

صاحب امتیاز: انجمن فناوری اطلاعات و ارتباطات ایران

مدیر مسؤول: دکتر مسعود شفیعی

سردبیر: دکتر حسن آقائی نیا

دبیر داخلی: دکتر هومان تحیری (دانشگاه شیراز)

شمارگان: ۱۰۰۰

حروفچینی و صفحه آرایی: سمیه آهاری

هیئت تحریریه:

فرخ حجت کاشانی (استاد)، دانشگاه علم و صنعت ایران

سید احمد رضا شرافت (استاد)، دانشگاه تربیت مدرس

مسعود شفیعی (استاد)، دانشگاه صنعتی امیرکبیر

محمد رضا 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 محمود کمره‌ای (رئیس انجمن)

دانشگاه تهران

دکتر علی دوست محمدی

دانشگاه صنعتی امیرکبیر

دکتر عباس محمدی

دانشگاه صنعتی امیرکبیر

دکتر سید علیرضا هاشمی گلپایگانی

دانشگاه صنعتی امیرکبیر

دکتر مهدی شجری

دانشگاه صنعتی امیرکبیر

دکتر وحید یزدانیان

دانشگاه علمی- کاربردی پست و مخابرات

دکتر اکبر کاری دولت آبادی

دانشگاه هوایی شهید ستاری

دکتر امیر شهاب شاهمیری

دانشگاه جامع علمی کاربردی

دکتر رمضانعلی صادق‌زاده

دانشگاه صنعتی خواجه نصیرالدین طوسی

مهندس ناصر یوسف‌پور

شرکت مخابرات همراه اول

مهندس سیدعلی صموتی

دانشگاه فنی و حرفه‌ای

• اعضای علی‌البدل:

مهندس فرهاد رضایی (شرکت کوشا تجارت نو پدید)

دکتر جمشید شیشه‌زاده (دانشگاه خوارزمی)

بازرسان:

• بازرس اصلی:

دکتر مسعود شفیعی (دانشگاه صنعتی امیرکبیر)

• بازرس علی‌البدل:

مهندس اکبر مصطفی صادقی (دانشگاه آزاد اسلامی)

براساس نامه شماره ۳/۴۸۱۷ مورخ ۱۳۸۶/۶/۱۱ دفتر امور پژوهشی وزارت
علوم، تحقیقات و فناوری و بیانیه رأی کمیسیون بررسی نشریات
علمی کشور در تاریخ ۱۳۸۶/۴/۲۳، درجه علمی- پژوهشی به این فصلنامه
اعطا شده است

فهرست

فصلنامه علمی - پژوهشی فناوری اطلاعات و ارتباطات ایران

سال هشتم، شماره‌های ۲۷ و ۲۸، بهار و تابستان ۱۳۹۵

-
- ۱ ■ تغییرات جدید الگوی دودویی محلی و طبقه‌بندی و قسمت‌بندی تصاویر بافتی بستر دریا
بابک گودرزی، رضا جاویدان، محمدجواد دهقانی
 - ۲۱ ■ ارائه روشی مناسب برای دسته‌بندی نامه‌های الکترونیکی تبلیغاتی بر مبنای پروفایل کاربران
محمد فتحیان، رحیم حضرتقلی‌زاده
 - ۳۷ ■ استخراج معماری فناوری اطلاعات بیمارستان با رویکرد قابلیت پیاده‌سازی در ایران
عاطفه‌السادات حقیقت‌حسینی، حسین بویرشاد، فاطمه ثقفی، هادی زارع
 - ۴۷ ■ ارائه مدلی جهت استفاده از عاملهای متحرک در سیستمهای تشخیص نفوذ توزیع‌شده مبتنی بر
تئوری بازی
امین نظارات، مهدی رجا، غلامحسین دستغیبی‌فرد
 - ۶۱ ■ تلفیق مدل تحلیل پوششی داده‌ها و درخت تصمیم به منظور ارزیابی واحدهای مبتنی بر
تکنولوژی اطلاعات
امیر امینی، علیرضا علی‌نژاد، سمیه شفق‌زاده
 - ۸۱ ■ درک اعتماد اولیه کاربران به شبکه‌های اجتماعی
محسن اکبری، کامران زاهدفر، زهرا ایاغ
 - ۹۵ ■ نگارش
 - ۹۷ ■ معرفی انجمن فناوری اطلاعات و ارتباطات
 - ۹۸ ■ چکیده انگلیسی مقالات
-

تغییرات جدید الگوی دودویی محلی و طبقه‌بندی و قسمت‌بندی تصاویر بافتی بستر دریا

*بابک گودرزی **رضا جاویدان ***محمد جواد دهقانی

*کارشناسی‌ارشد، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، ایران

**دانشیار دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، ایران

***دانشیار دانشکده مهندسی برق و الکترونیک، دانشگاه صنعتی شیراز، ایران

تاریخ دریافت: ۹۳/۰۹/۱۴ تاریخ پذیرش: ۹۶/۰۷/۰۶

چکیده

تحلیل بافت نقش مهمی را در پردازش تصویر ایفا می‌کند. با توجه به ظاهر فوق‌العاده بافتی تصاویر سونار، روش‌های تحلیل بافت یک انتخاب مناسب برای تحلیل تصاویر آکوستیکی دریا هستند. اپراتور الگوی دودویی محلی یک توصیف‌گر بسیار موثر و چند رزولوشنی بافت است. این توصیف‌گر، اطلاعات مناسب را از تغییر روشنایی و حالات تصویر بدست می‌آورد. با اینکه توسعه‌های زیادی از الگوی دودویی محلی ارائه شده است اما اپراتورهای الگوهای دودویی محلی موجود نسبت به نویز حساسند. همچنین گاهی اوقات منجر به توصیف الگوهای مختلف ساختاری با کد دودویی همگون می‌شوند که به ناچار قابلیت تمایز خود را کاهش می‌دهند. این تحقیق یک بررسی اجمالی پیرامون روش الگوی دودویی محلی را ارائه نموده که شامل چندین مورد از متغیرهای جدیدتر است. سپس برای غلبه بر ناکارآمدی‌های انواع الگوهای دودویی محلی، یک چارچوب قوی از الگوی دودویی بنام الگوی دودویی محلی تکمیل شده مقاوم ارائه شده که ارزش هر پیکسل مرکزی با میانگین مقادیر شدت خاکستری خانه‌ها از یک مربع سه در سه جایگزین می‌گردد. روش ارائه شده یک ابزار سریع با دقت بالا در طبقه‌بندی تصاویر بستر دریاست که مقایسه نتایج شبیه‌سازی با دیگر روش‌های مشهور، نشان‌دهنده کارایی الگوریتم ارائه شده است.

واژه‌های کلیدی: الگوی دودویی محلی، طبقه‌بندی، قسمت‌بندی، بستر دریا، تصاویر بافتی آکوستیکی.

۱. مقدمه

اشیاء در کف دریا دارد. از سوی دیگر، روش‌های کاشف زیردریایی با رزولوشن^۲ بالا و سنسورهای نوری، نقش مهمی را در سنجش زیر آب برای تقسیم‌بندی‌های^۳ سنجشی و طبقه‌بندی بستر دریا دارند [۱].

نیاز به طبقه‌بندی^۱ و مدیریت دقیق منابع در علوم مختلف از جمله زمین‌شناسی و بستر دریا هم احساس شده است. دانش‌برداری از بستر دریا نقش مهمی در درک محیط زیر آب از جمله بررسی‌های زمین‌شناسی، اکتشاف ژئوفیزیک، مهندسی اقیانوس (وسایل نقلیه زیر آب، نظارت بر خطوط لوله و کابل و...)، شبیه‌سازی انتشار صدا، برآورد پارامترهای فیزیکی، ناوبری، اندازه‌گیری مواد منفجره ضد معدن، ارتباطات داده‌ها و طبقه‌بندی

^۲. Resolution

^۳. segmentation

^۱. classification

استفاده می‌کند، یک ابزار قدرتمند برای تحلیل بافت می‌باشد. الگوی دودویی محلی که ابتدا توسط اوجالا و همکارانش پیشنهاد شد، بعلت مقاومتش نسبت به تغییرات روشنایی، پیچیدگی محاسباتی کم و توانایی کدگذاری جزئیات یکی از رایج ترین توصیف گرها می‌باشد [۸].

به عنوان یک روش غیرپارامتری، الگوی دودویی محلی به صورت خلاصه ساختارهای محلی تصاویر را به طور کارآمد با مقایسه هر کدام از پیکسل‌ها با پیکسل‌های مجاور نشان می‌دهد. الگوی دودویی محلی در اصل برای آنالیز بافت پیشنهاد گردیده است [۹] و هنوز به عنوان یک رویکرد قدرتمند برای توصیف ساختارهای محلی معرفی می‌شود.

اگرچه الگوی محلی دودویی و انواع آن، نتایج طبقه‌بندی موثری بر روی پایگاه داده‌های بافتی بدست آورده‌اند اما تاکنون معایب بالقوه زیادی از الگوی محلی دودویی باقی مانده است. به عنوان مثال الگوی محلی دودویی حساس به نویز است و اغلب طبقه‌بندهایش اختلاف الگویی بسیار زیادی را در کلاس‌های همگون دارند.

این مقاله با ارائه یک چارچوب قوی از الگوی محلی دودویی بنام الگوی محلی دودویی تکمیل شده مقاوم (CRLBP)^{۱۳} به حل این مشکلات بالقوه پرداخته است. در الگوی محلی دودویی تکمیل شده مقاوم، ارزش هر پیکسل مرکزی از یک همسایگی سه در سه با مقدار متوسط سطوح خاکستری جایگزین می‌گردد. در مقایسه با ارزش خاکستری، متوسط سطح خاکستری محلی، در مقابل نویز و انواع روشنایی‌ها قوی‌تر است. نتایج تجربی نشان می‌دهد که الگوی محلی دودویی تکمیل شده مقاوم، بهترین دقت طبقه‌بندی را در مقایسه با سایر الگوهای محلی دودویی می‌دهد و کمترین حساسیت را در مقابل نویز و انواع روشنایی دارد.

ساختار ادامه مقاله به صورت زیر است: در بخش دو الگوی دودویی محلی معرفی شده و در بخش سه توسعه‌های جدید آن تشریح شده است. در بخش چهار، دو عیب اساسی الگوی دودویی محلی عنوان شده و الگوی محلی دودویی تکمیل شده مقاوم و ادغام آن با دو کلاسه بند به عنوان راه کار ارائه شده است. نتایج

در سالیان اخیر نمونه‌های خاصی از روش‌های سونار^۴ مانند روش‌های خودکار تمایزی اکوستیکی RoxAnn و QTC-View توسعه یافته‌اند [۲]. با این وجود، نتایج متفاوتی برای مناطق یکسان ارائه می‌دهد. همچنین تعدادی از این روش‌ها حساس به نویزند. علاوه بر این، گاهی اوقات برچسب‌گذاری علائم اکوستیکی^۵ بستر دریا با یک روش طبقه‌بندی خاص، ممکن است لزوماً مشابه با علائم زمین‌شناسی نباشد. بنابراین، اغلب روش‌های جاری طبقه‌بندی بستر اکوستیکی، اساساً ابزارهایی تجربیند. پس، توسعه روش‌های تمایزی اکوستیکی زمین هنوز مورد علاقه محققان است [۳].

با توجه به ظاهر فوق‌العاده بافتی^۶ تصاویر سونار، فنون تحلیل بافت در واقع یک انتخاب مناسب برای تصاویر اکوستیکی بستر دریا هستند. اخیراً، محققین برمدل‌های بافت فضایی چند وضوحی^۷ و چند مقیاسی از جمله فیلتر گابور^۸ و تبدیلات موجکی^۹ تمرکز کرده‌اند [۴ و ۵]. در [۳ و ۶] نیز، یک روش ادغامی جدید بر اساس تبدیلات استاندارد موجکی و منطق فازی به عنوان پوششی برای طبقه‌بندی و بخش‌بندی تصاویر بافتی بستر دریا ارائه شده است. همچنین آنان یک الگوریتم تقسیم و ادغام اصلاح شده بر پایه محتوای تبدیلات کانتورلت زیرنمونه‌ای^{۱۰} برای این منظور ارائه داده‌اند [۷].

