می‌شود و سپس، مجموعه نودهای با کمترین مقدار LI بر اساس درجه به‌صورت نزولی مرتب می‌شوند و k نود به‌عنوان seed انتخاب می‌شوند. این الگوریتم، زمان اجرای مناسبی دارد ولی تقریب بهینه را تضمین نمی‌کند. سپس، اهجام و همکارانش الگوریتم HybridRank را مطرح کردند که نودهای تأثیرگذار بر اساس مرکزیت بردار ویژه و coreness انتخاب‌شده‌اند [۳۰]. در این الگوریتم، در انتخاب نودهای تأثیرگذار از پدیده‌ی Rich club اجتناب می‌شود. اگرچه این الگوریتم از پدیده‌ی Rich club اجتناب می‌کند، اما نودهای تأثیرگذار گسترش نفوذ مناسبی ارائه نمی‌دهند. نگوین و همکارانش، الگوریتم probDeg را مطرح کردند که نودهای بانفوذ را با بررسی چند گام از نود و درجه همسایگان آن نود انتخاب می‌کند [۳۱]. سپس، وو و همکارانش، الگوریتم زمان خطی LAIM را مطرح کردند که این الگوریتم یک رویکرد تکراری خطی برای مسئله حداکثر تأثیرگذاری در شبکه‌های بزرگ است [۳۲]. همچنین، این الگوریتم سربار حافظه‌ی مصرفی پایینی، در شبکه‌های اجتماعی با مقیاس بزرگ دارد. سپس، بنرجی و همکارانش الگوریتم ComBIM را مطرح کردند که نودهای تأثیرگذار را با توجه به بودجه جامعه، انتخاب می‌کند [۳۳]. همچنین، این الگوریتم تقریب بهینه را تضمین نمی‌کند. در ادامه، ژی و همکارانش، الگوریتم IRR را بر مدل MBIC مطرح کردند [۳۴]. در این الگوریتم انتشار شامل دو مرحله‌ی شروع و مرجع است. این الگوریتم، گسترش نفوذ بهتری به نسبت الگوریتم DegreeDiscount دارد. روی و همکارانش، الگوریتم RNR را مطرح کردند [۳۵]. این الگوریتم، با استفاده از تأثیر همسایگان، قدرت نفوذ نود را محاسبه می‌کند. در ادامه، کیو همکارانش، الگوریتم PHG را مطرح کردند که در این الگوریتم از رویکرد مبتنی بر تشخیص جامعه استفاده می‌کند. همچنین در این الگوریتم برای محاسبات گسترش نفوذ، از الگوریتم حریصانه ویژگی گراف استفاده می‌کند [۳۶]. همچنین در این الگوریتم، میزان گسترش نفوذ و زمان اجرا، به تعداد نودهای تأثیرگذار که توسط الگوریتم حریصانه انتخاب می‌شوند، وابسته است. پس‌ازآن، احمدی بنی و همکارانش، الگوریتم TI-SC را مطرح کردند که در این الگوریتم با استفاده از نودهای هسته در جوامع الگوریتم PHG را بهبود می‌بخشد [۱۱].

۳- روش پیشنهادی

در این مقاله الگوریتم جدیدی با نام ICIM-GREEDY ارائه‌شده است. در الگوریتم ICIM-GREEDY از مدل WC بهبودیافته استفاده شده است. در این الگوریتم ابتدا وزن نفوذ را محاسبه می‌شود و سپس محاسبات نفوذپذیری را بر اساس دو فاکتور مهم، قدرت نفوذ و حساسیت به نفوذ انجام می‌شود.

۳-۱- وزن نفوذ

وزن نفوذ، یک فاکتور مهم برای تشخیص نفوذ است، به همین دلیل از یک روش جدید برای محاسبه وزن نفوذ استفاده‌شده است. در مدل WC، وزن نفوذ، $\frac{1}{degree(v)}$ محاسبه می‌شود. اما این یک روش منطقی برای تعیین وزن نفوذ نیست به دلیل اینکه نودها دارای ویژگی‌های متفاوتی مثل ضریب خوشه‌بندی متفاوت، درجه همسایه‌های یک نود، پل محلی بودن یا در مجاورت پل محلی بودن و .. هستند، پس باید بین وزن‌های نفوذ در بین نودهای مختلف، تفاوت قائل شویم. بنابراین وزن نفوذ طبق فرمول زیر محاسبه می‌شود:

$$b_{uv} = \frac{D_u}{(DD_u log DD_u + CC_u) + 1} \quad (۲)$$

D_u درجه نود u است و D_u درجه نودهای همسایه نود u است و CC_u ضریب خوشه‌بندی نود u است. وزن نفوذ، به این صورت تعیین می‌شود هر چه درجه همسایگان و درجه همسایگان همسایگان نود v بیشتر باشد، نودهای

همسایه v می‌توانند نفوذ بیشتری بر روی v داشته باشند و همچنین هرچه همسایه‌های نود v دارای ضریب خوشه‌بندی کمتر باشند، تمایل بیشتری به برقراری ارتباط دارند. پس مجموع وزن نفوذ که یک گره u می‌تواند به همسایگان غیرفعال اعمال کند با استفاده از فرمول زیر محاسبه‌شده است:

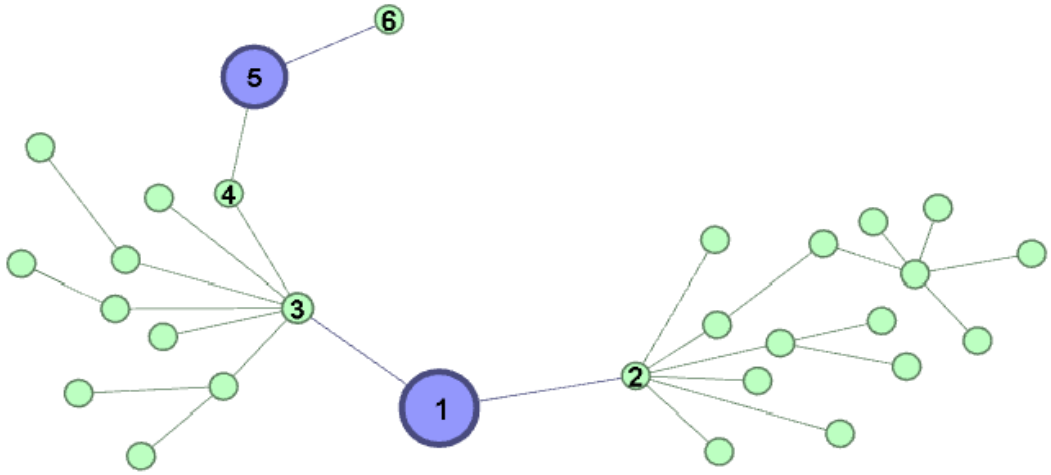
$$f(u) = \sum_{v \in N(u), v \notin A(u)} b_{uv} \quad (۳)$$

$A(u)$ مجموعه‌ای از نودهای فعال در میان همسایگان u است. در دنیای واقعی انتشار و نفوذ تحت تأثیر تمامی همسایگان نود است و در صورتی نودی می‌تواند فعال شود که همسایگان فعال نود بتواند نود موردنظر را تحت تأثیر قرار دهند. به همین دلیل معیار $f(u)$ به‌صورت مجموع نفوذ همسایگان محاسبه می‌شود. بنابراین طبق رابطه‌ی زیر احتمال نفوذ به‌صورت زیر محاسبه می‌شود:

$$p = 1 - e^{-f(u)} \quad (۴)$$

۲-۳- حساسیت به نفوذ

علاوه بر قدرت نفوذ، معیار حساسیت به نفوذ برای محاسبه نفوذ در نظر گرفته‌شده است. حساسیت به نفوذ مسئله‌ای بسیار مهم در حداکثر سازی نفوذ است زیرا در دنیای واقعی حساسیت به نفوذ متفاوت است ولی در تمامی کارهای قبلی حساسیت به نفوذ به‌صورت تصادفی تعیین‌شده است . برای نشان دادن اهمیت این موضوع، نگاهی به شکل ۱ داشته باشیم. همان‌طور که در این شکل می‌بینید، درجه نود ۱و نود ۵، برابر است ولی اهمیت نود ۱ بیشتر از نود ۵ است زیرا گره‌های مهم‌تری به نود ۱ متصل هستند. ممکن است بر اساس آستانه‌ی نفوذ تصادفی، نود ۵ فعال شود که می‌تواند نودهای کمتری را در ادامه فعال کند درصورتی‌که با فعال شدن نود ۱ ممکن است نودهای بیشتری را نسبت به نود ۵ فعال کند. در کارهای قبلی آستانه‌ی نفوذ به‌صورت تصادفی تعیین می‌شود، ولی در این الگوریتم آستانه نفوذ، بر اساس اهمیت نفوذ نود و نودهای پل/پل محلی، تعداد نودهای فعال‌شده، طول مسیر از مبدأ و همسایگان فعال نشده تعیین می‌شود.



شکل ۱ - گره‌های ۱و ۲ درجه یکسانی دارند ولی گره ۱ دارای اهمیت نفوذ بیشتری است.

اساس این ایده، آزمایش میلگرم است آزمایشی برای سنجش میزان فرمانبداری افراد از اقتدار[۱] در انجام کارهایی

مغایر با وجدان شخصی افراد بود که به ابتکار روانشناس معروف استنلی میلگرام صورت گرفت[۳۷]. آزمایش‌هایی که توسط میلگرام صورت گرفت نشان داد که حضور افراد دیگری که از فرمان‌ها پیروی نمی‌کردند، میزان فرمان‌برداری شرکت‌کننده را به نحو زیادی کاهش می‌دهد. وقتی کسانی دیگری در کنار شرکت‌کننده حضور داشتند که از اجرای فرمان‌ها سرباز می‌زدند، ۳۶ نفر از ۴۰ شرکت‌کننده از دادن ماکزیمم شوک خودداری کردند- [۳۷]. بر همین اساس، هرچه تعداد نودهای فعال‌شده با ایده یا محصول جدید، بیشتر باشند حساسیت به نفوذ در نود کمتر می‌شود و این باعث می‌شود ایده یا محصول جدید به‌راحتی پذیرفته شود.

برای محاسبه حساسیت به نفوذ فاکتورهای زیر بسیار مهم می‌باشند:

- تعداد نودهای فعال‌شده از همسایگان نود: هرچقدر تعداد نودهای فعال‌شده بیشتر باشند حساسیت به نفوذ پایین‌تر است.
- فاصله از نود مبدأ: هرچقدر فاصله نود از منبع شایعه دورتر باشد، حساسیت به نفوذ آن بیشتر می‌شود.(فاصله از نود تأثیرگذار)
- تعداد نودهای همسایگان فعال نشده از همسایگان نود: هرچقدر تعداد نودهای فعال نشده به نسبت تعداد نودهای فعال‌شده کمتر باشد، حساسیت به نفوذ نود کمتر می‌شود.
- تعداد نودهای فعال همسایه‌ی یک نود فعال: این فاکتور برای بیان اهمیت نودهای پل/پل محلی، مطرح‌شده است. زیرا این گره‌ها، نقش مهمی در فرایند انتشار اطلاعات دارند. هرچقدر تعداد نودهای فعال همسایه‌ی یک نود فعال، بیشتر باشد حساسیت به نفوذ پایین‌تر است.