مشکل متداول این الگوریتم‌ها، پیچیدگی محاسباتی است که ممکن است مشکلاتی برای پردازش زمان واقعی^{۱۱} ایجاد کنند. علاوه بر این اغلب آنان حساس به نویزند. اخیراً، اپراتور الگوی دوتایی محلی (LBP)^{۱۲} در چندین برنامه پردازش تصویر نظیر شناسایی صورت، کشف حالات چهره، تحلیل تصویر سنجش از راه دور، طبقه‌بندی تصاویر اثر انگشت و... مورد استفاده قرار گرفته است. از آنجایی که الگوی دودویی محلی، هم از مشخصه‌های آماری و هم ساختاری بافت

4. Sonar

۵. Acoustic

۶. texture

۷. Multi Resolution Space

۸. Gabor filter

۹. Wavelet transform

۱۰. SubSampled Contourlet

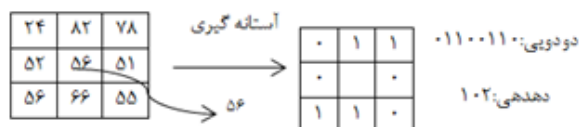
۱۱. Real time processing

۱۲. Local Binary Pattern

۱۳. Completed Robust LBP

مکانی پیکسل‌ها از بین رفته و فقط فراوانی مقادیر خاکستری پیکسل‌ها حساب می‌شود، موقعیت مکانی پیکسل‌ها نیز در نظر گرفته می‌شود. بعد از تشکیل ماتریس هم وقوعی بایستی بدنبال استخراج ویژگی‌های مناسبی بود که متمایزکننده یک کلاس از دیگری باشد. اثبات تفکیک‌کنندگی مناسب و دیگر ویژگی‌ها از قبیل تغییرناپذیری در تغییرات یکنواخت سطح خاکستری و کارایی محاسباتی، این روش را یکی از مناسب‌ترین روش‌ها برای آنالیز تصویر تبدیل کرده است.

اپراتور الگوی دودویی محلی اصلی پیکسل‌های تصویر را با اعداد دهدهی مشخص می‌نماید که تحت عنوان کدهای الگوی دودویی محلی خوانده شده می‌شود که می‌توانند ساختار محلی هرکدام از پیکسل‌ها را کدگذاری نمایند. به این ترتیب آن‌گونه که در شکل ۱ توضیح داده شده است هرکدام از پیکسل‌ها با هشت پیکسل همسایه در مجاورت سه در سه با کم کردن مقادیر پیکسل مرکزی مقایسه شده است و هر کدام از خانه‌ها مقدارش کمتر از مقدار مرکزی باشد صفر را نشان می‌دهد و موارد دیگر با یک همراه هستند. برای هرکدام از پیکسل‌های مشخص، یک عدد دودویی با استفاده از تمامی این مقادیر دودویی در یک جهت بلوکی ارائه می‌گردد که از قسمت‌های مجاور بالا سمت چپ شروع شده است.



شکل ۱: نمونه ای از اپراتور الگوی دودویی محلی

یکی از محدودیت‌های اپراتور الگوی دودویی محلی پایه این است که قسمت مجاور سه در سه کوچک آن نمی‌تواند ویژگی‌های غالب را با ساختارهای مقیاس وسیع حاصل نماید. برای ارتباط با بافت، مقیاس‌های متفاوت اپراتور به نحوی عمومی‌سازی شده است که می‌تواند از قسمت‌های مجاور با اندازه متفاوت استفاده نماید. قسمت مجاور محلی به عنوان مجموعه‌ای از نقاط نمونه‌گیری مطرح هستند که بر روی یک دایره قرار گرفته‌اند و در پیکسل‌های مشخص واقع شده‌اند و نقاط نمونه‌گیری که در پیکسل‌ها قرار نمی‌گیرند با استفاده از درون‌یابی دودویی حاصل شده‌اند و به این ترتیب اجازه می‌دهند تا هر شعاع و هر تعدادی از نقاط نمونه‌گیری گوناگون در مجاورت حاصل گردند. شکل ۲

تجربی آزمایشات در بخش پنج و نتیجه گیری در بخش ششم گزارش شده است.

۲. الگوهای دودویی محلی

الگوی دودویی محلی یکی از روش‌هایی است که در عین سادگی می‌تواند ویژگی‌های مناسبی برای طبقه‌بندی بافت تصویر با دقت بالا تولید کند که اخیراً تحقیقات در زمینه شناسایی چهره سبب ایجاد این توصیف‌گر قدرتمند بافتی شده است. این توصیف‌گر، اطلاعات مناسب را از تغییر روشنایی و حالات تصویر بدست می‌آورد. در روش معمول الگوی محلی دودویی، از هیستوگرام برای استخراج ویژگی‌ها استفاده می‌کند. برای طبقه‌بندی تصاویر بافتی می‌توان در مرحله آموزش الگوی محلی دودویی را پیکسل به پیکسل روی کل تصویر اعمال کرد و بردار ویژگی‌های مبنا را بدست آورد.

اولین توسعه الگوی دودویی محلی بنام الگوی یکنواخت^{۱۴} در سال ۲۰۰۰ صورت گرفت. اگر الگوی محلی دودویی حداکثر شامل دو گذار صفر به یک و یا یک به صفر باشد یکنواخت نامیده می‌شود. الگوی دودویی محلی مستقل از چرخش^{۱۵} توسعه‌ای دیگر از الگوی محلی دودویی است. زمانی که یک تصویر در صفحه می‌چرخد تمام همسایه‌ها حول پیکسل مرکزی در یک جهت خواهند چرخید. این اثر چرخش منجر به مقادیر متفاوتی برای الگوی دودویی محلی می‌شود. برای این منظور، محققین توسعه جدیدی از الگوی محلی دودویی که مستقل از چرخش می‌باشد را معرفی کرده‌اند. عبارت مستقل از چرخش در اینجا مربوط به تغییرات مکانی وابسته به تغییرات نور و یا اشیا متفاوت نیست. مستقل از چرخش برای برطرف کردن اثر نا مطلوب چرخش، از عملگر چرخشی بی‌تی دایره‌ای به سمت راست استفاده می‌شود تا با چندین تکرار تمامی کدهای دودویی که با این بیت‌ها می‌توان تولید کرد را بدست آورد و سپس کمترین مقدار دهدهی را از الگوی دودویی بدست آورد.

ماتریس هم وقوعی^{۱۶} نیز روشی از الگوی دودویی محلی است که در واقع بیان فرکانس‌ها می‌باشد که دو پیکسل همسایه با شدت خاکستری خاص خودشان و با فاصله ثابت روی تصویر دارند. اگر فاصله دو پیکسل برابر یک باشد زاویه‌های ممکن در چهار جهت بین دو پیکسل با صفر، ۴۵، ۹۰ و ۱۳۵ درجه بیان می‌شوند. در ماتریس هم وقوعی، بر خلاف هیستوگرام ساده که در آن اطلاعات

^{۱۴}. Uniform Patterns

^{۱۵}. Independent of the Rotation Local binary pattern

^{۱۶}. Co-Occurrence Matrix

دودویی محلی ثابت چرخشی در [۸] پیشنهاد شده است که آنرا در فرمول (۳) می‌بینید:

$$LBP_{P,R}^r i = \min\{ROR(LBP_{P,R}, i) | i = 0, 1, \dots, p-1\} \quad (3)$$

که در آن $ROR(x, i)$ به عنوان یک تغییر متغیر سمت راست دایره‌ای کوچک در تعداد P بیت به صورت (x, i) اعلان

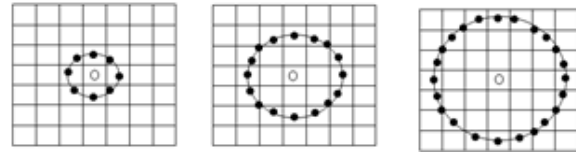
می‌گردد. اپراتور $LBP_{P,R}^r i$ مقادیر آماری الگوهای انتشاری متغیر غیر چرخشی گوناگون را نشان می‌دهد که به ریز ویژگی‌های مشخص شده درون تصویر بستگی دارد و در حقیقت می‌تواند الگوهایی را نشان دهد که به عنوان آشکار ساز ویژگی^{۱۸} لحاظ شده است. با این وجود در [۱۰] نشان داده شده است که چنین اپراتور الگوی محلی چرخشی ثابت ضرورتاً نمی‌تواند اطلاعات متمایزی را ارائه دهد زیرا تناوبات الگوهای انتشاری، جداگانه تلفیق شده و همان‌طور که در فرمول $LBP_{P,R}^r i$ نشان داده شده تفاوت‌های عمده‌ای با یکدیگر داشته و فقط قابلیت تعیین مقادیر خام فضای زاویه-دار در فواصل ۴۵ درجه امکان‌پذیر می‌گردد.

این نشان می‌دهد که الگوهای مجتمع شامل اطلاعاتی وسیع‌تر از این‌ها می‌باشد. این امکان وجود دارد که با استفاده از یک زیر مجموعه $(2P)$ تایی از الگوهای دودویی، بافت تصویر را توصیف کرد. اوجالا و همکارانش الگوهای $LBP_{P,R}^r i$ را یکنواخت نامیدند. مشاهده شده است که الگوهای یکنواخت برای ۹۰٪ تمامی الگوها در یک همسایگی (۱ و ۸) به صورت مجاور و حدود ۷۰٪ در یک مجاورت (۲ و ۱۶) در تصویر بافت ارائه شده‌اند [۱]. آزمایش مشابه در پایگاه داده‌های FERET انجام پذیرفته است و مشخص شده که ۹۰٫۶٪ الگوها در یک مجاورت (۱ و ۸) و ۸۵٫۲٪ در یک مجاور (۲ و ۸) یکنواخت هستند [۱۱].

۳. مروری بر کارهای انجام شده الگوهای دودویی محلی

الگوی دودویی محلی اخیراً با تفاوت‌های عمده‌ای از عملکرد بهبود یافته در کاربردهای متفاوت توسعه یافته است. این تفاوت‌ها با تاکید بر جنبه‌های متفاوت اپراتور الگوی دودویی محلی اصلی ایجاد شده‌اند که عبارتند از: افزایش قابلیت

نمونه‌ای از شکل توسعه‌یافته اپراتور الگوی دودویی محلی یعنی الگوی دودویی محلی طویل شده (ELBP)^{۱۹} را نشان می‌دهد که در آن علامت (P, R) نشان‌دهنده P نمونه مجاور به صورت نقاطی بر روی یک دایره با شعاع R می‌باشد.



شکل ۲: نمونه‌ای از اپراتور دودویی محلی طویل شده،

دوایر (۸ و ۱)، (۱۶ و ۲) و (۲۴ و ۳) در مجاورت مشاهده

می‌گردند.

به طور رسمی با احتساب یک پیکسل خاص در فرمول زیر، الگوی دودویی محلی حاصل می‌تواند در قالب دهدهی به صورت معادله زیر نشان داده شود:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c) 2^p \quad (1)$$

که در آن i, c و p بترتیب شامل مقادیر سطح خاکستری پیکسل مرکزی و پیکسل‌های اطراف و P تعداد پیکسل‌های اطراف بر روی دایره مجاور با شعاع R می‌باشد. تابع S نیز به صورت فرمول زیر نشان محاسبه می‌شود:

$$S(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (2)$$

با توجه به تعریف فوق، اپراتور الگوی دودویی محلی پایه برای تغییر شکل‌های مقیاس خاکستری یکنواخت تغییر نمی‌کند که می‌تواند رده تمرکز پیکسل را در مجاورت محلی حفظ نماید و تصویری از برجسب‌های الگوی دودویی محلی محاسبه شده در یک منطقه به عنوان توصیف‌گر بافت نشان دهد.