بنابراین برای محاسبه وزن نفوذ از رابطه‌ی زیر استفاده می‌کنیم:

$$\theta_v = 1 - e^{-\left(\frac{\sum_{i \epsilon NG_{inactive}, v \notin A(u)} K_i}{NG_{active} log NG_{active} * NNG_{active}}\right)} \quad (۵)$$

در رابطه ۵، $NG_{inactive}$تعداد نودهای همسایگان فعال نشده نود v است. NG_{active} تعداد نودهای فعال‌شده از همسایگان نود v است. NNG_{active} تعداد نودهای فعال همسایه‌ی نودهای فعال، v است. k_i درجه نودهای همسایگان نودv ، که فعال نشده‌اند، است. درواقع در این فرمول، اهمیت نودهای پل/پل محلی به نحو مناسبی در نظر گرفته شده است.

۳-۳- قدرت نفوذ

در رویکرد پیشنهادی، برای اینکه بین نفوذ نودهای مختلف تفاوتی وجود داشته باشد از مفهوم جدید توپولوژی شبکه که بر اساس تجزیه k-shell انجام خواهد شد، استفاده‌شده است. انتشار اطلاعات و الگوی رفتاری، وابسته به ساختار و توپولوژی شبکه است. به‌عنوان‌مثال، اگر یک hub در انتهای شبکه وجود داشته باشد، آن ممکن است حداقل گسترش را داشته باشد، درحالی‌که یک نود با کمترین اتصال در هسته شبکه، ممکن است گسترش قابل‌توجهی داشته باشد. در توپولوژی شبکه، موقعیت یک نود با استفاده از تجزیه K-Shell تعریف می‌شود. در k-shell ابتدا همه‌ی نودها با درجه k=1 حذف می‌شود، پس از هرس مرحله اول ممکن است گره‌هایی جدید با k=1 ظاهر شود. دوباره تمامی نودهای با درجه K=1 حذف می‌شوند. برای نودهای با درجه‌های متفاوت این روش ادامه می‌یابد. بنابراین نودهایی با درجه پایین در حاشیه و نودهایی با درجه بالا، در هسته قرار می‌گیرند. همان‌طور که در شکل ۲ قابل مشاهده است نود ۸ و ۹ درجه‌شان ۸ است ولی نود۹ به دلیل اینکه در هسته قرار دارد، گسترش نفوذ بهتری نسبت به نود ۸ دارد. با توجه به نتایج آزمایش‌ها، زمانی که یک نود اولیه را به‌صورت تصادفی فعال می‌شود، نودهایی که در لایه‌هایی با k_s

[1] authority

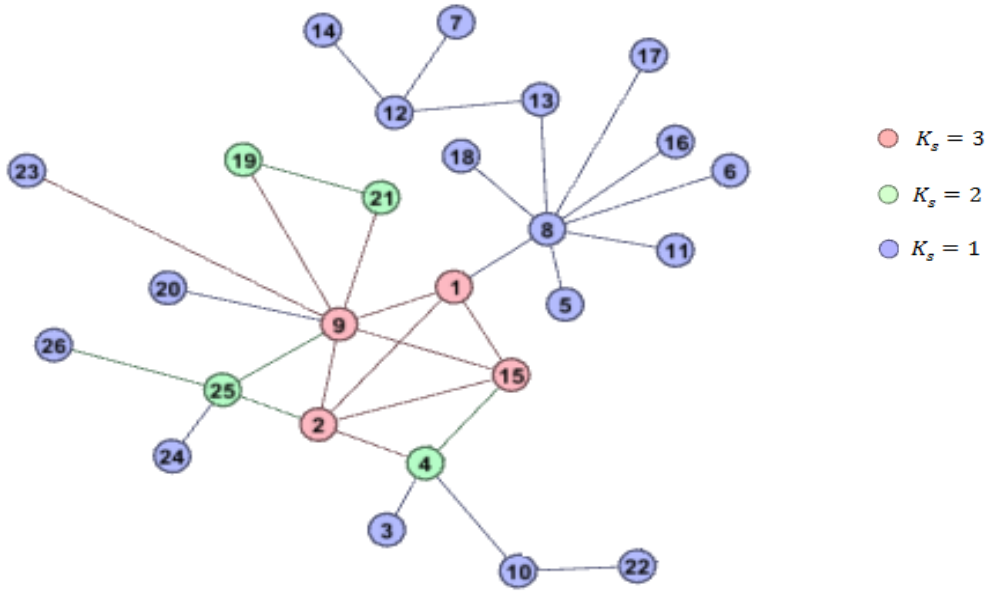
بالاتر قرار دارند، احتمال بیشتری دارند که فعال شوند. البته زمانی موقعیت نود مهم است که به‌عنوان یک نود منبع در نظر گرفته شود. بنابراین، نودهایی که در هسته-های داخلی (k_s بزرگ‌تری دارند) قرار دارند، نفوذ بیشتری دارند.

به دلیل موارد فوق، دو فاکتور k و k_s برای محاسبه قدرت نفوذ در نظر گرفته‌شده است. برای هر مجموعه از نودهای تأثیرگذار S، $\sigma(u|S) = \sigma(S \cup \{u\}) - \sigma(S)$ است. برای محاسبه قدرت نفوذ زمانی که $k > 1$ از فرمول زیر استفاده می‌شود:

$$PID(u) = \frac{\text{count} * K_s + 1}{N(k_s)} \qquad (۶)$$

در این رابطه، Count تعداد نودهای فعال‌شده است. $N(k_s)$ تعداد نودها در هر k_s است. علاوه بر این، در ابتدای کار نفوذ همه نودها برابر با1 در نظر گرفته‌شده است، به این معنی که اگر آن گره فعال شود، حداقل خودش را فعال کرده و یک گره فعال را نتیجه می‌دهد. مقادیر PID در یک بردار ذخیره می‌شود که قدرت نفوذ را برای هر نود مشخص می‌کند.

البته برای محاسبه قدرت نفوذ می‌توان مشخصه‌های رفتاری و یا غیر ساختاری مانند سن، جنسیت و زمان تبلیغات و غیره نیز در نظر گرفت. ولی در نظر گرفتن این مشخصه‌ها در شبکه اجتماعی کمی مشکل است زیرا در اکثر مواقع به دلایلی مانند حفظ حریم خصوصی این مشخصه‌ها قابل‌دسترس نیستند.



شکل ۲- مثالی از روش k-Shell. گره ۸ و ۹ هر دو درجه‌شان ۸ است ولی گره ۹ به دلیل اینکه در هسته قرار دارد، می‌تواند گسترش نفوذ بیشتری داشته باشد.

Algorithm 1:ICIM-GREEDY(G,k): Improved Cascade model for Influence Maximization.

Input: G: Social graph, k: size seed set, R=1

1:**for** j=0 **to** R **do**
2: **if**($v\ is\ active.$)
3: **for** 1 **to** H **do**
4: $theta[e.v] = 1 - e^{-(\frac{\sum_{i\epsilon NG_{inactive}, v\notin A(u)} K_i}{NG_{active} log NG_{active} * NNG_{active}})}$
5: $p[e.v] = 1 - e^{-f(u)}$
6: **if**$((theta[e.v] - p) < 0)$ **then**
7: n_a ←number of active node
8: count←count+n_a
9: **end if**
10: **end for**
11: **end if**
12:**end for**
13: **return** $PID = \frac{k_s * count}{N(k_s)}$

۳-۴- الگوریتم حریصانه

ما می‌توانیم به‌سادگی برای انتخاب مجموعه تأثیرگذار با اندازه **k**، الگوریتم ۲ را k بار فراخوانی کنیم. به دلیل اینکه الگوریتم حریصانه عمومی بسیار وقت‌گیر است، از الگوریتم NewGreedyWC استفاده‌شده است. در الگوریتم ICIM-GREEDY وزن نفوذ و حد آستانه به‌صورت تصادفی انتخاب نمی‌شوند به همین دلیل از شبیه‌سازی مونت‌کارلو استفاده نمی‌شود (به‌عبارتی‌دیگر R=1 است). به همین دلیل الگوریتم NewGreedyWC نسبت به الگوریتم حریصانه عمومی از لحاظ زمان اجرا بهتر می‌شود. در این الگوریتم باید توجه شود، نودهایی که گسترش نفوذ مناسبی ندارند، باعث ایجاد سربار محاسباتی می‌شوند. به همین دلیل برای کاهش سربار محاسباتی گسترش نفوذ باید نودهایی که گسترش نفوذ مناسبی ندارند در محاسبات گسترش نفوذ نادیده گرفته شوند. بدین ترتیب، برای محاسبات گسترش نفوذ، $\frac{m}{\sqrt{n}}$ نود با بالاترین درجه در نظر گرفته می‌شوند که m تعداد یال‌های گراف و n تعداد نودهای گراف است. در مدل WC گراف جهت‌دار در نظر گرفته می‌شود. اگر d_v درجه v در گراف G و $\overline{uv}$ یال در گراف باشد. در این مدل، اگر نود u فعال‌شده باشد، آنگاه نود v با احتمال $\frac{1}{d_v}$ در گام $i+1$ ام، فعال می‌شود [۱۹]. در خط ۲ از الگوریتم ۲ ابتدا الگوریتم ICIM-GREEDY فراخوانی می‌شود. در خط ۴ -۵ الگوریتم به این صورت است که برای هر یال $\overrightarrow{uv} \in E$، با احتمال $1-p$ از G پاک می‌شود. RanWC(G) نشان‌دهنده این فرایند است و $G' = \text{RanWC(G)}$ نتایج را بر روی گراف جهت‌دار نشان می‌دهد. در خط ۶-۱۱ در الگوریتم ۲ از الگوریتم کوهن برای تخمین تعداد نودهای قابل‌دسترس از هر نود استفاده‌شده است که در اولین گام در گراف G' تمام کامپونت‌های قویاً متصل را محاسبه‌شده است. $G^{*\prime}$ گراف جهت‌دار غیر مدور(DAG) است و V^* نشان‌دهنده مجموعه نودهای $G^{*\prime}$ است.

الگوریتم ۱ تنها یک‌بار اجرا می‌شود تا بردار PID را مشخص کند. بردار PID قدرت نفوذ را برای هر نود ارائه می‌دهد. بردار PID به‌عنوان ورودی برای الگوریتم NewGreedyWC است. بنابراین پیچیدگی کلی الگوریتمNewGreedyWC برابر است با $O(kTn)$ بطوری‌که k تعداد نودهای seed است و m تعداد یال‌ها است. در واقع T=5 است زیرا برآورد خوبی از گسترش نفوذ را می‌دهد.

برای محاسبه بردار PID نیاز به محاسبه $N(k_s)$ و k_s است که این دو مقدار با استفاده از الگوریتم *K-Shell* محاسبه می‌شود [38]. پیچیدگی زمانی *K-shell*، $O(m)$ است [38]. به دلیل اینکه به‌صورت پویا

تعداد نودهای همسایه‌ی فعال، تعداد نودهای همسایه‌ی غیرفعال و تعداد نودهای همسایگان همسایگان نود فعال محاسبه می‌شود، پیچیدگی محاسبات $O(2n)$ است که n تعداد نودها است. پیچیدگی الگوریتم۱، $O(2m+2n)=O(m+n)$ می‌شود. مرتبه زمانی کلی الگوریتم به‌صورت $O(kTn+2m+2n)$ است که همان‌طور که در بالا اشاره شد T=5 است پس پیچیدگی زمانی کلی به‌صورت $O(kn+2m+2n)=O(kn+m)$ می‌شود.