اپراتور فرمول (۲) می‌تواند مقادیر خروجی متفاوتی مربوط به الگوهای محلی متفاوت را توسط P پیکسل مجاور نشان دهد. در صورتیکه تصویر بچرخد این پیکسل‌های اطراف، در هرکدام از قسمت‌های مجاور به همین ترتیب در مجاورت دایره حرکت می‌کنند و به این ترتیب نتیجه آن یک مقدار الگوی دودویی متفاوت می‌باشد و تنها استثنای الگوها یک و صفر خواهد بود. به منظور حذف چرخش تصویر، یک الگوی

^{۱۸}. Feature Detector

^{۱۹}. Elongated LBP

۱-۳- افزایش قابلیت تمایز

اپراتور الگوی دودویی محلی تعداد مشخصی از الگوها را برای توصیف ساختارهای محلی نشان می‌دهد. افزایش قابلیت تمایز آنها، نیازمند الگوهای بیشتری از اطلاعات قابل کدگذاری است. ژائو و همکارانش اپراتور الگوی دودویی محلی را برای توصیف اطلاعات ساختاری محلی بیشتری در شرایط مشخص اصلاح نموده‌اند [۱۲]. به ویژه اینکه الگوی دودویی محلی بهبودیافته یا میانه (MLBP/ILBP)^{۱۹} را پیشنهاد داده‌اند که می‌تواند تمامی پیکسل‌ها شامل پیکسل مرکزی را با میانگین شدت تمامی پیکسل‌ها در مسیر مورد نظر مقایسه نماید.

وجود همسایگی سه‌تایی در الگوی دودویی محلی بهبود یافته با ۵۱۱ الگو (۱-۲^۹ عدد صفر و یک) برای استفاده از قسمت‌های مجاور با هر اندازه خاص به جای مورد اصلی رگه سه در سه توسعه یافته است [۱۳]. تقریباً در همین زمان یک ساختار مشابه برای توسعه تبدیلات هماهنگ به عنوان تبدیلات هماهنگ اصلاح شده استفاده شده است [۱۴] که میانگین الگوی دودویی محلی ارائه گردیده مشابه با الگوی دودویی محلی بهبود یافته می‌باشد با این تفاوت که در این مورد پیکسل مرکزی مد نظر قرار نگرفته است.

گو و همکاران روش الگوی دودویی محلی هامینگ^{۲۰} را برای بهبود قابلیت تمایز الگوی دودویی محلی اصلی پیشنهاد داده‌اند [۱۵]. آنها الگوهای غیر یکنواخت را مجدداً بر مبنای فاصله همینگ طبقه‌بندی نموده‌اند که به جای جمع‌آوری آنها در یک مجموعه، جداگانه این کار را می‌کنند. در الگوی دودویی محلی همینگ، این الگوهای غیر یکنواخت به صورت الگوهای یکنواخت موجود با به حداقل رسانیدن فاصله همینگ بین آنها تلفیق شدند. برای نمونه می‌توان به الگوهای غیر یکنواخت (۱۰۰۰۱۱۱۰) که به الگوی یکنواخت (۱۰۰۰۱۱۱۱) تبدیل می‌گردند اشاره کرد، زیرا فاصله همینگ آنها یک می‌باشد. هنگامی که چندین الگوی یکنواخت فاصله همینگ مشابهی را با الگوی غیر یکنواخت

ایجاد تمایز، افزایش قدرت، انتخاب مجاورت، توسعه به داده‌های سه بعدی و ترکیب با رویکردهای دیگر. در این بخش تفاوت‌های مشخص الگوهای محلی دودویی را مورد بررسی قرار می‌دهیم (برای مرور جدول ۱ را بررسی کنید).

جدول ۱: فهرستی از تغییرات جدید الگوهای

دودویی محلی

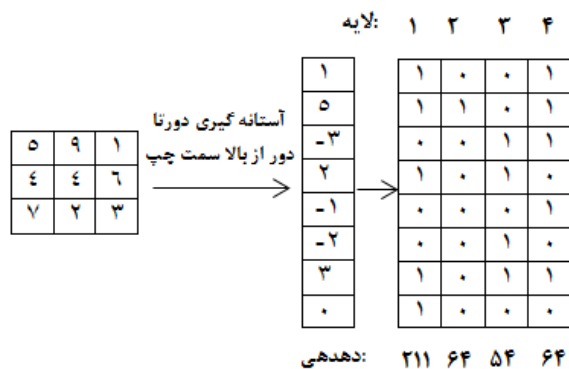
بخش فرعی	نوع الگو	خصوصیات
افزایش قابلیت ایجاد تمایز	الگوی دودویی محلی بهبود یافته (الگوی دودویی محلی میانه)	در نظر گرفتن اثرات پیکسل‌ها، ارائه الگوهای ساختاری کامل
	الگوی دودویی محلی همینگ	تلفیق الگوهای غیر یکنواخت به صورت الگوهای یکنواخت
	الگوی دودویی محلی توسعه یافته	تمایز ساختن الگوهای دودویی محلی که سبب ایجاد ابعاد عمده می‌گردد
بهبود قدرت	الگوی دودویی محلی تکمیل شده	شامل اطلاعات علامت و اطلاعات دامنه یک منطقه محلی مشخص می‌باشد
	الگوهای چهارگانه محلی	در یک آستانه جدید وارد می‌گردد، عدم وجود هیچ گونه متغیر کاملاً ثابت در تبدیلات سطح خاکستری
انتخاب مجاورت	الگوی دودویی محلی طولی شده	هیچ گونه متغیری برای تغییرات مقیاس خاکستری یکنواخت مشاهده نمی‌گردد، سبب ایجاد پیچیدگی محاسباتی بالا خواهد شد
	الگوی دودویی محلی چند بلوکی	استخراج اطلاعات غیر مشابه و از دست رفتن اطلاعات غیر مشابه، عدم تفاوت در فرآیند چرخش
	الگوی دودویی محلی سه‌چهار دسته‌ای	دستیابی به اطلاعات ساختاری خرد و کلان
توسعه به داده‌های سه بعدی	الگوی دودویی محلی سه بعدی	کد گذاری دسته‌ای اطلاعات بافتی
	الگوی دودویی محلی حجمی (الگوی دودویی محلی قله)	توسعه الگوی محلی دودویی به داده‌های حجمی سه بعدی
ترکیب با دیگر رویکردها	ترکیب الگوی دودویی محلی و موجک گابور	توصیف بافت پویا که سبب ایجاد بعد پذیری بالا می‌گردد
	ترکیب الگوی دودویی محلی و تبدیلات ویژگی مقیاس ثابت	ترکیب مزایای الگوی دودویی محلی و موجک گابور یعنی افزایش هزینه زمانی و ایجاد ابعاد بالا به طور همزمان
	ترکیب الگوی دودویی محلی و هیستوگرام فوری	بهره‌مندی از مزایای تبدیلات مقیاسی ویژگی‌های یکنواخت، کاهش طول بردار ویژه
		دستیابی به عدم تغییرپذیری چرخشی برای کل تصویر

^{۱۹}. Improved LBP (Mean LBP)

^{۲۰}. Hamming LBP

را توصیف می‌نماییم اگرچه الگوی دودویی محلی لایه اول به قدر کافی متمایز نیست، اطلاعات کدگذاری شده در لایه‌های اضافی دیگر برای تشخیص آنها مورد استفاده قرار گرفته است. حرکت رو به پایین آن نشان می‌دهد که الگوی دودویی محلی توسعه یافته می‌تواند ابعاد ویژگی‌ها را تا حد زیادی افزایش دهد.

اخیرا گو و همکارانش یک الگوی دودویی محلی کامل را پیشنهاد داده‌اند که کارش مشابه با الگوی دودویی محلی توسعه یافته می‌باشد [۱۷]. بعلاوه، الگوی دودویی محلی کامل شده شامل هر دو مورد نشانه و گرادیان دوم بین یک پیکسل مرکزی مشخص و قسمت‌های مجاور به منظور بهبود نیروی تمایز اپراتور الگوی دودویی محلی اصلی می‌باشد. بر خلاف استراتژی کدگذاری بیت دودویی که توسط الگوی دودویی محلی توسعه یافته مورد استفاده قرار می‌گیرد، الگوی دودویی محلی کامل شده مقادیر مطلق گرادیان را با پیکسل مرکزی مشخص برای ایجاد یک کد دودویی محلی گونه ترکیب می‌نماید.



شکل ۳. نمونه‌ای از اپراتور الگوی دودویی محلی توسعه یافته

دارد موردی با حداقل فاصله یو کلیدی^{۲۱} قابل انتخاب خواهد بود.

الگوی دودویی محلی توسعه یافته (E-LBP)^{۲۲} یکی دیگر از رویکردها برای بهبود قابلیت تمایز الگوی دودویی محلی می‌باشد [۱۶]. اپراتور الگوی محلی توسعه یافته نه تنها مقایسه دودویی و دوگانه‌ای را بین پیکسل‌های مرکزی و قسمت‌های مجاور آن انجام می‌دهد بلکه می‌تواند تفاوت‌های دقیق ارزش خاکستری آن‌ها را با استفاده از برخی از واحدهای دودویی اضافی کدگذاری نماید. می‌توان گفت که ویژگی الگوی دودویی محلی توسعه یافته از چندین مورد از ویژگی‌های الگوهای دودویی محلی تشکیل شده که در چندین لایه، کدگذاری را انجام می‌دهند و می‌توانند مقدار گرادیان^{۲۳} بین پیکسل مرکزی و همچنین پیکسل‌های مجاور را کدگذاری کنند.

همان‌گونه که در شکل ۳ نشان داده می‌شود اولین لایه الگوی دودویی محلی توسعه یافته در واقع به عنوان کد الگوی دودویی محلی اصلی خوانده می‌شود که می‌تواند نشانه گرادیان را کدگذاری نماید. لایه‌های زیر از الگوی دودویی محلی توسعه یافته می‌توانند مقادیر مطلق گرادیان را کدگذاری کنند. در هر سمت مقادیر مطلق گرادیان برای مرتبه اول در نمونه دودویی کدگذاری شدند و در ادامه تمامی مقادیر دودویی در یک لایه مشخص به الگوی دودویی محلی اضافی ختم می‌شود. برای نمونه در شکل ۳، اولین لایه به عنوان کد الگوی دودویی محلی اصلی خوانده شده است که می‌تواند نشانه گرادیان را کدگذاری کند و به این ترتیب مقدار دهدهی ۲۱۱ را از شکل دودویی (۱۱۰۱۰۰۱۱) حاصل نماید. با استفاده از یک ساختار وزنی مشخص الگوی محلی در تمامی بیت‌های دودویی، کد الگوی دودویی محلی توسعه یافته لایه مربوط به آن ایجاد می‌گردد، برای مثال در لایه دوم مقدار دهدهی آن ۶۴ می‌باشد و لایه سوم ۵۴ و در پایان مقدار دهدهی لایه چهارم ۲۳۴ می‌باشد. در نتیجه هنگامی که بافت‌های محلی مشابه

^{۲۱}. Euclidian

^{۲۲}. Extended-LBP

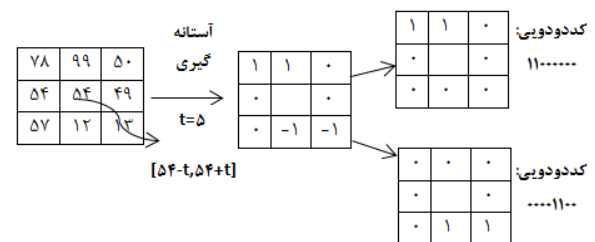
^{۲۳}. Gradient

۳-۲- بهبود قدرت

الگوی دودویی محلی به نويز حساس است زیرا آستانه اپراتور دقیقاً مقدار پیکسل مرکزی می‌باشد. برای رفع این مسئله تان و تریگس نسخه الگوی دودویی محلی اصلی را به نسخه‌ای با کدهای دارای سه مقدار متفاوت توسعه دادند که تحت عنوان الگوهای سه‌گانه محلی خوانده می‌شود (LTP)^{۲۴} [۱۸]. در الگوی سه‌گانه محلی شاخص S(X) معادله (۱)، توسط فرمول (۴) جایگزین می‌گردد:

$$s(i_n, i_c, t) = \begin{cases} 1, & i_n \geq i_c + t \\ 0, & |i_n - i_c| < t \\ -1, & i_n \leq i_c - t \end{cases} \quad (4)$$

کدهای الگوهای سه‌گانه محلی مقاومت بیشتری به نويز دارند اما عدم تحریک‌پذیری را برای تغییر شکل‌های سطح خاکستری نشان نمی‌دهند. یک رویکرد کدگذاری برای تقسیم‌بندی هرکدام از الگوهای سه‌گانه به دو قسمت تقسیم می‌شود که شامل مقادیر مثبت و مقادیر منفی می‌باشد (به شکل ۴ توجه کنید). یکی از مشکلات الگوهای سه‌گانه محلی تنظیم آستانه می‌باشد که چندان ساده نیست.



شکل ۴. نمونه‌ای از اپراتور الگوی چهارگانه محلی

محلی

الگوی دودویی محلی نرم (SLBP)^{۲۵} که در [۱۹] معرفی شده است بجای معادله (۲)، دو معادله عضویت فازی (۵) و (۶) را بکار می‌گیرد:

$$S_{1,d}(x) = \begin{cases} 0, & x < -d \\ 0.5 + 0.5 \frac{x}{d}, & -d \leq x \leq d \\ 1, & x > d \end{cases} \quad (5)$$

$$S_{0,d}(x) = 1 - S_{1,d}(x) \quad (6)$$

پارامتر d مقادیر فازی بودن را کنترل می‌نماید که توسط تابع فازی انجام می‌پذیرد. هنگامی که مجاور محلی از P نمونه تشکیل شده است، هیستوگرام با یک اپراتور الگوی یکنواخت تحت عنوان $(0, 1, \dots, 2^p - 1)$ می‌باشد. سهم هر پیکسل جداگانه (x_c, y_c) برای مشخص نمودن h از هیستوگرام برابر با فرمول (۷) خواهد بود:

$$SLBP(x_c, y_c, h) = \prod_{p=0}^{p-1} [b_p(h) \cdot S_{1,d}(i_p - i_c) + (1 - b_p(h)) \cdot S_{0,d}(i_p - i_c)] \quad (7)$$

که در آن $b_p(h) \in \{0, 1\}$ نشان‌دهنده مقادیر عددی بیت P ام نمونه دودویی h می‌باشد. همچنین نشان داده شده است که الگوهای مشخص، محتوی اطلاعات بیشتری نسبت به موارد دیگر می‌باشند.

با احتساب الگوی دودویی محلی نرم، یک پیکسل می‌تواند در بیش از یک قسمت و بسته ^{۲۶} شرکت کند اما نسبت مجموع سهم پیکسل‌ها به تمامی بسته‌ها همیشه برابر یک می‌باشد. الگوی دودویی محلی نرم به افزایش قدرت در شرایطی می‌انجامد که تغییر مقادیر در تصویر ورودی تنها سبب تغییر محدودی در خروجی گردد. اگرچه تغییرناپذیری را از دست می‌دهد ولی با افزایش پیچیدگی محاسبات، تغییرات آن یکنواخت می‌گردد. مقادیر مناسب d در معادله (۵) همانند الگوی دودویی محلی نرم تنظیم می‌گردد.

۳-۳- انتخاب مجاور

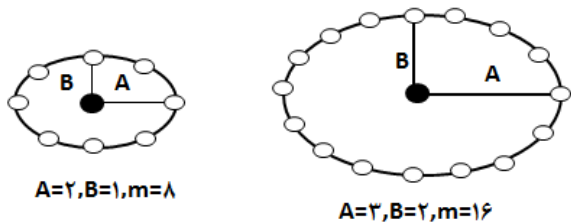
انتخاب یک مجاور مناسب برای روش دودویی محلی محور، تاثیر عمده‌ای بر روی غالب نهایی دارد. این مورد شامل تعدادی از نقاط نمونه‌گیری، توزیع نقاط نمونه‌گیری، شکل قسمت مجاور و اندازه قسمت مجاور می‌باشد.

پیکسل‌های مجاور در الگوی دودویی محلی اصلی به صورت دایره توسط لیانو و همکاران نشان داده شدند و این مبحث را بیان نمودند که دلیل اصلی برای تعریف قسمت مجاور در چنین حالت، یکسانی، همگرایی و دستیابی به چرخش غیرمتناوب برای توصیف بافت می‌باشد [۲۰]. اگرچه این مورد برای تمامی کاربردها مناسب نیست اما در مقابل اطلاعات غیر یکنواخت می‌تواند به عنوان یک ویژگی مهم به حساب آید. در نتیجه الگوی دودویی محلی طویل شده را با

^{۲۴}. Local Ternary Patterns

^{۲۵}. Soft LBP

^{۲۶}. Bin



شکل ۵: نمونه ای از اپراتور الگوی دودویی محلی طویل شده

۴-۳- توسعه به الگوی محلی سه بعدی

در چندین مورد محققان تلاش نموده‌اند تا الگوی دودویی محلی را از صفحه دو بعدی به سه بعدی یعنی الگوی دودویی حجمی (VLBP)^{۲۰} تبدیل نمایند [۲۳ و ۲۴ و ۲۵]. این مورد چندان ساده نیست و آن گونه که در نگاه اول به چشم می‌آید دو مشکل اساسی مطرح می‌گردد: اول این که مساوات در نمونه‌گیری‌ها در محدوده سه بعدی کار مشکلی است و نکته دوم اینکه به سختی می‌توان رده‌ای را برای این نقاط نمونه‌گیری مشخص نمود که نکته حائز اهمیت برای دستیابی به عدم تغییرپذیری در هنگام چرخش می‌باشد. برای نشان دادن قابلیت دستیابی الگوی دودویی محلی به اطلاعات بافت پویا، ژائو و پی‌تیکانن قسمت‌های مجاور الگوی دودویی محلی را از صفحه دو بعدی به فضای سه بعدی توسعه داده‌اند [۲۵]. اپراتور تحت عنوان الگوی دودویی محلی حجمی یا سه بعدی (VLBP یا 3D-LBP) معرفی شده است که حرکت و اطلاعات وضعیت ظاهری را با یکدیگر تلفیق نموده و به این ترتیب می‌توان از آن برای آنالیز توالی تصویر یا ویدیوها استفاده نمود. لازم به ذکر است که این رویکرد سبب می‌شود که استفاده از آنالیز بافت پویای سری‌های زمانی دو بعدی به جای داده‌های حجمی کلی سه بعدی امکان‌پذیر گردد. ویژگی‌های الگوی دودویی محلی حجمی تنها به نگاشت و چرخش حساس نیستند (چرخش حول محور) بلکه جهت تغییرات مقیاس خاکستری یکنواخت مقاوم هستند. VLBP(L,P,R) در مقایسه با LBP(P,R)، زمان غالب را به حساب می‌آورد و پارامتر L نشان‌دهنده طول فاصله زمانی می‌باشد. از یک مجاورت

پیکسل‌های مجاور قرار گرفته بر روی یک بیضی مطابق شکل ۵ پیشنهاد داده‌اند که این شکل دو مثال از الگوی دودویی محلی طویل شده را نشان می‌دهد که در آن A و B نشان‌دهنده محور طولانی تر افقی و محور کوتاه تر عمودی بوده و m نشان‌دهنده تعداد پیکسل‌های مجاور می‌باشد.

با توجه به الگوی دودویی محلی اصلی، روش درون‌یابی خطی برای پیکسل‌های مجاور مورد استفاده قرار می‌گیرد که دقیقاً در پیکسل واقع نمی‌شود. اپراتور الگوی دودویی محلی طویل شده می‌تواند حول محور مرکزی بچرخد و یک زاویه ویژه را برای تعیین نمودن ساختار محلی طویل شده در جهات متفاوت برای دستیابی به آنالیز چند جهتی ایجاد نماید. به منظور دستیابی به این ساختارها و همچنین ساختارهای بزرگتر، ژو و همکاران الگوی دودویی محلی چند بلوکی (MBLBP)^{۲۷} را پیشنهاد داده‌اند که به جای مقایسه پیکسل‌ها، میانگین تمرکز قسمت‌های فرعی مجاور را مقایسه می‌کند [۲۱]. الگوی دودویی محلی اصلی می‌تواند به عنوان یک مورد خاص از الگوی دودویی محلی چند بلوکی مورد توجه قرار گیرد. منطقه فرعی می‌تواند به عنوان مربع یا چهار گوش مشخص گردد. میانگین تمرکز در بلوک‌ها می‌تواند به شکل مشخصی با استفاده از جدول منطقه جمع‌بندی و محاسبه گردد و یا انتگرال تصویر محاسبه شود. یک طرح مشابه در [۲۲] ارائه شده است: الگوی دودویی محلی سه‌رگه (TP-LBP)^{۲۸} و چهاررگه (FP-LBP)^{۲۹} که برای مقایسه فاصله بین کل بلوک‌ها (رگه‌های) مورد نظر، بجای هر پیکسل میانگین متمرکز ارائه شده هر تابع فاصله می‌تواند مورد استفاده قرار گیرد (برای تفاضلات سطح خاکستری).

^{۲۷} . Multi Block LBP

^{۲۸} . Three Patch LBP

^{۲۹} . Four Patch LBP

^{۲۰} . Volume LBP

۳-۵- تلفیق با ویژگی‌های روش‌های دیگر

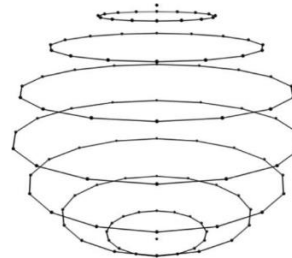
به عنوان روشی برای توصیف ویژگی‌های محلی، الگوی دودویی محلی می‌تواند با رویکردهای دیگر تلفیق گردد. برای نمونه مجموعه‌ای از رویکردها برای تلفیق موجک‌های گابور و ویژگی‌های الگوی دودویی محلی با استفاده از روش‌های متفاوت پیشنهاد شده‌اند [۲۶]. همچنین در [۲۷] محاسبه شده است که موجک گابور و ویژگی الگوی دودویی محلی محور به صورت متقابل حالت مکمل دارند زیرا الگوی دودویی محلی می‌تواند جزئیات وضعیت ظاهری محلی را ایجاد نموده در حالی که موجک گابور اطلاعات شکل را در یک محدوده وسیع‌تر از مقیاس نشان می‌دهد. یک استراتژی جفت شدن ساده برای استخراج اولیه ویژگی‌های گابور و الگوی دودویی محلی به طریقی موازی مطرح شده‌اند و در ادامه دو نوع متفاوت از ویژگی‌ها در سطح ویژگی در کنار یکدیگر قرار گرفته‌اند که با سطح امتیاز یا سطح تصمیم هماهنگ می‌باشند [۲۷ و ۲۸].

یکی دیگر از روش‌ها، تلفیق و ترکیب استراتژی سریالی است که در اصل از کاربرد فیلترهای گابور تشکیل شده و در ادامه الگوی دودویی محلی برای تصویر خام مورد استفاده قرار می‌گیرد [۲۹]. نگاشت‌های ویژه گابور (GFM) [۳۲] چندگانه توسط تصاویر ورودی در هم پیچیده به کمک فیلترهای گابور چند مقیاسی و چند جهتی نمود یافته و محاسبه شده است. هر نگاشت ویژه گابور به مناطق بدون هم‌پوشانی کوچک تقسیم‌بندی شده‌اند که هیستوگرام‌های دودویی محلی از آنها حاصل شده و در نهایت می‌توانند درون یک هیستوگرام دارای ویژگی‌های مشخص قرار گیرند. الگوهای هیستوگرام چند رزولوشنی متغیر محلی (MHLVP) [۳۳] [۲۰] و همچنین الگوی هیستوگرام دودویی گابور محلی [۳۰] بر مبنای چنین فرایندی پیشنهاد شده‌اند. اخیراً نویسندگان یک روش سریالی مشابه را با استفاده از هر دو مورد موجک و الگوی دودویی محلی پیشنهاد داده‌اند که در آغاز از موجک برای تجزیه تصاویر خام به درون تصاویر چهار

محلی کوچک حجمی، مقایسه پیکسل‌های مجاور با پیکسل‌های مرکزی در برخی از واحدهای دودویی حاصل می‌شود و مقادیر وزنی برای این واحدها به عنوان یک خط مارپیچ ارائه شده است.