Algorithm 2 NewGreedyWC(G , k)

Input: G: Social graph, k: size seed set
Output: selected seed set
1: initialize $S = \emptyset$
2: call algorithm ICIM-GREEDY
3: **for** i= 1 to k **do**
4: set $s_v = 0$ for all vertices.
5: obtain $G' = RanWC(G)$
6: compute DAG G'^* and weights $w(v^*)$ for all $v^* \in V^*$
7: **for** $\ell = 1$ to T **do**
8: for each $v^* \in V^*$, generate random value $X_{v^*}^{\ell}$ from The exponential distribution with mean $1/w(v^*)$
9: **for each** $v^* \in V^*$, compute $Y_{v^*}^{\ell}=min_{u^* \epsilon R_{G'^*}(S^* \cup \{v^*\})} X_{v^*}^{\ell}$
10: for each $v^* \in V^*, s_{v^*}^{\ell} += Y_{v^*}^{\ell}$
11: **end for**
12: for each $v \in V \backslash S, s_v += (T-1)/s_{v^*}^{\ell}$
13: **end for**
15: $S = S \cup \{argmax_{v \in V \backslash S}\{s_v\}$
16: **end for**
17: output S

۴- آزمایش‌ها و ارزیابی

ما ابتدا الگوریتم ICIM-Greedy را بر روی ۸ مجموعه داده (دیتاست) مورد ارزیابی قرار می‌دهیم. دیتاست‌ها بدون جهت می‌باشند. اندازه‌ی دیتاست‌ها بزرگ و متوسط می‌باشند. هر ۸ دیتاست از سایت KONECT دانلود شده-اند.

NetHEPT: این شبکه از بخش "High Energy Physics-Theory" است که مقالات مربوط سال ۱۹۹۱ تا ۲۰۰۳ است هر نود در شبکه یک نویسنده را ارائه می‌دهد و یال‌های بین دو نود برابر با همکاری دو نویسنده در یک مقاله است [۱۹]. این شبکه شامل 15k نود و 58k یال است.

Physics: این شبکه شامل تمامی مقالات بخش "Physics" است. در شبکه Physics نیز هر نود در شبکه یک نویسنده را ارائه می‌دهد و یال‌های بین دو نود برابر با همکاری دو نویسنده در یک مقاله است [۱۹]. این شبکه شامل 15k نود و 231k یال است.

Email: دیتاست Email شبکه‌ای از ایمیل در دانشگاه Rovira i Virgili است که اگر افراد به یکدیگر ایمیلی ارسال کرده باشند، یال بدون جهت بین ایمیل‌ها تشکیل می‌شود [۳۹]. شبکه شامل 1k نود و 5k یال است.

Netscience : این یک شبکه از همکاری‌های نویسندگان در حوزه علوم شبکه است[۳۹]. این شبکه شامل 1k نود و 2k یال است.

M-Fo115: شبکه M-Fo115 از مدل جنگل آتش با احتمال اتصال p = 0.115 تولیدشده است که شامل 10k نود و 23k یال است [۴۰].

M-Fo120: با استفاده از مدل جنگل آتش، شبکه M-Fo120 با احتمال p=0.120 که شامل 10k نود و 29k ساخته می‌شود[۴۰].

M-Fo-0.1: شبکه M-Fo-0.1 از مدل جنگل آتش با احتمال اتصال p = 0.1 تولیدشده است که شامل 5k نود و 7k یال است [۴۰].

M-Fo-0.2: با استفاده از مدل جنگل آتش، شبکه M-Fo-0.2با احتمال p=0.120 که شامل 5k نود و 23k ساخته می‌شود [۴۰].

جدول ۱- مشخصات دیتاست‌های موردنظر برای آزمایش

Data set	NetHEPT	Physics	M-Fo115	M-Fo120	M-Fo-0.1	M-Fo-0.2	Email	NetScience
Node	317k	6k	10k	10k	5k	5k	1k	1k
Edge	1M	13k	23k	29k	7k	23k	5k	2k
Max Degree	343	1459	155	229	20	336	209	43
Min Degree	1	1	1	1	1	1	1	1

۴-۱- پارامترهای مورداستفاده در آزمایش:

الف: تولید احتمال انتشار. به‌صورت کلی بر روی الگوریتم‌های مقایسه شده توزیع احتمال انتشار به‌صورت توزیع یکنواخت است، که از مدل زیر برای تولید این احتمال یکنواخت استفاده می‌کنیم:

ب: استفاده از مدلWC. مدل آبشاری وزن‌دار توسط کمپ و همکارانش در سال ۲۰۰۳ پیشنهادشده است [۱۹].احتمال بر روی یال‌ها به‌صورت $\frac{1}{\deg(v)}$ است که $\deg(v)$ درجه نود v است. گراف انتشار بر روی این مدل به‌صورت جهت‌دار است به دلیل اینکه احتمال انتشار $\frac{1}{\deg(v)}$ از یال (u,v) به یال (v,u) متفاوت است و تعداد تکرار شبیه‌سازی مونت‌کارلو ۲۰۰۰۰ تکرار است.

۴-۲- الگوریتم‌های مورد مقایسه

الگوریتم ICIM-GREEDY را با دو نوع از الگوریتم‌های حریصانه و هیوریستیک مقایسه کردیم. لیستی از الگوریتم‌های مقایسه شده را در زیر ارائه داده‌ایم:

- ICIM-GREEDY: الگوریتم ۱
- Greedy: الگوریتم حریصانه معمولی که از شبیه‌سازی مونت‌کارلو استفاده می‌کند [۱۹].
- MixedGreedy: ترکیبی از الگوریتم حریصانه که در دور اول از NewGreedy و در دور دوم از CELF استفاده می‌کند، که R=20000 وT=5 است[۱۹].
- Degree: الگوریتم اکتشافی که k نود با بالاترین درجه را انتخاب می‌کند [۱۹].
- SingleDiscount: الگوریتم اکتشافی ساده است وقتی‌که هر یک از همسایه‌های که به‌عنوان تأثیرگذار انتخاب می‌شوند درجه کاهش آن یک است [۱۹].
- Distance: یک الگوریتم اکتشافی ساده است که k نود با کوچک‌ترین میانگین فاصله را انتخاب می‌کند [۵].
- Random: الگوریتمی است که k نود را به‌صورت تصادفی انتخاب می‌کند [۵].
- LIR: الگوریتمی که با استفاده از توپولوژی شبکه k نود را انتخاب می‌کند [۲۹].
- ProbDeg: الگوریتمی که با تأثیر درجه همسایگان، k نود را انتخاب می‌کند [۳۱].
- TI-SC: الگوریتمی که با استفاده از تشخیص جامعه، k نود را انتخاب می‌کند [۱۱].

۴-۳- فرایند آزمایش

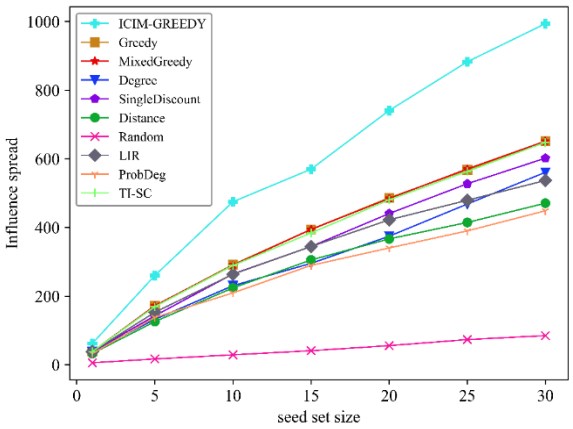
در این مقاله ، آزمایش شبیه‌سازی توسط پایتون را بر روی سروری با پردازنده‌ی Intel Xeon E5410 و 32 گیگابایت حافظه انجام‌شده است که شش مجموعه داده شبکه اجتماعی و ده الگوریتم مقایسه اتخاذشده است.

۴-۴ نتایج

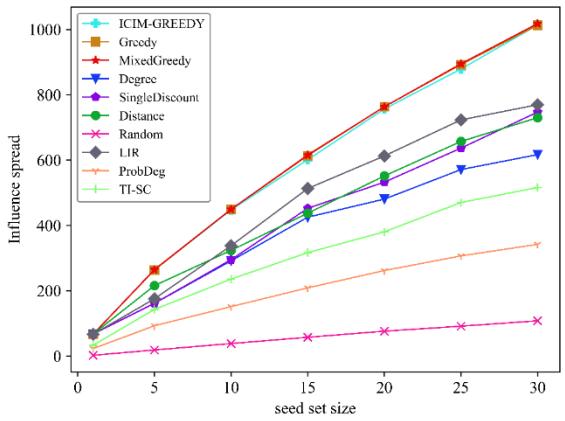
برای این الگوریتم‌ها گسترش نفوذ را برای مجموعه تأثیرگذار در محدوده‌ی ۱تا ۳۰ مقایسه می‌کنیم. همچنین زمان اجرا را برای k=30 مقایسه می‌کنیم.

گسترش نفوذ و زمان اجرا. نتایج گسترش نفوذ بر ۸ دیتاست بررسی‌شده است. اندازه‌ی مجموعه تأثیرگذار رنجی بین ۱ تا ۳۰ است. شکل‌های ۳ تا ۱۰ نتایج گسترش نفوذ را بر ۸

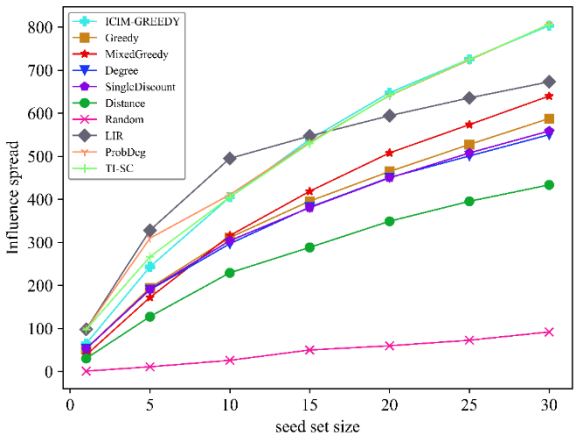
دیتاست با مدل WC را نشان می‌دهد و شکل ۴ زمان اجرای الگوریتم‌های متفاوت را نشان می‌دهد. . در این آزمایش، هدف اصلی مقایسه الگوریتم‌های حریصانه با روش پیشنهادی ICIM-GREEDY می باشد زیرا روش‌هایی مانند Random و Degree به خاطر ضعف بارز در گسترش نفوذ به دلیل عدم انتخاب آگاهانه، الگوریتم‌های خوبی نیستند. بر روی دیتاست NetHEPT (شکل ۳) الگوریتم ICIM-GREEDY گسترش نفوذ بهتر و چشمگیرتری نسبت به بقیه‌ی الگوریتم‌ها دارد و Random بدترین گسترش نفوذ را دارد. در دیتاست Physics (شکل ۴) الگوریتم‌های Greedy و MixedGreedy و ICIM-GREEDY گسترش نفوذ بهتری نسبت به بقیه دارند که هر سه ازنظر گسترش نفوذ تقریباً یکسان و بعضاً با اختلاف بسیار ناچیز ظاهرشده‌اند. در دیتاست Physics، روش Random بدترین گسترش نفوذ را دارد. همچنین، شکلهای ۵ تا ۱۰ نیز نشان می دهد که روش پیشنهادی بر روی مابقی دیتاست‌ها از نظر گسترش نفوذ، نتایج بهتری نسبت به الگوریتم‌های دیگر دارد.