به منظور ایجاد سادگی محاسباتی الگوی دودویی محلی حجمی و آسانی توسعه آن، فقط هم‌وقوعی در سه بعد جداگانه مدنظر قرار می‌گیرد. بافت‌ها همراه با هیستوگرام‌های دودویی محلی موجود مدل‌سازی شده‌اند و از سه صفحه عمود بر یکدیگر گرفته شده‌اند و به این ترتیب این نسخه ساده از الگوی دودویی محلی حجمی تحت عنوان الگوی دودویی محلی قله^{۳۱} خوانده می‌شود. نمونه‌گیری مرسوم مدور توسط یک بیضی به نحوی جایگزین شده است که پارامترهای شعاعی متفاوت در حوزه فضایی و زمانی قابل جایگزینی خواهند بود.

فهر تغییر شکل هارمونیک و موزون کروی را برای ایجاد یک پایه متعامد در دو کره معرفی نموده و در ادامه ویژگی‌های الگوی دودویی محلی را در حوزه تناوب محاسبه نموده است [۲۵]. این روش بر هر دو مورد مشکلات فوق غلبه می‌نماید. محققین راه‌حل‌های دیگری را برای الگوی دودویی محلی در سه بعد پیشنهاد داده‌اند. آنها از برخی دوایر برای نشان دادن کره استفاده نموده و پارامتر S را اضافه کرده و به این ترتیب اپراتور $LBP(S,P,R)$ را معرفی کردند که در شکل ۶ می‌توانید آن را ببینید و همچنین قوانین یکنواخت را به صورت دو بعدی ارائه داده‌اند. این روش سبب می‌شود که مسئله تمایز بافت مشابه با توصیف الگوی دودویی محلی پیش رود.



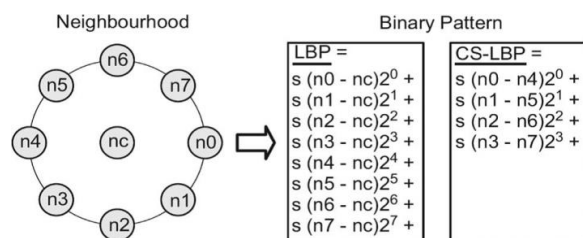
شکل ۶. یک مثال از الگوی دودویی محلی سه‌بعدی ($S=9$) و ($R=2$ و $P=16$)

^{۳۲} . Gabor Feature Maps

^{۳۳} . Multi Resolution Histograms of Local Variation Patterns

^{۳۱} . LBP-TOP

همچنین اخیراً برای تلفیق الگوی دودویی محلی و تبدیلات ویژگی ثابت مقیاسی به طریق سریالی پیشنهاداتی را ارائه داده‌اند. اولین مورد، ارائه اپراتور الگوی دودویی محلی اصلی در تصویر ورودی با استفاده از مقیاس متفاوت برای استخراج چندین مورد تصاویر که الگوی دودویی محلی چند مقیاسی (MS-LBP) نامیده می‌شود و در ادامه تبدیلات ویژگی ثابت مقیاسی با تصاویر حاصل از الگوی دودویی محلی چند مقیاسی برای ایجاد تطابق محلی مورد استفاده قرار گرفته است.



شکل ۷. ویژگی‌های الگوی دودویی محلی متقارن مرکزی و

الگوی دودویی محلی برای مجاورت هشت پیکسل

همچنین آهونن رویکردی را پیشنهاد داد که تحت عنوان ویژگی‌های فوریه هیستوگرام الگوی دودویی محلی (LBP-HF)^{۳۹} نام‌گذاری شده‌اند تا به این ترتیب بتوانند الگوی دودویی محلی و تغییر شکل فوریه متمایز (DFT)^{۴۰} را با یکدیگر تلفیق کنند. برخلاف ویژگی‌های الگوی دودویی محلی چرخشی ثابت، توصیف‌گر فوریه هیستوگرام الگوی دودویی محلی با مقایسه یک هیستوگرام الگوی محلی در کل منطقه ایجاد شده است و در ادامه ویژگی‌های غیر متغیر چرخشی از هیستوگرام با تغییر شکل متمایز فوریه ایجاد گردیده است. این مورد به این معناست که عدم تغییرپذیری چرخش به طور کلی حاصل شده و به این ترتیب ویژگی‌ها برای چرخش‌های کل سیگنال ورودی بدون تغییر باقی می‌مانند اما هنوز هم اطلاعاتی در رابطه با توزیع نسبی جهت‌گیری‌های متفاوت یک الگوی دودویی محلی یکنواخت مطرح است.

فرکانسی استفاده گردیده است که برای نمونه می‌توان به فرکانس افقی بالا با تناوب کم، فرکانس بالای عمودی و فرکانس بالای دو وجهی اشاره نمود که به عنوان نهاده‌هایی برای الگوی دودویی محلی اصلی مطرح هستند.

با انگیزه‌ای که توسط الگوی دودویی محلی قله و الگوی هیستوگرام دودویی گابور محلی (LGBP)^{۳۴} برای آنها ایجاد گردیده است یک حجم تصویر محور گابور رده سوم را ایجاد نموده و در ادامه از الگوی دودویی محلی اصلی برای سه صفحه متعامد برای استخراج اطلاعات متمایز نه تنها در حوزه فضایی بلکه در حوزه‌های جهت‌گیری و تناوب گابور استفاده نموده‌اند. برای کاهش پیچیدگی محاسباتی، اپراتور محلی دودویی متغیر سراسری (GV-LBP)^{۳۵} برای توصیف تغییرات صورت گرفته در حوزه‌های جهت‌گیری و تناوب فضایی به صورت هم زمان با تعریف جهت‌گیری‌ها و نقاط مجاور مقیاس در تصاویر گابور متفاوت پیشنهاد شده‌اند.

یک الگوی دودویی محلی متقارن مرکزی (CS-LBP)^{۳۶} تنها توسط مقایسه‌ای از جفت پیکسل‌های مجاور پیشنهاد شده است که در قطر مشابه دایره قرار می‌گیرند [۳۱]. چنین تفاضلاتی می‌توانند الگوی دودویی محلی را با تبدیلات ویژگی ثابت مقیاسی (SIFT)^{۳۷} ترکیب نمایند و به این ترتیب واحدهای دودویی کمتری را نسبت به الگوی دودویی محلی اصلی ایجاد می‌نمایند. تفاوت بین الگوی دودویی محلی متقارن مرکزی و الگوی دودویی محلی با هشت پیکسل مجاور در شکل ۷ ارائه شده است.

در ادامه، دیگر محققین الگوهای دودویی مرکزی (CBP)^{۳۸} را معرفی نموده‌اند که یک اصلاح محدود را برای این ویژگی سبب می‌گردد و پیکسل مرکزی را با میانگین مقادیر تمامی پیکسل‌ها در مجاورت برای ایجاد یک واحد دودویی اضافی تلفیق می‌نماید که بیشترین مقادیر وزنی را برای تاکید بر تاثیرات بیت دودویی جدید به خود اختصاص می‌دهد.

^{۳۴} . local Gabor Binary Pattern Histogram

^{۳۵} . Global Variable LBP

^{۳۶} . Center-symmetric LBP

^{۳۷} . Scale-invariant Feature Transform

^{۳۸} . Centralized Binary Patterns

^{۳۹} . LBP Histogram Fourier Features

^{۴۰} . Discrete Fourier Transform

۳-۶- انتخاب ویژگی الگوی دودویی محلی

در اغلب فعالیت‌های صورت گرفته الگوی دودویی محلی، تصویر ورودی به مناطق کوچک‌تر تقسیم‌بندی می‌گردد و هیستوگرام حاصل از آن به صورت هیستوگرام‌های محلی بهم پیوسته طولانی به بردار ویژگی‌های طولانی‌تر با ابعاد $O(10^3)$ توسعه یافته‌اند. بعلاوه برخی از متغیرهای جدید سبب افزایش طول بردار ویژگی به طور چشمگیری خواهند شد که از آن جمله می‌توان به الگوی دودویی محلی طویل شده، الگوی دودویی حجمی و الگوی دودویی محلی موجکی گابور اشاره نمود. اعتقاد بر این است که بردار ویژگی الگوهای محلی و محور حاصل می‌تواند یک نمونه کامل را با اطلاعات فراوان ارائه نماید که به عنوان مورد متمایزتر و متراکم تر مطرح می‌گردد. بعلاوه هنگام ایجاد سیستم‌های زمان واقعی دارا بودن یک نمونه الگوی محلی دودویی با کاهش طول ویژه مطلوب است و به تمامی این دلایل، مسئله انتخاب ویژگی الگوی دوتایی اخیراً در بسیاری از منابع پیشنهاد داده شده است. می‌توانیم این روش را به دو طبقه‌بندی تقسیم نماییم: اولین مورد کاهش‌دهنده طول ویژگی‌ها بر مبنای برخی از قوانین همانند الگوهای یکنواخت می‌باشد در حالیکه مورد دیگر استفاده کردن از روش‌های انتخاب ویژگی برای انتخاب الگوهای تمایزی می‌باشد. هر دو مورد جریان‌ها نکات مثبت و نکات منفی منحصر به فرد خود را دارند. اولین مورد ساده است اما قابلیت انتخاب ویژگی محدودی دارد و در مقابل مورد دوم ظرفیت انتخاب ویژگی بهتری را نشان می‌دهد اما معمولاً نیازمند آموزش آفلاین می‌باشد که می‌تواند از نقطه‌نظر محاسباتی گران‌قیمت باشد. فراگیری افزایشی یک روش کارآمد را برای انتخاب ویژگی‌ها نشان می‌دهد. با تغییر و تعیین مقیاس تصویر در زیر دریچه‌ها، مناطق فرعی بیشتری برای دستیابی به هیستوگرام محلی الگوهای دودویی حاصل شده است و فاصله بین هیستوگرام‌های مربوطه از دو تصویر به عنوان یک ویژگی تمایزی بیان شده است. روش آدابوست^{۴۱} برای فراگیری تعداد محدودی از کارآمدترین ویژگی‌ها مطرح گردیده است. این رویکرد می‌تواند به عملکردی دست یابد

که اندکی بهتر است اما تعداد محدودی از هیستوگرام‌های محاسبه شده از مناطق محلی را دارا می‌باشد. از آنجا که تمامی پنجره‌ها ضرورتاً برای تامین اطلاعات مفید مطرح نیستند آزمایشات آنها نشان می‌دهد که بین‌های الگوهای انتخابی می‌توانند نمونه متراکم‌تری را با طول کاملاً کاهش یافته بردار ویژگی ارائه نمایند در حالیکه عملکرد بهتری را ایجاد می‌کنند. آدابوست تا حد زیادی برای انتخاب ویژگی‌های الگوهای دودویی محلی در وظایف گوناگون مطرح شده است [۳۴، ۳۳، ۳۲، ۳۱ و ۳۵].

فراگیری فضای فرعی و یا کاهش بعدپذیری، مجموعه داده‌ها را از فضای دارای ابعاد بالا به یک فضای کم بعد تبدیل می‌نماید و به این ترتیب می‌توان از آن برای ویژگی‌های الگوی دودویی محور و برای ارائه یک نمونه متراکم کم بعد استفاده نمود. برای نمونه با تلفیق موجک گابور و ویژگی‌های الگوهای محلی دودویی برای تشخیص چهره، تان و تریگز در آغاز بردارهای ویژگی اصلی را در آنالیز اجزای اصلی (PCA)^{۴۲} به صورت فضای خاص لحاظ نموده و در ادامه از بردارهای معمول متمایز هسته‌ای (KDCV)^{۴۳} برای استخراج ویژگی‌های متمایز استفاده کرده‌اند [۲۸].