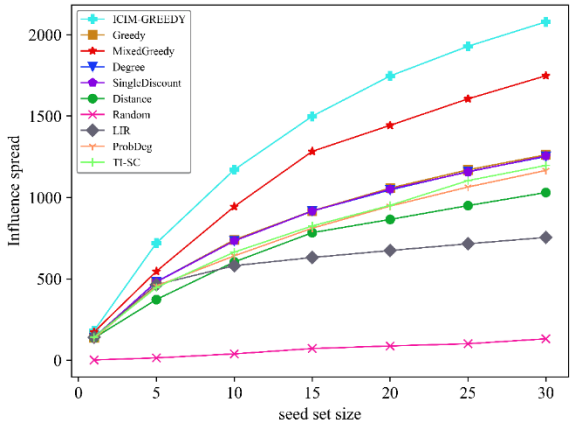
شکل ۳- گسترش نفوذ بر روی الگوریتم‌های متفاوت بر روی NetHEPT



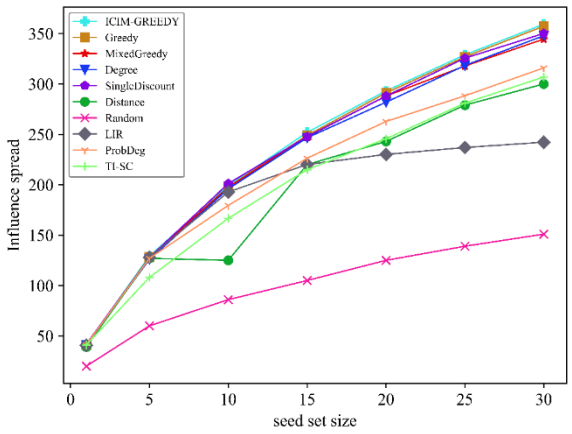
شکل ۴- گسترش نفوذ بر روی الگوریتم‌های متفاوت بر روی Physics



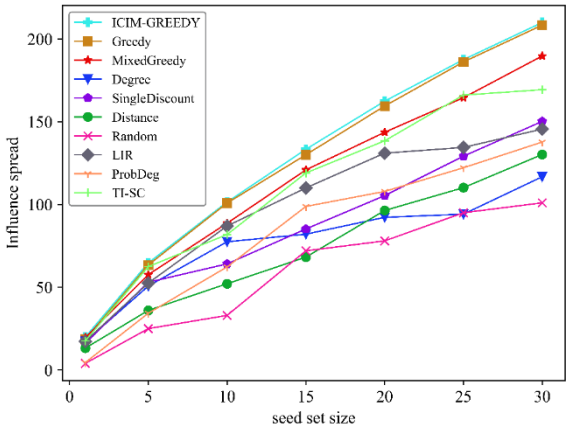
شکل ۵- گسترش نفوذ بر روی الگوریتم‌های متفاوت بر روی M-Fo115



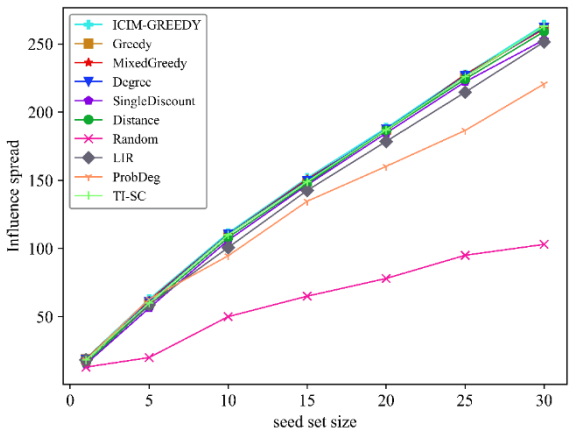
شکل ۶- گسترش نفوذ بر روی الگوریتم‌های متفاوت بر روی M-Fo120



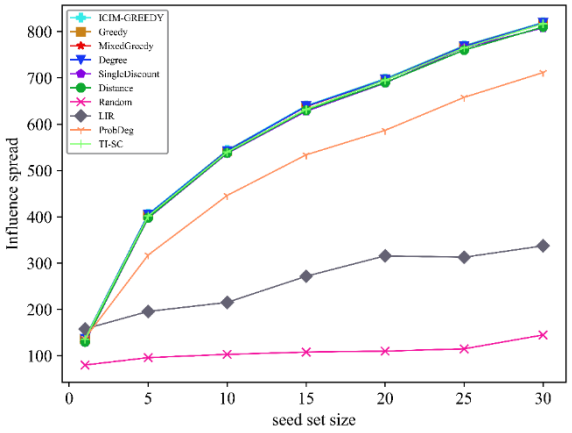
شکل ۷- گسترش نفوذ بر روی الگوریتم‌های متفاوت بر روی Email



شکل ۸- گسترش نفوذ بر روی الگوریتم‌های متفاوت بر روی NetScience



شکل ۹- گسترش نفوذ بر روی الگوریتم‌های متفاوت بر روی M-Fo-0.1

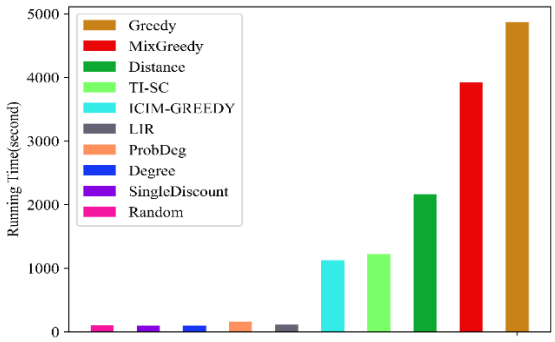


شکل ۱۰- گسترش نفوذ بر روی الگوریتم‌های متفاوت بر روی M-Fo-0.2

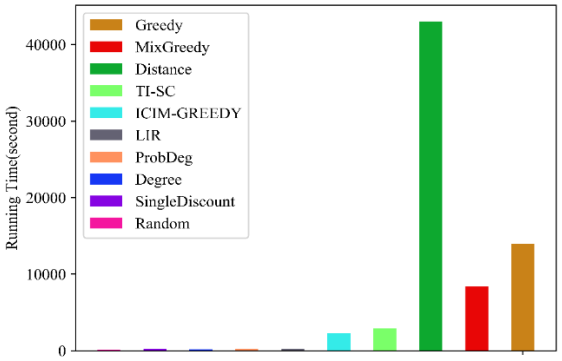
در شکل ۷ گسترش نفوذ بر دیتاست Email مورد بررسی قرارگرفته است که الگوریتم‌های ICIM-GREEDY و Greedy گسترش نفوذ تقریباً نزدیکی دارند و به نسبت دیگر الگوریتم‌ها گسترش نفوذ بهتری ارائه می‌دهند. البته الگوریتم ICIM-GREEDY با اختلاف بسیار کم گسترش نفوذ بهتری از الگوریتم Greedy ارائه می‌دهد. در شکل ۸، گسترش نفوذ بر دیتاست NetScience مورد بررسی قرارگرفته است که الگوریتم ICIM-GREEDY گسترش نفوذ بهتری ارائه می‌دهد. در این دیتاست الگوریتم Greedy گسترش نفوذ تقریباً نزدیکی به الگوریتم ICIM-GREEDY ارائه می‌دهد ولی گسترش نفوذ در الگوریتم ICIM-GREEDY بیشتر از الگوریتم Greedy است. به عنوان‌مثال در الگوریتم ICIM-Greedy گسترش نفوذ در k=25 و k= 30 نسبت به الگوریتم Greedy بهتر است. در شکل ۹ و ۱۰ گسترش نفوذ بر دیتاست‌های M-Fo-0.1 و M-Fo-0.2 مورد بررسی قرارگرفته است که الگوریتم‌های ICIM-GREEDY و Greedy گسترش نفوذ تقریباً نزدیکی دارند و به نسبت دیگر الگوریتم‌ها گسترش نفوذ بهتری ارائه می‌دهند.

در شکل‌های ۱۱ تا ۱۸، الگوریتم‌ها از نظر زمان اجرا روی دیتاستهای مذکور مورد آزمایش قرارگرفته‌اند. همان‌طور که در این شکل‌ها قابل‌مشاهده است، الگوریتم پیشنهادی نسبت به الگوریتم‌های حریصانه مورد مقایسه زمان اجرای

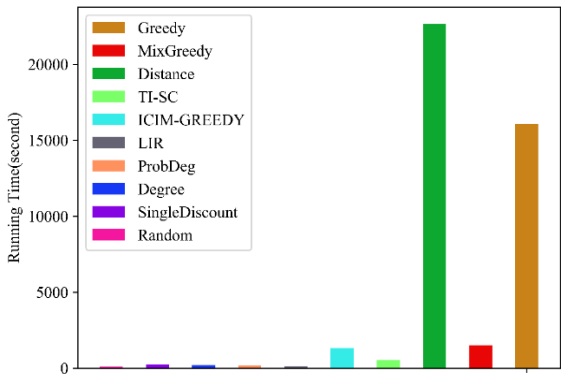
کمتری دارد. البته الگوریتم‌های Random و Degree دارای کمترین مرتبه‌ی زمانی هستند ولی این الگوریتم‌ها دارای دقت بسیار پایینی در محاسبه گسترش نفوذ دارند.



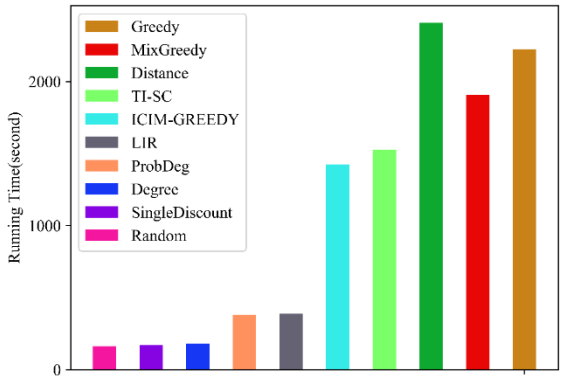
شکل ۱۱- زمان اجرای الگوریتم‌های متفاوت بر روی دیتاست NetHEPT



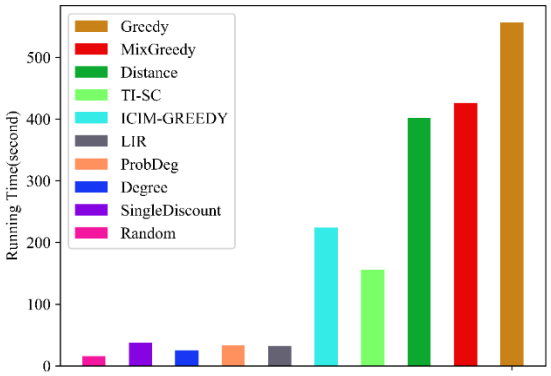
شکل ۱۲- زمان اجرای الگوریتم‌های متفاوت بر روی دیتاست Physics



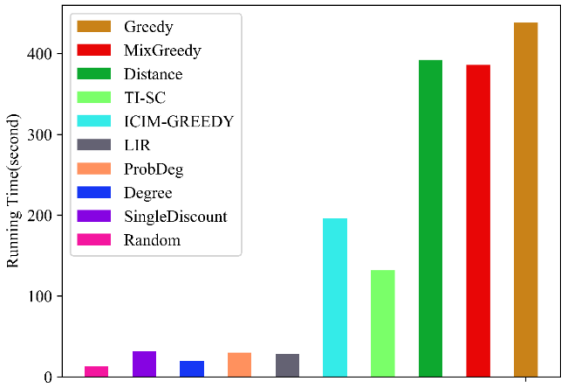
شکل ۱۳- زمان اجرای الگوریتم‌های متفاوت بر روی دیتاست M-FO115



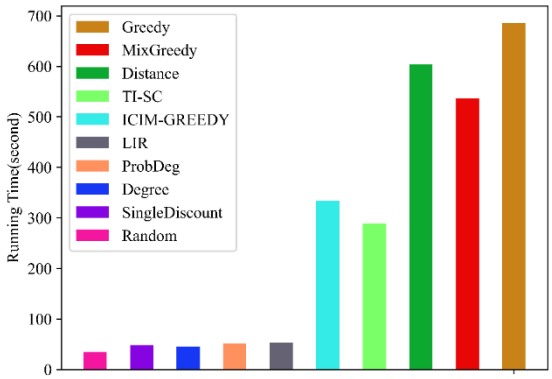
شکل ۱۴- زمان اجرای الگوریتم‌های متفاوت بر روی دیتاست M-FO120



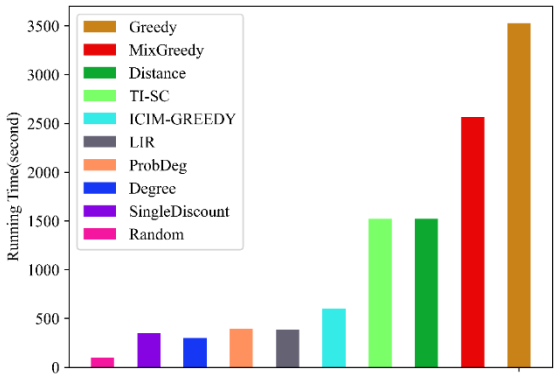
شکل ۱۵- زمان اجرای الگوریتم‌های متفاوت بر روی دیتاست Email



شکل ۱۶- زمان اجرای الگوریتم‌های متفاوت بر روی دیتاست Netscience



شکل ۱۷- زمان اجرای الگوریتم‌های متفاوت بر روی دیتاست M-Fo-0.1



شکل ۱۸- زمان اجرای الگوریتم‌های متفاوت بر روی دیتاست M-Fo-0.2

در این آزمایش، هدف اصلی مقایسه الگوریتم‌های حریصانه با ICIM-GREEDY است زیرا روش‌هایی مثل روش Random و Degree به خاطر ضعف بارز در گسترش نفوذ به دلیل عدم انتخاب آگاهانه، الگوریتم‌های خوبی نیستند و در این حوزه هم چندان موردتوجه نیستند.

۵- نتیجه‌گیری

در این مقاله بر اساس معیارهای جدید حساسیت به نفوذ و قدرت نفوذ یک الگوریتم بهینه برای مسئله حداکثر سازی نفوذ ارائه‌شده است.در الگوریتم ICIM-GREEDY از معیارهای که در دنیای واقعی وجود دارد، برای محاسبه حساسیت به نفوذ و قدرت نفوذ استفاده‌شده است. الگوریتم‌های حریصانه برای دقت بیشتر نیاز به تکرار بیشتر توسط شبیه‌سازی مونت‌کارلو دارند درصورتی‌که الگوریتم ICIM-GREEDY تنها با یک‌بار اجرا دارای دقت بالایی در گسترش نفوذ است به همین دلیل در این الگوریتم زمان اجرا نسبت به الگوریتم حریصانه کاهش داده‌شده است و دقت انتخاب مجموعه نودهای تأثیرگذار بر اساس گسترش نفوذ افزایش داده‌شده است. زمانی که زمان اجرا اهمیت دارد الگوریتم‌های SingleDiscount پیشنهاد می‌شود و زمانی که دقت اجرا اهمیت دارد الگوریتم‌های ICIM-GREEDY، MixedGreedy و TI-SC پیشنهاد می‌شود.

منابع

1. P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 57-66: ACM.
2. D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137-146: ACM.
3. S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *International workshop on web and internet economics*, 2007, pp. 306-311: Springer.
4. C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the spread of misinformation in social networks," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 665-674.
5. W. Chen *et al.*, "Influence maximization in social networks when negative opinions may emerge and propagate," in *Proceedings of the 2011 siam international conference on data mining*, 2011, pp. 379-390: SIAM.
6. X. He, G. Song, W. Chen, and Q. Jiang, "Influence blocking maximization in social networks under the competitive linear threshold model," in *Proceedings of the 2012 siam international conference on data*

leaping algorithm to identify influential nodes for influence maximization in social networks," *Knowledge-Based Systems*, vol. 187, p. 104833, 2020.
17. A. Zareie, A. Sheikhahmadi, and M. Jalili, "Identification of influential users in social network using gray wolf optimization algorithm," *Expert Systems with Applications*, vol. 142, p. 112971, 2020.
18. J. Leskovec *et al.*, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 420-429: ACM.
19. W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 199-208: ACM.
20. R. Narayanam and Y. Narahari, "A shapley value-based approach to discover influential nodes in social networks," *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 1, pp. 130-147, 2010.
21. M. Kitsak *et al.*, "Identification of influential spreaders in complex networks," *Nature physics*, vol. 6, no. 11, p. 888, 2010.
22. S. Cheng, H. Shen, J. Huang, G. Zhang, and X. Cheng, "Staticgreedy: solving the scalability-accuracy dilemma in influence maximization," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 509-518: ACM.
23. A. Goyal, W. Lu, and L. V. Lakshmanan, "Simpath: An efficient algorithm for influence maximization under the linear threshold model," in *2011 IEEE 11th international conference on data mining*, 2011, pp. 211-220: IEEE.
24. Y.-C. Chen, W.-Y. Zhu, W.-C. Peng, W.-C. Lee, and S.-Y. Lee, "CIM: Community-based influence maximization in social networks," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 2, p. 25, 2014.
25. W. Chen, T. Lin, Z. Tan, M. Zhao, and *mining*, 2012, pp. 463-474: SIAM.
7. W. Lu, W. Chen, and L. V. Lakshmanan, "From competition to complementarity: comparative influence diffusion and maximization," *arXiv preprint arXiv:1507.00317*, 2015.
8. H. Ma, Y. Zhu, D. Li, D. Kim, and J. Liang, "Improving the influence under IC-N model in social networks," *Discrete Mathematics, Algorithms and Applications*, vol. 7, no. 03, p. 1550037, 2015.
9. N. Pathak, A. Banerjee, and J. Srivastava, "A generalized linear threshold model for multiple cascades," in *2010 IEEE International Conference on Data Mining*, 2010, pp. 965-970: IEEE.
10. C. Wang, W. Chen, and Y. Wang, "Scalable influence maximization for independent cascade model in large-scale social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 545-576, 2012.
11. H. A. Beni and A. Bouyer, "TI-SC: top-k influential nodes selection based on community detection and scoring criteria in social networks," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-20, 2020.
12. Y. Chen, Q. Qu, Y. Ying, H. Li, and J. Shen, "Semantics-aware influence maximization in social networks," *Information Sciences*, vol. 513, pp. 442-464, 2020.
13. Q. He *et al.*, "CAOM: A community-based approach to tackle opinion maximization for social networks," *Information Sciences*, vol. 513, pp. 252-269, 2020.
14. M. M. Keikha, M. Rahgozar, M. Asadpour, and M. F. Abdollahi, "Influence maximization across heterogeneous interconnected networks based on deep learning," *Expert Systems with Applications*, vol. 140, p. 112905, 2020.
15. J. Ko, K. Lee, K. Shin, and N. Park, "MONSTOR: An Inductive Approach for Estimating and Maximizing Influence over Unseen Social Networks," *arXiv preprint arXiv:2001.08853*, 2020.
16. J. Tang, R. Zhang, P. Wang, Z. Zhao, L. Fan, and X. Liu, "A discrete shuffled frog-

X. Zhou, "Robust influence maximization," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 795-804: ACM.
26. J.X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, "Identifying a set of influential spreaders in complex networks," *Scientific reports*, vol. 6, p. 27823, 2016.
27. J. Shang, S. Zhou, X. Li, L. Liu, and H. Wu, "CoFIM: A community-based framework for influence maximization on large-scale networks," *Knowledge-Based Systems*, vol. 117, pp. 88-100, 2017.
28. F. Morone, B. Min, L. Bo, R. Mari, and H. A. Makse, "Collective influence algorithm to find influencers via optimal percolation in massively large social media," *Scientific reports*, vol. 6, p. 30062, 2016.
29. D. Liu, Y. Jing, J. Zhao, W. Wang, and G. Song, "A fast and efficient algorithm for mining top-k nodes in complex networks," *Scientific reports*, vol. 7, p. 43330, 2017.
30. S. Ahajjam and H. Badir, "Identification of influential spreaders in complex networks using HybridRank algorithm," *Scientific reports*, vol. 8, no. 1, p. 11932, 2018.
31. D.-L. Nguyen, T.-H. Nguyen, T.-H. Do, and M. Yoo, "Probability-based multi-hop diffusion method for influence maximization in social networks," *Wireless Personal Communications*, vol. 93, no. 4, pp. 903-916, 2017.
32. H. Wu, J. Shang, S. Zhou, Y. Feng, B. Qiang, and W. Xie, "LAIM: A linear time iterative approach for efficient influence maximization in large-scale networks," *IEEE Access*, vol. 6, pp. 44221-44234, 2018.
33. S. Banerjee, M. Jenamani, and D. K. Pratihar, "ComBIM: A community-based solution approach for the Budgeted Influence Maximization Problem," *Expert Systems with Applications*, vol. 125, pp. 1-13, 2019.
34. G. Xie, Y. Chen, H. Zhang, and Y. Liu, "MBIC: A Novel Influence Propagation Model for Membership-Based Influence Maximization in Social Networks," *IEEE Access*, vol. 7, pp. 75696-75707, 2019.
35. X. Rui, F. Meng, Z. Wang, and G. Yuan, "A reversed node ranking approach for influence maximization in social networks," *Applied Intelligence*, vol. 49, no. 7, pp. 2684-2698, 2019.
36. L. Qiu, W. Jia, J. Yu, X. Fan, and W. Gao, "PHG: A Three-Phase Algorithm for Influence Maximization Based on Community Structure," *IEEE Access*, vol. 7, pp. 62511-62522, 2019.
37. S. Milgram, "Behavioral study of obedience," *The Journal of abnormal and social psychology*, vol. 67, no. 4, p. 371, 1963.
38. V. Batagelj and M. Zaversnik, "An O (m) algorithm for cores decomposition of networks," *arXiv preprint cs/0310049*, 2003.
39. J. Kunegis, "Konect: the koblenz network collection," in *Proceedings of the 22nd International Conference on World Wide Web*, 2013, pp. 1343-1350.
40. A.L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509-512, 1999.