۴. معایب الگوی دودویی محلی و ارائه راه‌کار

از آنجا که مقدار سطح خاکستری پیکسل مرکزی، مستقیماً به عنوان حد آستانه^{۴۴} بکار گرفته می‌شود، الگوی محلی دودویی در مقابل نویز حساس است، به خصوص قسمت‌های تقریباً غیر یکنواخت تصاویر. از آنچه که در شکل ۸ نشان داده شده است در می‌یابیم که تغییرات کوچک در پیکسل مرکزی (مثلاً از ۷۰ به ۷۵) می‌تواند تاثیر زیادی در کدهای الگوی محلی دودویی داشته باشد.

^{۴۲} . Principal component analysis

^{۴۳} . Kernel discriminative common vector

^{۴۴} . Threshold

^{۴۱} . AdaBoost

مرکزی نیرومندتر عمل می‌کند [۳۶]. سپس در فرایند الگوی دودویی محلی، ALG به عنوان حد آستانه بکار گرفته می‌شود. این روش الگوی دودویی محلی نیرومند نامیده می‌شود. به عنوان یک نتیجه ما الگوی دودویی محلی

نیرومند را به صورت زیر تعریف می‌کنیم:

$$RLBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - ALG_c) 2^p = \quad (9)$$

$$\sum_{p=0}^{P-1} s \left(g_p - \frac{\sum_{i=1}^8 g_{ci} + g_c}{9} \right) 2^p$$

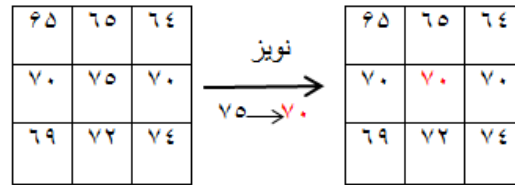
در اینجا g_c مقدار سطح خاکستری پیکسل مرکزی و $g_p (p=0, \dots, 8)$ مقدار سطح خاکستری پیکسل‌های همسایه بر روی یک دایره به شعاع R است و P مجموع تعداد همسایگان می‌باشد. همچنین $g_{ci} (i=0, \dots, 8)$ نمایان گر مقدار سطح خاکستری پیکسل‌های مجاور g_c است. واضح است که الگوی دودویی محلی نیرومند حساسیت کمتری به نویز دارد زیرا مقدار متوسط سطح خاکستری به عنوان حد آستانه بکار گرفته می‌شود. علاوه بر این، دو الگوی متفاوت با کد دودویی محلی سنتی یکسان، ممکن است کد دودویی محلی نیرومند شده متفاوت داشته باشند زیرا همسایگان هر پیکسل مجاور نیز در نظر گرفته می‌شوند. بنابراین الگوی دودویی محلی نیرومند می‌تواند بر دو مشکل مزبور الگوی محلی دودویی سنتی چیره شود.

ALG مقدار خاص پیکسل‌های منحصر به فرد را نادیده می‌گیرد. اغلب اوقات اطلاعات پیکسل‌های مرکزی مورد نیاز است. یک توازن بین عدم نویزپذیری و پیکسل‌های منحصر به فرد وجود دارد. پس ما سطح خاکستری وزن دار (WLG)^{۴۷} محلی را به صورت زیر تعریف می‌کنیم:

$$WLG = \frac{\sum_{i=1}^8 g_i + \alpha g}{8 + \alpha} \quad (10)$$

G و g_i در معادله (۸) تعریف شده‌اند. α پارامتری است که توسط کاربر تنظیم می‌شود. توجه داشته باشید که سطح خاکستری وزن دار معادل ALG متداول است اگر α مجموعه‌ای از یک‌ها باشد. حال الگوی دودویی محلی نیرومند را به صورت زیر تعریف می‌کنیم:

$$RLBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - WLG_c) 2^p = \quad (11)$$



کد دودویی: ۰۰۰۱۱۱۰۱ کد دودویی: ۰۰۰۰۰۰۰۰

شکل ۸: یک مثال از الگوی محلی دودویی حساس به نویز. عیب دیگر الگوی دودویی محلی این است که با ساختارهای الگویی متفاوت از بافت‌ها، کدهای دودویی یکسانی ممکن است نتیجه گردد. همان طور که در شکل ۹ نشان داده شده است دو الگو، کد دودویی محلی یکسانی دارند با اینکه سخت است بگوییم این دو ساختار محلی به هم شبیه‌اند.

۴	۹	۹	۱۰۰	۱۸۰	۱۸۰
۸	۸	۸	۲۴۰	۱۸۰	۲۴۰
۰	۶	۴	۱۷۹	۵۷	۳۰

دودویی: ۰۱۱۱۰۰۰۱ دودویی: ۰۱۱۱۰۰۰۱

شکل ۹. یک مثال از دو الگوی ساختاری متفاوت با کد

دودویی یکسان

به منظور حل مشکلات فوق‌الذکر در این بخش یک چارچوب قوی برای الگوی محلی دودویی ارائه می‌کنیم که هم مزایای الگوهای محلی دیگر را به ارث می‌برد و هم می‌تواند بر ایرادات آنها چیره شود.

۴-۱- الگوی دودویی محلی مقاوم (RLBP)^{۴۵}

هدف، یافتن یک آستانه است که نسبت به نویز حساسیت نداشته باشد و در مقابل تبدیلات سطح خاکستری یکنواخت، ثابت باشد. ما یک سطح خاکستری متوسط (ALG)^{۴۶} را به صورت زیر تعریف می‌کنیم:

$$ALG = \frac{\sum_{i=1}^8 g_i + g}{9} \quad (8)$$

در اینجا g مقدار خاکستری پیکسل مرکزی را بیان می‌کند و $g_i (i=0, \dots, 8)$ مقدار سطح خاکستری پیکسل‌های همسایه هست. ALG متوسط سطح خاکستری بافت محلی را ارائه می‌دهد که قطعا در مقابل نویز نسبت به خود پیکسل

^{۴۵} . Robust LBP

^{۴۶} . Average Local Gray Level

^{۴۷} . Weighted Local Gray Level

سطح خاکستری وزن‌دار است. دامنه الگوی دودویی محلی نیرومند (RLBP_M)^{۴۹} با استفاده از واریانس محلی سطح خاکستری وزن‌دار به صورت زیر تعریف می‌شود:

$$RLBP_{MP,R} = \sum_{P=0}^{P-1} s(m_P - c) 2^P \quad (۱۳)$$

در اینجا میانه مقادیر m_P به عنوان حد آستانه c تنظیم می‌گردد. پیکسل مرکزی که بیان‌کننده سطح خاکستری مرکزی تصویر است، شامل اطلاعات ایجاد تمایز است. بنابراین ما می‌توانیم یک اپراتور بنام الگوی دودویی محلی نیرومند مرکزی (RLBP_C)^{۵۰} به صورت زیر تعریف کنیم تا اطلاعات مرکزی محلی را استخراج کنیم:

$$RLBP_{CP,R} = s(WLG_C - C_i) \quad (۱۴)$$

در اینجا حد آستانه C_i به عنوان متوسط سطح خاکستری محلی در هر قسمت از کل تصویر تنظیم می‌شود. همان طور که در [۱۳] نشان داده شده است، از راه‌های مختلفی برای ترکیب سه اپراتور الگوی دودویی محلی مقاوم، دامنه الگوی دودویی محلی نیرومند و الگوی دودویی محلی نیرومند مرکزی استفاده کرد.

اقدامات زیادی برای تمایز قائل شدن و عدم تجانس بین دو هیستوگرام ارائه شده است. در این مقاله ما از آماره X^2 برای این کار استفاده کرده‌ایم. اگر برای $(i=1, \dots, B)$ ، $H=\{h_i\}$ و $K=\{k_i\}$ نمایانگر دو هیستوگرام باشند پس آماره X^2 به صورت زیر تعریف می‌شود:

$$L(S, M) = \sum_{b=1}^B (S_b * \log(M_b)) \quad (۱۶)$$

در اینجا b یک عدد bins و B مجموع تمامی bins است، که با توجه به استفاده از هیستوگرام در زبان برنامه‌نویسی مطلب و روش این مقاله، $B=256$ و b هم بین صفر تا ۲۵۵ هست. در معادله (۱۶)، L یک شبه روش است که احتمالات نمونه S را که از طریق کلاس‌های بافتی تناوبی جایگزین بدست ی‌د، اندازه‌گیری می‌کند که بر پایه احتمالات درستی از ارزش‌های خصیصه با مدل‌های بافتی از قبل رده‌بندی شده M می‌باشد. در مورد توزیع اتصالات LBP/VAR، معادله (۱۶) در یک روش درست با ردگیری و جستجو از

$$\sum_{P=0}^{P-1} s \left(g_P - \frac{\sum_{i=1}^8 g_{ci} + \alpha g_c}{8 + \alpha} \right) 2^P$$

g_P ، g_c و g_{ci} در معادله (۹) تعریف شده‌اند. α پارامتر سطح خاکستری وزن‌دار است. به عنوان مجموعه‌ای از آزمایشاتی که بر روی پایگاه داده بافتی انتخابی بزرگ انجام شد پارامتر α بهینه در این مقاله موجود است. تصاویر مجموعه داده‌ای CURET دچار تغییرات قابل توجه در شدت نور، تنوع در نقطه دید و تغییرات مقیاسی هستند. بنابراین آزمایشات را بر روی مجموعه داده‌ای CURET و مجموعه داده‌ای CURET نویزی (SNR=5)^{۴۸} انجام شده است. توابع پاسخ برای α های متفاوت بر روی مجموعه داده‌ای CURET و مجموعه داده‌ای CURET نویزی نشان می‌دهد که الگوی دودویی محلی نیرومند با α های ۹،۸ و ۱۰ روی پایگاه داده نرمال بهتر عمل می‌کند. همچنین زمانی $\alpha=1$ است و نویز وجود ندارد الگوی دودویی محلی نیرومند بهتر عمل می‌کند و همچنین زمانی که α بزرگتر از هشت است الگوی دودویی محلی نیرومند حساسیت بیشتری نسبت به نویز دارد. بنابراین در روش ما، با توجه به نتایج تجربی α یک یا هشت تنظیم می‌گردد. عبارت دیگر، الگوی دودویی محلی نیرومند با $\alpha=1$ در مقابل الگوی دودویی محلی نیرومند با $\alpha=8$ حساسیت کمتری نسبت به نویز دارد در صورتیکه الگوی دودویی محلی مقاوم با $\alpha=8$ در مقابل نورهای مختلط و بغرنج و همچنین در مقابل حالات مختلف نقطه دید پایدارتر است زیرا اطلاعات سطح خاکستری را هم از مجموعه همسایگان محلی و هم از پیکسل‌های منحصر به فرد استخراج می‌کند.