راهنمای نگارش
مقالات فصلنامهٔ علمی- پژوهشی فنّاوری اطلاعات و ارتباطات ایران

فصلنامهٔ علمی-پژوهشی ***«فنّاوری اطلاعات و ارتباطات ایران»***، نشریه‌ای علمی است که با هدف ایجاد ارتباط میان پژوهشگران و اندیشمندان عرصهٔ فنّاوری اطلاعات و اعتلای سطح دانش نظری این حوزهٔ علمی منتشر می‌شود. این فصلنامه در پی آن است که زمینه‌ای مناسب را برای ارائهٔ آخرین یافته‌های علمی محققان در حوزه‌های مختلف فنّاوری اطلاعات در داخل و خارج از ایران پدید آورد.

فصلنامه پذیرای مقاله‌هایی است که حاصل پژوهشهای اصیل نظری، کاربردی و توسعه‌ای در زمینه‌های تخصصیِ فنّاوری اطلاعات و فنّاوری ارتباطات باشد، از این رو از همهٔ پژوهشگران و متخصصان دعوت می‌شود تا نتایج تحقیقات خود را برای انتشار به فصلنامه ارسال کنند. دراین خصوص اشاره به نکات زیر ضروری است.

۱- مقاله‌های ارسالی به فصلنامه باید دربرگیرندهٔ نتایج نو در هر یک از حوزه‌های تخصصی فصلنامه باشد و نباید قبلاً در هیچ نشریهٔ دیگری به چاپ رسیده یا به طور همزمان به سایر نشریه‌های داخلی و خارجی برای داوری ارسال شده باشد. شایان ذکر است انتشار مقالهٔ کاملی که خلاصهٔ نتایج آن پیشتر در مجامع علمی ارائه و چاپ شده باشد، بلامانع است.

۲- مقاله‌های ارسالی به فصلنامه باید حاوی مطالب جدید و واجد نتایج، روشها، مفاهیم، کاربردها یا ترکیبی از این موارد باشد.

۳- فصلنامه از مقاله‌های مروری، تحلیلی و توصیفی (با دعوت از صاحبنظران برجسته) و نیز نامه به سردبیر (در نقد، تحلیل و تشریح مقاله‌ای خاص) استقبال می‌کند.

۴- مقاله‌ها باید بتوانند اطلاعات کافی را به خواننده منتقل کنند، از این رو لازم است حاوی بخشهای زیر باشند:

الف- عنوان مقاله: گویای اصلی‌ترین نکته و یافتهٔ مقاله (طول عنوان نباید از ۱۵ واژه بیشتر باشد).

ب- مشخصات نویسندگان(گان): شامل نام و نام خانوادگی، شغل، وابستگی سازمانی، نشانی پستی و الکترونیکی.

ج- چکیده: متنی حاوی حداکثر ۲۰۰ واژه که بیانگر نکات اصلی مقاله باشد.

د- کلید واژگان: شامل حداکثر ۵ واژهٔ اصلی که بتواند در نمایه‌گذاری مقاله، ویژگیهای آن را توصیف کند.

ه- بدنهٔ مقاله: شامل مقدمه، تبیین مسئلهٔ اصلی تحقیق، روش کاربردی برای حل مسئله و نتایج آن. این بخش باید چنان کامل باشد که خواننده با مطالعهٔ آن به اقناع علمی دست یابد از سویی لازم است نگارندگان از تطویل کلام بپرهیزند و بخشهای غیر ضروری را از مقاله حذف کنند.

و- مراجع: شامل همهٔ کتابها، مقاله‌ها، گزارشها و صفحه‌های وبِ مورد استناد در مقاله.

یادآوری۱: شکلها، جدولها و مراجع باید به ترتیبی که در متن آمده‌اند شماره‌گذاری شود. (عنوان شکل در زیر آن و عنوان جدول در بالای آن ذکر می‌شود).

یادآوری۲: شیوهٔ نگارش مراجع و ارجاع به آن در متن، مطابق استاندارد نشریات IEEE است.

یادآوری۳: در مقاله‌های فارسی لازم است همهٔ اصطلاحها، اسامی و اعداد کاربردی در متن مقاله، جدولها و شکلها به زبان فارسی نوشته شود و در صورت نیاز معادل فارسی آن در زیرنویس صفحه آورده شود. مبنای انتخاب معادل فارسی، مصوّبات فرهنگستان زبان و ادب فارسی است.

یادآوری۴: شیوهٔ نگارش فارسیِ فصلنامه براساس «دستور خط فارسی» مصوّب فرهنگستان زبان و ادب فارسی است و نگارندگان لازم است اصول نگارش آن را در مقاله رعایت کنند.

۵- مقاله‌های دریافتی به هر دو زبان فارسی و انگلیسی قابل پذیرش است. فصلنامه شماره‌های ویژه‌ای را به مقاله‌های انگلیسی زبان اختصاص خواهد داد. به هر حال همهٔ مقاله‌ها (خواه فارسی، خواه انگلیسی) باید دارای چکیدهٔ دو زبانه باشد.

۶- لازم است نگارندگان مقالهٔ خود را در محیط MS-Word به صورت دو ستونه، با فاصلهٔ خطوط ۱ (Single) و با قلم بی-لوتوس (B Lotus) ۱۱ (برای مقالهٔ فارسی) و Times New Roman 11 (برای مقالهٔ انگلیسی) و با رعایت حاشیهٔ بالا ۳ سانتی‌متر، پایین ۱ سانتی‌متر، چپ ۲,۵ سانتی‌متر، راست ۲,۵ سانتی‌متر و با توجه به نکات بند ۴، آماده و به دفتر نشریه

ارسال کنند. (در ضمن در صورت امکان، PDF آن را نیز ارسال شود.)

یادآوری: ضروری است نشانی کامل پستی والکترونیکی و نیز شماره تلفن نویسندهٔ عهده‌دار مکاتبات (Corresponding Author) درمقالهٔ ارسالی معلوم باشد تا امکان ایجاد ارتباط با نگارندگان فراهم آید.

۷- حق نشر مقاله‌ها برای فصلنامه محفوظ است؛ از این رو ضروری است نویسندهٔ عهده‌دار مکاتبات به نمایندگی از همهٔ نویسندگان «مجوّز حق‌انتقال نشر» را امضا کند، درغیر این صورت فصلنامه از چاپ مقالهٔ پذیرفته شده معذور است.

۸- هر مقاله پس از بررسی سردبیر به سه داور ارسال می‌شود. جمعبندی نظریات داوران برعهدهٔ سردبیر فصلنامه است. فصلنامه حق پذیرش، ویرایش یا ردّ مقاله‌ها را برای خود محفوظ می‌دارد.

۹- فصلنامه متعهد می‌شود در ازای چاپ هر مقاله، تعداد پنج نسخه از بازچاپ مقاله (یا پنج نسخه از مجله) را برای نویسندهٔ عهده‌دار مکاتبات ارسال کند. ضمناً دسترسی به نسخهٔ الکترونیکی مقاله برای همهٔ نویسندگان مقاله امکان‌پذیر خواهد بود.

فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال دهم، شماره‌های ۳۷و۳۸ ، پاییز و زمستان ۱۳۹۷

معرفی انجمن فناوری اطلاعات و ارتباطات ایران

اهداف

انجمن فنّاوری اطلاعات و ارتباطات ایران درسال ۱۳۸۳ با اهداف زیر تشکیل گردید:

- ایجاد ارتباط با پژوهشگران و افراد خبره در حوزه‌های مرتبط با فنّاوری اطلاعات (IT)، فنّاوری ارتباطات (CT) و فراهم نمودن زمینه‌های شناسایی و همکاری مشترک بین آنها.
- همکاری با نهادهای اجرائی، علمی و پژوهشی مرتبط با فنّاوری اطلاعات و ارتباطات، اقدام درجهت پیوند و همکاری دستگاههای اجرائی با مجموعه‌های علمی- تحقیقاتی.
- توسعه فرهنگ استفاده مطلوب از توانمندیهای فنّاوری اطلاعات و ارتباطات.
- ارزیابی وتحلیل عوامل رشد کشورهای مشابه دیگر و استفاده مطلوب از تجربیات دیگران در توسعه کشور درحوزهٔ ارتباطات و فنّاوری اطلاعات.

برنامه‌ها

انجمن به‌منظور تحقق اهداف خود، برنامه‌های زیر را در اولویت فعالیتهایش قرار داده است.

- ایجاد و تقویت ارتباط علمی، پژوهشی و آموزشی در سطوح ملّی، منطقه‌ای و بین‌المللی.
- برگزاری همایشهای علمی درسطح ملّی، منطقه‌ای و بین‌المللی.
- ایجاد تسهیلات برای نشر نتایج آخرین دستاوردهای علمی و پژوهشی مرتبط با نیازهای کشور درحوزهٔ ارتباطات و فنّاوری اطلاعات (انتشار کتب و نشریات علمی).
- برگزاری کارگاههای تخصصی درحوزهٔ ارتباطات و فنّاوری اطلاعات.

برنامه‌ریزی فعالیتها

- انتشار خبرنامه انجمن
- تشکیل کمیته‌های تخصصی به‌منظور تحقق اهداف انجمن
- اقدامات وسیع برای معرفی انجمن در همایشهای علمی، سازمانهای اجرائی و دانشگاهها
- انتشار نشریه علمی- پژوهشی ICT
- همکاری در برگزاری همایشهای علمی از قبیل چهارمین کنفرانس فنّاوری اطلاعات و دانش (IKT)
- تشکیل جلسات با مسئولین مملکتی ذیربط جهت شناساندن اهداف انجمن و مشخص‌شدن نحوه همکاری

کمیته‌های تخصصی انجمن

- کمیته پذیرش و اطلاع‌رسانی
- کمیته آموزش
- کمیته پژوهش
- کمیته مالی و پشتیبانی
- کمیته ارتباط صنعت دانشگاه
- کمیته جامعه اطلاعاتی

از کلیه اندیشمندان، اساتید دانشگاه، پژوهشگران و سایر علاقه‌مندان در حوزهٔ ارتباطات و فنّاوری اطلاعات دعوت می‌شود با عضویت در انجمن، ما را درتحقق اهداف والای آن یاری نمایند.