۴-۲- الگوی دودویی محلی تکمیل‌شده مقاوم

به خاطر الگوهای محلی گیج‌کننده مختلف دودویی، الگوی دودویی محلی نیرومند، چارچوبی از الگوی دودویی محلی را به ارث برده است. مقدار m_P به صورت زیر تعریف می‌گردد:

$$m_P = |WLG_P - WLG_C| = \left| \frac{\sum_{i=1}^8 g_{Pi} + \alpha g_P}{8 + \alpha} - \frac{\sum_{i=1}^8 g_{ci} + \alpha g_c}{8 + \alpha} \right| \quad (۱۲)$$

g_{Pi} ($i=0, \dots, 8$) در معادله (۹) تعریف شده‌اند. نشان‌دهنده شدت خاکستری پیکسل‌های مجاور g_P و α پارامتر

^{۴۹} . Robust LBP Magnitude

^{۵۰} . Robust LBP-Center

^{۴۸} . Signal-to-Noise Ratios

داده‌های تست استفاده می‌کنیم. متوسط میزان طبقه‌بندی برای بیش از صد جز تصادفی، در جدول ۲ لیست شده است:
جدول ۲: میزان طبقه‌بندی در پایگاه داده CURET بر

حساب درصد

شعاع و تعداد مجاور	R=۱,P=۸		R=۲,P=۱۶		R=۳,P=۲۴	
	۲۳	۶	۲۳	۶	۲۳	۶
تعداد نمونه مدل	۲۳	۶	۲۳	۶	۲۳	۶
LBP	۷۴,۴۰	۵۸,۰۵	۷۰,۳۹	۵۶,۳۰	۷۳,۷۲	۵۹,۸۱
CLBP_S/M/C	۹۱,۳۰	۷۳,۴۴	۹۱,۳۴	۷۵,۴۱	۹۱,۷۷	۷۶,۱۶
RLBP($\alpha = 8$)	۶۸,۷۵	۵۴,۰۹	۷۰,۳۹	۵۶,۳۰	۷۳,۷۲	۵۹,۸۱
CRLBP($\alpha = 8$)	۹۱,۳۳	۷۶,۵۶	۹۱,۸۵	۷۷,۷۹	۹۱,۸۳	۷۸,۴۳
RLBP($\alpha = 1$)	۶۶,۳۴	۵۳,۴۴	۶۶,۳۵	۵۳,۱۳	۶۸,۱۸	۵۴,۵۰
CRLBP($\alpha = 1$)	۸۹,۴۷	۷۳,۲۲	۹۱,۱۰	۷۶,۳۹	۹۰,۷۳	۷۶,۳۴

با توجه به جدول ۲، اولاً درمی‌یابیم که الگوی دودویی محلی تکمیل شده مقاوم با $\alpha=8$ بهترین طبقه‌بندی را نسبت به سایر روش‌ها دارد، به خصوص زمانی که داده‌های مدل کمتر باشند. دوم اینکه، وقتی تعداد داده‌های مدل به اندازه کافی باشد، الگوی دودویی محلی تکمیل شده مقاوم با $\alpha=1$ اجرای بدتری نسبت به الگوی دودویی محلی تکمیل شده دارد. زمانی که داده‌های مدل کمتر از ۲۳ عدد و شعاع بزرگتر از یک باشد الگوی دودویی محلی تکمیل شده مقاوم با $\alpha=1$ کمی بهتر از الگوی دودویی محلی تکمیل شده عمل می‌کند. بنابراین الگوی دودویی محلی تکمیل شده مقاوم با $\alpha=8$ پایدارتر از الگوی دودویی محلی تکمیل شده نیرومند با $\alpha=1$ است زیرا هم اطلاعات سطح خاکستری و هم اطلاعات خاص پیکسل‌های منحصر به فرد را برمی‌گرداند. در جدول ۳ الگوی ارائه شده را با الگوی دودویی محلی واریانس^{۵۲} [۱۵]، الگوی دودویی محلی انطباقی^{۵۳} [۳۸]، الگوی دودویی محلی ادغامی با داده‌کاوی^{۵۴} [۳۹] و الگوی دودویی محلی مونوژنیک^{۵۵} [۴۰] نیز مقایسه کرده‌ایم که نتایج مشابهی را ارائه می‌کند.

میان هیستوگرام‌های دوبعدی توسعه یافت. حال معادله (۱۷) را ببینید. هرچه خروجی این معادله کمتر باشد، احتمال اینکه این دو تصویر نمونه و مرجع هم کلاس باشند بیشتر است.

$$G(S, M) = 2 \sum_{b=1}^B \left(S_b * \log \left(\frac{S_b}{M_b} \right) \right) \quad (17)$$

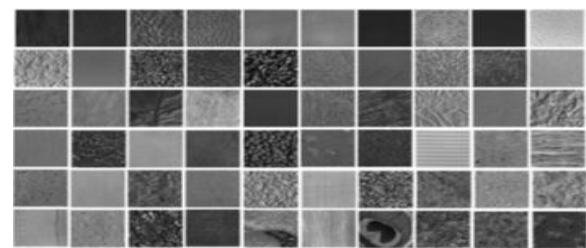
$$2 \sum_{b=1}^B (S_b * \log(S_b) - S_b * \log(M_b))$$

۵. ارزیابی نتایج تجربی

برای ارزیابی روش ارائه شده، ما یک سری از آزمایشات را روی پایگاه داده بافتی CURET [۳۷] و Outex انجام داده‌ایم تا دقت طبقه‌بندی الگوی دودویی محلی تکمیل شده مقاوم بررسی گردد. برای هدایت آزمایشات روی تصاویر نویزی، هر تصویر بافتی با افزودن نویز گوسین^{۵۱} با میانگین و انحراف معیار صفر خراب می‌گردد که با توجه به نسبت سیگنال به نویز تعیین می‌گردد. همچنین قسمت‌بندی تصاویر آکوستیکی بستر دریا به کمک روش ارائه شده و مقایسه نتایج با روش‌های قبل، راه کار مقاله را مورد ارزیابی قرار می‌دهد.

۱-۵- مشاهده نتایج تجربی روی پایگاه داده CURET

پایگاه داده CURET شامل ۶۰ تصویر بافتی با نقطه دیدها و جهات نور متفاوت می‌باشد (شکل ۱۰ را ببینید). در هر کلاس، ۹۲ تصویر از تصاویر عکسی با زوایای دید کمتر از ۶۰ درجه انتخاب می‌شود.



شکل ۱۰. تصاویر بافتی از مجموعه داده‌ای CURET

برای مجموعه مدل، N تصویر به صورت تصادفی از هر کلاس انتخاب می‌کنیم و سپس از $(92-N)$ تصویر باقیمانده بعنوان

^{۵۲}. Log-likelihood ratio

^{۵۳}. Gaussian noise

^{۵۴}. LBP-Variance

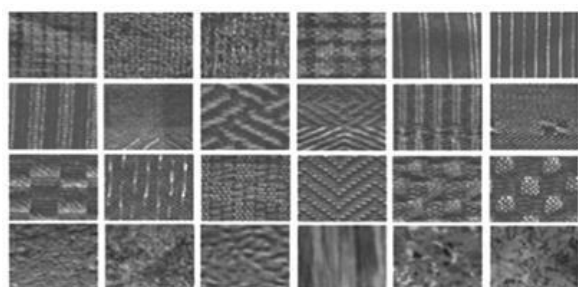
^{۵۵}. Adaptive-LBP

⁵¹. Chi-square statistic

جدول ۳. مقایسه میزان طبقه‌بندی در پایگاه داده

CURET با کارهای دیگر

شعاع و تعداد مجاور	R=۱,P=۸		R=۲,P=۱۶		R=۳,P=۲۴	
	۴۶	۲۳	۴۶	۲۳	۴۶	۲۳
تعداد نمونه مدل	۴۶	۲۳	۴۶	۲۳	۴۶	۲۳
LBP/Var	۹۳,۶۵	۷۰,۷۰	۹۳,۹۰	۷۰,۸۲	۹۳,۹۰	۷۰,۹۱
LBPV	۸۸,۲۳	۷۱,۵۶	۸۹,۷۷	۷۳,۱۰	۹۱,۰۹	۷۴,۲۶
LBPV_GM	۹۳,۱۹	۷۵,۲۶	۹۴,۱۵	۷۹,۶۶	۹۲,۹۷	۷۶,۶۹
ALBP	-	۵۷,۴۰	-	۶۷,۳۰	-	۶۵,۵۰
ALBPF	-	۶۶,۱۰	-	۷۲,۷۰	-	۷۱,۷۰
LBP/LDDP ²	۷۲,۸۵		۷۰,۷۴		۶۶,۴۰	
LBP/LDDP ³	۶۷,۹۱		۶۸,۶۴		۶۷,۴۱	
M_LBP	۸۸,۹۶ (نمونه ۲۳) و ۹۳,۷۶ (۴۶ نمونه)					
CRLBP($\alpha = 8$)	۹۵,۳۹	۹۱,۳۳	۹۵,۸۸	۹۱,۸۵	۹۶,۲۷	۹۱,۸۳
CRLBP($\alpha = 1$)	۹۴,۵۵	۸۹,۴۷	۹۴,۷۸	۹۱,۱۰	۹۵,۳۵	۹۰,۷۳



شکل ۱۱. تصاویر بافتی از مجموعه داده‌ای OUTEX

برای آزمایشات انجام شده بر روی پایگاه داده Outex، کارایی روش ما با روش‌های متعددی مانند الگوی دودویی محلی واریانس [۱۵]، الگوی دودویی محلی ادغامی با داده کاوی [۳۹]، الگوی دودویی محلی وزن‌دار [۴۱] و الگوی دودویی محلی غالب برجسب‌دار و بدون برجسب [۴۲] مقایسه شده است. جدول ۵ لیستی از نتایج آزمایشات با روش‌های مختلف است که ما می‌توانیم به یافته‌های زیر دست یابیم. به استثنای آزمایشات با $R=2$ و $P=16$ روی مجموعه داده‌ای TC10، الگوی دودویی محلی تکمیل شده مقاوم، اجرای بهتری دارد. باید توجه داشت که الگوی دودویی محلی تکمیل شده مقاوم با $\alpha=8$ نتایج بهتری تحت روشنایی‌های tl84 و horizon را نسبت به سایر روش‌ها می‌دهد طوری که ثابت کرده است که الگوی دودویی محلی تکمیل شده نیرومند با $\alpha=8$ در مقابل تغییرات روشنایی از بقیه قوی‌تر است. همچنین نتایج مشابه‌ای برای الگوی دودویی محلی تکمیل شده نیرومند با $\alpha=1$ حاصل شده است. الگوی دودویی محلی تکمیل شده

برای هدایت تصاویر بافتی نویزی مجموعه داده‌ای CURET، از هر کلاس به طور تصادفی ۴۶ نمونه مدل انتخاب می‌کنیم. این بخش نیز بیش از صد بار به طور مستقل با $(R=24$ و $P=3)$ اجرا شده است. متوسط نتایج تصاویر نویزی در جدول ۴ ارائه شده است:

جدول ۴: دقت طبقه‌بندی در پایگاه داده CURET با

افزودن نویز گوسین

	SNR 100	SNR 30	SNR 15	SNR 10	SNR 5
CLBP	۹۵,۵۱	۹۵,۸۷	۸۷,۲۳	۷۲,۷۷	۵۱,۳۵
CRLBP($\alpha = 8$)	۹۶,۳۴	۹۶,۱۸	۹۲,۳۰	۸۲,۸۸	۶۴,۹۷
CRLBP($\alpha = 1$)	۹۶,۰۶	۹۵,۹۰	۹۳,۵۶	۸۵,۵۸	۶۹,۶۷

با توجه به تصاویر نویزی، به یافته‌های زیر از جدول ۴ دست می‌یابیم. اول اینکه، الگوی دودویی محلی تکمیل شده نیرومند همچنان با $\alpha=8$ میزان دقت بیشتری نسبت الگوی دودویی محلی تکمیل شده عمل می‌کند. دوم اینکه، الگوی دودویی محلی تکمیل شده نیرومند با $\alpha=1$ نتایج ضعیف‌تری نسبت به الگوی دودویی محلی تکمیل شده برای تصاویر نرمال دارد اما نتایج بهتری را در پایگاه داده‌های نویزی دارد.

۵-۲- مشاهده نتایج تجربی روی پایگاه داده Outex

وقتی مجموعه آزمایشات بر روی مجموعه داده‌ای Outex هدایت می‌شود (شکل ۱۱) ما از داده‌های آزمایشی مناسب Outex (TC10) Outex_TC_0010 و Outex (TC12) Outex_TC_0012 استفاده می‌کنیم که TC10 و TC12 شامل ۲۴ کلاس از تصاویر بافتی با سه

۵۶. Data Mining

۵۷. Monogenic-LBP

موجکی و فازی در کنار نتایج راه کار ارائه شده برای یکی تصاویر بستر دریا مشاهده می‌شود:



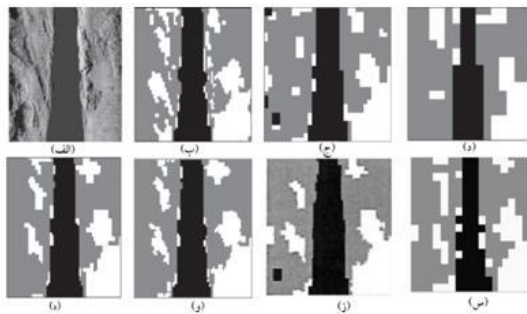
الف ب ج

شکل ۱۳: نتایج حاصل از قسمت‌بندی. (الف) تصویر اصلی.