نشانی دبیرخانه: تهران- خیابان حافظ- دانشگاه صنعتی امیرکبیر- ساختمان ابوریحان- طبقه ششم- اتاق ۶۱۲
تلفن: ۶۶۴۸۵۸۵۶-۰۲۱ ، تلفکس: ۶۶۴۹۵۴۳۳- ۰۲۱

اعضای حقوقی انجمن در سال ۹۲	
• عصر ارتباطات و انتقال داده‌های سپاهان امیر صفوی شماره عضویت: ۹۲۴۸۶۵۱۸۴	• موسسه آموزش عالی غیر انتفاعی نوین مجید میرزا محمدی ۹۲۴۴۵۳۱۸۵ شماره عضویت:
• جامع علمی و کاربردی خانه کارگر استان اردبیل ودود مبینی شماره عضویت: ۹۲۴۴۵۳۱۸۶	
اعضای پیوسته انجمن در سال ۹۲	
انجمن فناوری اطلاعات و امیر شهاب شاهمیری- ارتباطات شماره عضویت : ۸۹۱۲۱۳۵۴۰	سید مهدی حاتمیان- وزارت فناوری اطلاعات و ارتباطات ایران شماره عضویت : ۸۹۱۲۱۲۵۴۳
شیرین گیلکی – جهاد دانشگاهی شماره عضویت : ۹۰۳۲۱۵۵۸۹	فریدون عبدی – دانشگاه افسری امام علی (ع) شماره عضویت : ۹۱۳۲۱۳۶۷۰
نرگس رضایی ملک - شرکت هوای فشرده ایمن شماره عضویت : ۹۲۱۲۱۶۶۸۴	محمد نوری مطلق – شماره عضویت : ۹۲۱۵۸۶۶۸۶
احمد استیری - شماره عضویت : ۹۲۱۲۱۶۶۸۶	محمدآبادی – دانشگاه آزاد اسلامی واحد مشهدالهام میرزایی شماره عضویت : ۹۲۱۵۱۶۶۸۶
ساسان عظیمی - دفتر مطالعه و نوآوری شماره عضویت : ۹۲۱۲۱۳۷۰۵	زهرا عزیزی - شرکت ایزایران شماره عضویت : ۹۲۱۲۱۳۷۰۶
غلامحسین قاسمی –دانشکده برق و کامپیوتر، دانشکده فنی قائن شماره عضویت : ۹۲۱۵۶۳۷۰۷	مهریزی - موسسه آموزش عالی علمی زهرا ملا محمدعلیان کاربردی هلال ایران شماره عضویت : ۹۲۱۲۱۳۷۰۸
رباب قاسم شربیانی- شرکت مهندسی مشاور نیروی آذربایجان (منا) شماره عضویت : ۹۲۱۴۱۵۷۰۹	سید رئوف خیامی- دانشگاه صنعتی شیراز ۹۲۱۷۱۳۷۱۰ شماره عضویت :
سجاد رضائیان- دانشگاه صنعتی شیراز ۹۲۱۷۳۳۷۱۱ شماره عضویت :	پانیذ علیپوراقدم- ۹۲۱۲۱۵۷۱۲ شماره عضویت:
پریسا پورحسن- ۹۲۱۱۲۳۷۱۳ شماره عضویت :	سامان سیادتی- ۹۲۱۲۱۳۷۱۴ شماره عضویت :
عرفانه غروی- ۹۲۱۷۱۳۷۱۵ شماره عضویت :	کامبیز رضایی- پردیس شهید رجایی فارس ۹۲۱۷۱۳۷۱۶ شماره عضویت :
ایمان قاصدیان- بانک تجارت استان فارس ۹۲۱۷۱۶۷۱۷ شماره عضویت :	شریفی- دانشگاه علوم پزشکی شیرازمژگان شیال ۹۲۱۷۱۳۷۱۸ شماره عضویت :
سید هادی موسوی- دانشگاه صنعتی بیرجند ۹۲۱۵۶۳۷۱۹ شماره عضویت :	نوبری- دانشکده تربیت بدنی و علوم ورزشی دانشگاه دل‌بابک زنده تهران ۹۲۱۲۱۳۷۲۰ شماره عضویت :

سهیل افراز- دانشگاه آزاد اسلامی واحد اردبیل ۹۲۱۴۵۳۷۲۱ شماره عضویت :	عبدالله مهدوی- دانشگاه علوم پزشکی اردبیل ۹۲۱۴۵۳۷۲۲ شماره عضویت :
فاطمه حورعلی- شماره عضویت :۹۲۱۵۷۳۷۳۳	علی موثق- شماره عضویت :۹۲۱۲۱۳۷۲۴
محمدرضا کنعانی- اداره کل حفاظت محیط زیست استان مازندران شماره عضویت :۹۲۱۱۵۳۷۲۵	لیلا انصاری فرد- شماره عضویت :۹۲۱۴۵۳۷۲۶
اعضای دانشجویی انجمن در سال ۹۲	
محسن گلبند- شماره عضویت : ۹۲۳۲۱۳۶۳۳	فراهانی-دانشگاه صنعتی شریف‌الهام درمنکی شماره عضویت : ۹۲۳۲۱۳۶۳۴
الهه پوریان- شماره عضویت :	شاهرخ بحتوئی ۹۲۳۷۱۳۶۳۵ شماره عضویت :
محمدکاظم حیدری- شماره عضویت : ۹۲۳۷۱۳۶۳۶	پیمان غلامی- دانشگاه آزاد اسلامی اراک ۹۲۳۸۶۳۶۳۷ شماره عضویت :
نژادسهیلا بارچی ۹۲۳۳۴۳۶۳۸ شماره عضویت :	میثم افتخاری- دانشگاه آزاد اسلامی واحد اراک ۹۲۳۸۶۳۶۳۹ شماره عضویت :
آیت قلندری- دانشگاه آزاد اسلامی واحد اراک ۹۲۳۶۶۳۶۴۰ شماره عضویت :	موسوی- مدرسه راهنمایی فرزانگان ۲وجیهه لوح ۹۲۳۳۱۳۶۴۱ شماره عضویت :
آبادی-دانشگاه قم‌ملک‌زهره گلی ۹۲۳۳۳۳۶۴۲ شماره عضویت :	بیدگلی- پردیس فنی دانشگاه تهران‌علی بلالی ۹۲۳۲۱۳۶۴۳ شماره عضویت :
محمدمهدی فیروزی- دانشگاه صنعتی امیرکبیر ۹۲۳۲۱۳۶۴۴ شماره عضویت :	مسعود نجفی- دانشگاه آزاد اسلامی نجف‌آباد اصفهان ۹۲۳۷۷۳۶۴۵ شماره عضویت :
سپهر قاسمی-پردیس دانشکده فنی دانشگاه تهران ۹۲۳۲۱۳۶۴۶ شماره عضویت :	سالار محتاج- دانشگاه صنعتی امیرکبیر ۹۲۳۲۱۳۶۴۷ شماره عضویت :
لیلا آذری- دانشگاه آزاد اسلامی واحد علوم و تحقیقات خوزستان ۹۲۳۲۱۳۶۴۸ شماره عضویت :	زنجانی- دانشگاه صنعتی امیرکبیرمحمد دریایی ۹۲۳۶۱۳۶۴۹ شماره عضویت :
محسن مرادی- دانشگاه صنعتی مالک اشتر ۹۲۳۲۱۳۶۵۰ شماره عضویت :	امیر اسماعیلی- دانشگاه صنعتی امیرکبیر ۹۲۳۲۱۳۶۵۱ شماره عضویت :
آتنا کاویان- دانشگاه صنعتی امیرکبیر ۹۲۳۲۱۳۶۵۲ شماره عضویت :	آزاده پیرحیاتی- دانشگاه آزاد اسلامی همدان ۹۲۳۶۶۳۶۵۳ شماره عضویت :
ناهید فتوحی- دانشگاه سلمان فارسی کازرون ۹۲۳۷۱۳۶۵۴ شماره عضویت :	بابالی- وزارت فناوری اطلاعات و ارتباطات میراسماعیل میرنبی مازندران ۹۲۳۱۱۳۶۵۵ شماره عضویت :
ندا بینش ۹۲۳۲۳۳۶۵۶ شماره عضویت :	محمدمهدی محصولی- دانشگاه صنعتی امیرکبیر ۹۲۳۲۱۳۶۵۷ شماره عضویت :

خواه- دانشگاه علوم پزشکی شیرازمهدی شرف ۹۲۳۷۱۳۶۵۸ شماره عضویت :	حمیدرضا زارعیان- بانک تات ۹۲۳۲۱۳۶۵۹ شماره عضویت :
عباس خسروانی- دانشگاه صنعتی امیرکبیر ۹۲۳۲۱۳۶۶۰ شماره عضویت :	بابایی‌فاطمه حاجی ۹۲۳۵۱۳۶۶۱ شماره عضویت :
رامین ساروقی- دانشگاه سلمان فارسی کازرون ۹۲۳۷۱۳۶۶۲ شماره عضویت :	جم- دانشگاه تبریزمعصومه رضائی ۹۲۳۸۷۳۶۶۳ شماره عضویت :
بهرام هدایتی ۹۲۳۷۳۳۶۶۵ شماره عضویت :	اصغر ترکی- فولاد مبارکه اصفهان ۹۲۳۳۱۳۶۶۶ شماره عضویت :
فاطمه قنبریان ۹۲۳۷۱۳۶۶۶ شماره عضویت :	محمد ضیاءالدینی- دانشگاه باهنر ۹۲۳۳۴۳۶۶۷ شماره عضویت :
خلجی‌فاطمه کیماسی ۹۲۳۳۷۳۶۶۸ شماره عضویت :	وحید ظهیرپور- دانشگاه علم و صنعت ایران ۹۲۳۲۱۳۶۶۹ شماره عضویت :
داریوش سرخه- دانشگاه آزاد اسلامی شوش ۹۲۳۶۴۳۶۷۰ شماره عضویت :	محسن رمضانی ۹۲۳۸۷۳۶۷۱ شماره عضویت :
فرزانه کرمی- دانشگاه صنعتی امیرکبیر ۹۲۳۲۱۳۶۷۲ شماره عضویت :	محسن یعقوبی سورکی ۹۲۳۱۵۳۶۷۳ شماره عضویت :
امیرحسین خانیکی ۹۲۳۵۱۳۶۷۴ شماره عضویت :	محمدی‌علیرضا ملک ۹۲۳۳۸۳۶۷۵ شماره عضویت :
رضا سالارمهر- دانشگاه صنعتی امیرکبیر ۹۲۳۲۱۳۶۷۶ شماره عضویت :	فاطمه مجیدی- دانشگاه جامع علمی کاربردی خانه کارگر ۹۲۳۴۵۳۶۷۷ شماره عضویت :
میثم معمرپور- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۷۸ شماره عضویت :	تیراقلویوسف عطائی ۹۲۳۴۵۳۶۷۹ شماره عضویت :
سلمان نظیری- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۸۰ شماره عضویت :	خواه- موسسه غیرانتفاعی مقدس اردبیلی‌شبنم وطن ۹۲۳۴۵۳۶۸۱ شماره عضویت :
امید تقوائی- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۸۲ شماره عضویت :	جمادی- موسسه غیرانتفاعی مقدس اردبیلی‌فریده نعمتی ۹۲۳۴۵۳۶۸۳ شماره عضویت :
مجتبی لاله- شورای اسلامی شهر اردبیل ۹۲۳۴۵۳۶۸۴ شماره عضویت :	هاجر کوثرزاده- دانشگاه آزاد اسلامی واحد اردبیل ۹۲۳۴۵۳۶۸۵ شماره عضویت :
میلاد راستگو- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۸۶ شماره عضویت :	میلاد مددپور- دانشگاه آزاد اسلامی واحد اردبیل شماره عضویت : ۹۲۳۴۵۳۶۸۷
سحر ولی نژادروح بخش- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۸۸ شماره عضویت :	گلناز هاشمی- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۸۹ شماره عضویت :
رضا فرجی- موسسه غیرانتفاعی مقدس اردبیلی شماره عضویت : ۹۲۳۴۵۳۶۹۰	بهاءالدین کاوه‌پور- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۹۱ شماره عضویت :
مریم قاسمی- موسسه غیرانتفاعی مقدس اردبیلی	فاطمه جعفرزاده- موسسه غیرانتفاعی مقدس اردبیلی

۹۲۳۴۵۳۶۹۲ شماره عضویت :	۹۲۳۴۵۳۶۹۳ شماره عضویت :
زینب تیزپر- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۹۴ شماره عضویت :	توحید فرهادی- دانشگاه آزاد اسلامی اردبیل ۹۲۳۴۵۳۶۹۵ شماره عضویت :
محمد وجدانی- دانشگاه آزاد اسلامی اردبیل ۹۲۳۴۵۳۶۹۶ شماره عضویت :	جواد مؤذنی- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۹۷ شماره عضویت :
سید مسعود سجادی- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۹۸ شماره عضویت :	حجت همپایی- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۶۹۹ شماره عضویت :
عباسی- موسسه غیرانتفاعی مقدس برحامد بیرون اردبیلی ۹۲۳۴۵۳۷۰۰ شماره عضویت :	سجاد آخربین- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۳۴۵۳۷۰۱ شماره عضویت :
زاده- موسسه غیرانتفاعی مقدس اردبیلی‌مهدی بیک ۹۲۳۴۵۳۷۰۲ شماره عضویت :	زاده- موسسه غیرانتفاعی مقدس اردبیلی‌مینا عالی ۹۲۳۴۵۳۷۰۳ شماره عضویت :
اعضای وابسته انجمن در سال ۹۲	
سعید صفوی ۹۱۵۲۱۵۳۴۸ شماره عضویت :	الهه قنبریان ۹۲۲۷۱۶۳۸۴ شماره عضویت :
علیرضا ابریشمی‌شیرازی ۹۲۲۷۱۶۳۸۵ شماره عضویت :	نقی زندیان‌اجیرلو- موسسه غیرانتفاعی مقدس اردبیلی ۹۲۲۴۵۳۳۸۶ شماره عضویت :
سید عباس پیغمبرزاده- هنرستان غیرانتفاعی ایران مهارت شماره عضویت: ۹۲۲۴۵۳۳۸۷	حمید نصیری ۹۲۲۴۵۳۳۸۸ شماره عضویت :
جعفر زاهدی‌اول ۹۲۲۵۷۳۳۸۹ شماره عضویت:	حسین رهبرفر ۹۲۲۴۵۳۳۹۰ شماره عضویت :
محمدرضا سیدپیروی ۹۲۲۴۵۳۳۹۱ شماره عضویت:	مهدی جانفشار-دانشگاه آزاد اسلامی‌واحد اردبیل ۹۲۲۴۵۳۳۹۲ شماره عضویت :
توحید ابراهیمی-اداره کل راه و شهرسازی اردبیل ۹۲۲۴۵۲۳۹۳ شماره عضویت :	سمیرا نجفی ۹۲۲۴۵۳۳۹۴ شماره عضویت :

Determining the factors affecting the collective financing of knowledge-based IT companies

Ali Hajigholamsaryazdi Ali Rajabzadeh ghotri Alinaghi Mashayekhi
Alireza Hassanzade

Abstract:

The method of crowdfunding in the world has expanded rapidly due to the need for financing in the early stages of start-up businesses as well as advances in information technology. In Iran, several financing platforms have been established so far, some of which have been successful and some of which have been unsuccessful. Therefore, it is necessary to help the development of this method by examining the factors affecting it. Since mass financing is a new and new phenomenon, it is necessary to increase its awareness in the society in an appropriate way while determining the factors affecting this method. The collective modeling method is based on social networks and the Web 2 with the aim of recognizing new phenomena. Therefore, in this article, using collective modeling, the factors affecting crowdfunding in Iran in order to support start-up companies in the field of IT are discussed.

Keywords: Crowdfunding; IT start-ups; Collective modeling; Systems dynamics approach; Telegram.

Provide methodology for selecting IT project management strategy

Mona Jamipour Seyed Mohammadbagher Jafari Marziye Nasrollahi

Abstract:

The strategic approach to project management is a new approach that has received more and more attention from business thinkers in the last decade. Having a strategic approach in the field of information technology projects requires managers to choose the right strategy in the first step of project management, taking into account several factors influencing the choice of strategy. Logical and adaptive approaches are the two dominant strategies in project management, each of which requires different tools, resources, and procedures. Despite the increase in investment in IT projects and the dependence of organizations on such technologies, still No study has been conducted to assist IT managers in choosing the appropriate strategy for such projects. Therefore, the main purpose of this study is to identify the approach of selecting the appropriate strategy in the management of information technology projects. In order to achieve this goal, in the first stage, in order to identify the factors influencing the choice of IT project management strategy, after a comprehensive review of the literature in the relevant field, the opinions of experts in the focus group have been used. In the second stage, in order to validate the extracted factors, and in the third stage, in order to weight each of the factors in selecting the two dominant logical and comparative approaches, experts are surveyed using the survey method. In the developed methodology, the factors that influence the choice of strategy are: process factors, human factors, system related factors and contextual factors. The final strategy of IT project management is obtained as a range of logical strategy and adaptive strategy.

Keywords: Project management, IT project, project management strategy, logical approach, adaptive approach

Design of FPGA-based processor for SHA-2 series cryptographic algorithms

Neda Sedghahrabi Mohammad ALI Jabraeil Jamali

Abstract:

Secure decryption algorithms are a type of cryptographic algorithms whose importance in today's society has been highlighted by applications such as the use of personal digital tools to maintain confidentiality. On the other hand, with the advancement of technology, the need to implement these algorithms on flexible platforms can be challenging. Reducing the area and speeding up the execution of operations are the main challenges for designing and implementing these algorithms. This paper proposes a new architecture for the FPGA-based processor for SHA-2 series cryptographic algorithms. In the proposed processor, the use of memory units and multi-port data path, followed by parallel processor performance, has reduced the use of resources and increased the speed of data processing. Processor architecture for SHA-2 cryptographic algorithms is modeled in VHDL and implemented on the FPGA platform in the Virtex series by ISE software. Implementation results show that the proposed compact processor compared to previous tasks with similar objectives, was able to increase the operating frequency for the SHA-256 cryptographic algorithm by 25% and occupy 55% less space for the SHA-512 cryptographic algorithm to the desired level of operational power and efficiency. Also maintain. The proposed processor is suitable for applications such as trusted mobile platforms (TMP), digital currency (Bitcoin) and secure on-chip network routing (NoC).

Keywords:Secure decoder algorithms, SHA-2 series encryption algorithms, CPU, VHDL, FPGA

Learning to Rank for the Persian Web Using the Layered Genetic Programming

Amirhossein Keyhanipour

Abstract:

Learning to rank (L2R) has emerged as a promising approach in handling the existing challenges of Web search engines. However, there are major drawbacks with the present learning to rank techniques. Current L2R algorithms do not take into account to the search behavior of the users embedded in their search sessions' logs. On the other hand, machine-learning as a data-intensive process requires a large volume of data about users' queries as well as Web documents. This situation has made the usage of L2R techniques questionable in the real-world applications. Recently, by the use of the click-through data model and based on the generation of click-through features, a novel approach is proposed, named as MGP-Rank. Using the layered genetic-programming model, MGP-Rank has achieved noticeable performance on the ranking of the English Web content. In this study, with respect to the specific characteristics of the Persian language, some suitable scenarios are presented for the generation of the click-through features. In this way, a customized version of the MGP-Rank is proposed of the Persian Web retrieval. The evaluation results of this algorithm on the dotIR dataset, indicate its considerable improvement in comparison with major ranking methods. The improvement of the performance is particularly more noticeable in the top part of the search results lists, which are most frequently visited by the Web users.

Keywords: Learning to Rank, Layered Genetic Programming, Click-through Features, Persian Web Content, dotIR dataset.

Scheduling tasks in cloud environments using mapping framework - reduction and genetic algorithm

Nima Khezr Nima jafari novimipour

Abstract:

Task scheduling is a vital component of any distributed system such as grids, clouds, and peer-to-peer networks that refer tasks to appropriate resources for execution. Common scheduling methods have disadvantages such as high time complexity, inconsistent execution of input tasks, and increased program execution time. Exploration-based scheduling algorithms to prioritize tasks from. They use different policies that result in high execution times on heterogeneously distributed computing systems. Therefore, it is appropriate to prioritize it to produce a minimum execution time. Genetic algorithm is used as one of the evolutionary methods to optimize NP-complete problems. In this paper, a parallel genetic algorithm using a mapping-reduction framework for scheduling tasks on cloud computing using multiple priority queues is presented. The main idea of this article is to use a mapping-reduction framework to reduce the execution time of the entire program. Test results on A set of directional graphs without random rounds indicates that the proposed method improves the execution time of all two existing methods with high convergence speed.

Keywords: Cloud Computing, Task Scheduling, Execution Reduction, Genetic Algorithm, Mapping-Reduction, Hadoop

A greedy new method based on the cascade model to calculate maximizing penetration in social networks

Asgarali Bouyer Hamid Ahmadi

Abstract:

In the case of penetration maximization, the goal is to find the minimum number of nodes that have the most propagation and penetration in the network. Studies on maximizing penetration and dissemination are becoming more widespread. In recent years, many algorithms have been proposed to maximize the penetration of social networks. These studies include viral marketing, spreading rumors, innovating and spreading epidemics, and so on. Each of the previous studies has shortcomings in finding suitable nodes or high time complexity. In this article, we present a new method called ICIM-GREEDY to solve the problem of maximizing penetration. In the ICIM-GREEDY algorithm, we consider two important criteria that have not been considered in the previous work, one is penetration power and the other is penetration sensitivity. These two criteria are always present in human social life. The proposed method is evaluated on standard datasets. The obtained results show that this method has a better quality in finding penetrating nodes in 30 seed nodes than other compared algorithms. This method also performs better in terms of time compared to the comparative algorithms in terms of relatively fast convergence.

Keywords: Independent cascade model, penetration maximization, diffusion, social network

Contents

Iranian Journal of
Information Technology & Communication
No. 37-38, Vol.10, September- March 2019

- **Identifying Factors Affecting on Crowdfunding of Knowledge Based IT Companies Through Crowd Model Building** 1
  A. Haji Gholam Saryazdi, A. Rajabzadeh Ghotri, A.N. Mashayekhi, A.R. Hassanzadeh
- **introducing a methodology of strategy selection in IT project management** 17
  M. Jami Pour, M.B. Jafari, M. Nasrollahi
- **Design of a FPGA-based Processor for the Cryptographic Algorithms SHA-2** 35
  N. Sedgh Ahrabi, M.A. Jabraeil Jamali
- **Learning to Rank for the Persian Web Using the Layered Genetic Programming** 45
  A.H. Keyhanipour
- **Task scheduling in cloud environments using MapReduce framework and genetic algorithm** 71
  N. khezr, N. Jafari
  - **A new greedy method based on cascade model for the influence maximization problem in social networks.** 85
  A.A. Bouyer, H. Ahmadi