(ب) قسمت‌بندی حاصل از تبدیلات زیرنمونه‌ای [۷].

(ج) قسمت‌بندی با راهکار ارائه شده.

در شکل ۱۴، نتایج حاصل از تبدیلات موجکی یک و دو سطحی، آشکار ساز لبه فازی و ادغامی از آنها، همچنین نتایج تبدیلات کانتورلتی زیرنمونه‌ای، در کنار قسمت‌بندی با راهکار ارائه شده، برای یکی دیگر از تصاویر بستر دریا نمایش داده شده است. از آنجا که رزولوشن این تصاویر ۲۵۶ در ۲۵۶ می‌باشد، اندازه هر بلوک ۱۶ در ۱۶ در نظر گرفته شده است که با توجه به آزمایشات صورت گرفته مناسب‌ترین سایز برای بلوک‌ها در راهکار ارائه شده می‌باشد.



شکل ۱۴: نتایج حاصل از قسمت‌بندی. (الف) تصویر اصلی.

(ب) تقسیم‌بندی با آشکار ساز لبه فازی. (ج) و (د)

تقسیم‌بندی سطح یک و دو موجکی بترتیب. (ه) ادغام

(ب) و (ج) و (د) و (و) پالایش مرزهای (ه) [۳]. (ز) قسمت‌بندی

حاصل از تبدیلات کانتورلتی زیرنمونه‌ای [۷].

(ج) قسمت‌بندی با راه کار ارائه شده.

۶. نتیجه‌گیری

با توجه به ظاهر با کیفیت عکس‌های بافتی سونار، فنون تجزیه و تحلیل بافت یک انتخاب متداول برای عکس‌های آکوستیکی کف دریاست. الگوی دودویی محلی یک توصیف گر قدرتمند بافت برای تحلیل تصاویر می‌باشد. در این مقاله پیشرفت‌ها و توسعه‌های جدیدی که در الگوی

نیرومند با $\alpha=1$ کارایی بهتری نسبت به الگوهای دودویی محلی روی مقدار متوسط به خصوص تحت روشنایی‌های متفاوت دارد. در پایان، با مقایسه الگوی دودویی محلی تکمیل شده نیرومند با $\alpha=8$ با الگوی دودویی محلی تکمیل شده نیرومند با $\alpha=1$ خواهیم دریافت که زمانی که شعاع برابر یک است روش اولی کارایی بهتری دارد و دومی بیشترین میزان طبقه‌بندی در سایر شعاع‌ها دارد. در یک کلمه، الگوی دودویی محلی تکمیل شده نیرومند بیشترین میزان طبقه‌بندی را نسبت به سایر روش‌ها دارد و در مقابل تغییرات روشنایی حساسیت کمتری را دارد.

جدول ۵: دقت طبقه‌بندی در پایگاه داده Outex بر حسب

درصد

شعاع و تعداد مجاوران نمونه‌های آزمایشی الگو	R=1,P=8				R=2,P=16				R=3,P=24			
	TC10		TC12		TC10		TC12		TC10		TC12	
	T184	Hr.	T184	Hr.	T184	Hr.	T184	Hr.	T184	Hr.		
LBPV_GM _{ES}	۷۳٫۶۴	۷۲٫۴۷	۷۶٫۵۷	۹۳٫۹۰	۹۰٫۲۵	۹۴٫۲۸	۹۷٫۷۶	۹۵٫۳۹	۹۵٫۵۷			
LBPV_GM _{PD1}	۶۵٫۵۹	۶۵٫۳۰	۶۸٫۳۳	۸۹٫۶۰	۸۶٫۳۹	۹۰٫۱۱	۹۵٫۰۷	۸۸٫۷۲	۸۹٫۰۲			
LBPV_GM _{PD2}	۷۲٫۹۹	۷۲٫۱۹	۷۶٫۱۵	۹۲٫۹۹	۸۹٫۴۹	۹۳٫۹۵	۹۷٫۵۵	۹۴٫۲۳	۹۴٫۱۸			
LBP/LDDP ²	۸۸٫۱۶	۷۷٫۵۳	۷۸٫۲۵	۹۷٫۲۵	۹۱٫۹۵	۸۸٫۵۳	۹۷٫۴۹	۹۱٫۶۲	۸۷٫۹۲			
LBP/LDDP ³	۹۰٫۶۵	۸۰٫۲۶	۷۸٫۹۹	۹۳٫۳۶	۸۴٫۸۸	۸۱٫۱۳	۹۶٫۶۱	۸۹٫۸۳	۸۵٫۸۷			
WLBP	۹۱٫۸۲	۸۳٫۴۰	۸۱٫۸۵	۹۶٫۴۳	۸۹٫۳۹	۸۹٫۲۸	۹۷٫۳۹	۹۳٫۷۲	۹۲٫۰۱			
Labeled DLBP	۷۶٫۷۰											
unLabeled DLBP	۶۸٫۲۵											
CRLBP($\alpha=8$)	۹۷٫۵۵	۹۱٫۹۴	۹۲٫۴۵	۹۸٫۵۹	۹۵٫۸۸	۹۶٫۴۱	۹۹٫۳۵	۹۶٫۸۳	۹۶٫۱۶			
CRLBP($\alpha=1$)	۹۶٫۵۴	۹۱٫۱۶	۹۲٫۰۶	۹۸٫۸۵	۹۶٫۶۷	۹۶٫۹۷	۹۹٫۴۸	۹۷٫۵۷	۹۷٫۳۴			

۳-۵- مشاهده نتایج تجربی تقسیم‌بندی بستر دریا

در این زیربخش برای ارزیابی راه کار ارائه شده، نتایج حاصل از قسمت‌بندی تصاویر آکوستیکی بستر دریا، با نتایج حاصل از تبدیلات زیرنمونه‌ای کانتورلتی [۷] و تبدیلات موجکی چند سطحی و آشکار ساز لبه فازی [۳] مقایسه می‌شود. ابعاد تصاویر انتخابی ۲۵۶ در ۲۵۶ می‌باشد. ابتدا تصاویر بلوک‌بندی می‌گردند و سپس هم الگوی دودویی محلی سنتی و هم الگوی دودویی محلی تکمیل شده مقاوم به تصاویر اعمال شده و در نهایت تمامی بلوک‌ها ادغام می‌شوند که شکل اصلی و نتایج حاصل از آن، در شکل ۱۲ نمایش داده شده است:

همان طور که در شکل بالا می‌بینید الگوی دودویی محلی تکمیل شده مقاوم، بسیار قوی‌تر از الگوی دودویی محلی عمل می‌کند. در شکل ۱۳، نتایج حاصل از ادغام تبدیلات

کانتورلتی زیر نمونه‌ای، نشان داده شده است که توانایی بهبود تقسیم‌بندی‌ها از نقطه نظر سرعت اجرا، پیچیدگی محاسبات و دقت، نتایج بهتری نسبت به سایر روش‌های شناخته شده دارد. مزیت اصلی روش ارائه شده سادگی روش و سرعت کافی برای تعبیه در روش‌های زمان واقعی تشخیص بستر دریاست.

به عنوان یک راهکار برای آینده، می‌توان توسعه جدیدی از الگوی دودویی محلی را ارائه کرد طوری که منجر به توسعه اهداف کاربردی، بهبود قدرت در یک یا چند جنبه الگوی دودویی محلی و افزایش قابلیت تمایز در الگوی دودویی محلی به خصوص در تصاویر دارای نویز گردد. همچنین برای طبقه‌بندی و قسمت‌بندی، می‌توان از دیگر طبقه‌بندها استفاده کرد و به شیوه‌های مختلف آن طبقه‌بندها را با انواع الگوهای دودویی محلی ادغام کرد.

منابع

1. M. Mignotte and C. Collet, "Markov Random Field and Fuzzy Logic Modeling in Sonar Imagery: Application to the classification of underwater floor," *Computer Vision and Image Understanding*, Vol. 79, No. 1, pp. 4-24, 2000.
2. R. Javidan and H. J. Eghbali, "Automatic Classification of Persian Gulf Bottom Based on Acoustic Images," *6th International Conference on Coasts, Ports and Marine Structures*, 2004.
3. R. Javidan and H. J. Eghbali, "Automatic seabed texture segmentation and classification based on wavelet transform and fuzzy approach," *International Journal of the Society for Underwater Technology*, Vol. 27, No. 2, pp. 51-55, 2007.
4. A. Dejebari and F. Bereksi-Reguig, "A new chirp based wavelet for heart sounds time-frequency analysis,"

دودویی دریافتیم که الگوی سنتی صرفاً از الگوهای یکنواخت استفاده می‌کند و تمام الگوهای غیر یکنواخت را در یک طبقه ادغام می‌کند و نسبت به نویز حساس است. توصیف‌گرهای الگوی دودویی محلی، دارای ناکارآمدی واضح دیگری هم می‌باشند. اینکه گاهی اوقات منجر به توصیف الگوهای مختلف ساختاری با کد دودویی همگون می‌شود که به ناچار قابلیت تمایز خود را کاهش می‌دهد. برای غلبه بر این ناکارآمدی‌ها، یک چارچوب قوی از الگوی دودویی محلی بنام الگوی دودویی محلی کامل شده مقاوم در این مقاله ارائه شد.

به عنوان یک نتیجه، پس از اجرای پردازش، با تکیه بر رویکردهای اعمال شده به تصویر اصلی و ادغام الگوی دودویی محلی، آماره مربع کای و نسبت احتمالات لگاریتمی و همچنین مقایسه نتایج با تبدیلات چند سطحی موجکی، تقسیمات اصلاح شده، آشکار ساز لبه فازی و تبدیلات

international journal on communication antenna and propagation (IRECAP), Vol. 1, N. 1, pp. 92-102, 2011.

5. K. K. Gupta and R. Gupta, "wavelet based speckle filtering of the sar image," *international review on computers and software (IRECOS)*, Vol. 1, No. 3, pp. 224-232, 2006.

6. R. Javidan, "wavelet-based acoustic seabed ground discrimination system," *In proceeding of IEEE international conference on information & communication technologies: from theory to application*, Paris, July 7-10 2010.

7. R. Javidan, M. A. Masnadi-Shirazi and Z. Azimifar, "Contourlet-Based Seabed Image Texture Segmentation and Classification," *International Journal on Communications Antenna and Propagation (IRECAP)*, Vol. 1, No. 4, August 2011.

8. T. Ojala, M. Pietikainen and T.

- Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transaction on Pattern Anal. Mach. Intell.*, Vol. 24, No. 7, pp. 971–987, 2002.
9. A. Fathi and A. R. Naghsh-Nilchi, "Noise tolerant local binary pattern operator for efficient texture analysis," *Pattern Recognition Letters*, Vol. 33, No. 9, pp. 1093-1100, 2012.
10. M. Pietikainen, T. Ojala, and Z. Xu, "Rotation-invariant texture classification using feature distributions," *Pattern Recog.*, Vol. 33, No.1, pp. 43–52, 2000.
11. T. Ahonen, A. Hadid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transaction on Pattern Anal. Mach. Intell.*, Vol. 28, No. 12, pp. 2037–2041, 2006.
12. G. Zhao, T. Ahonen, J. Matas and M. Pietikainen, "Rotation invariant image and video description with local binary pattern features," *IEEE Transactions on Image Processing*, Vol. 21, No. 4, pp. 1465-1477, 2010.
13. Z. Guo, L. Zhang and D. Zhang, "A Completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, Vol. 19, No. 6, pp. 1657-1663, 2010.
14. J. Ruiz-del-Solar and J. Quinteros, "Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different pre-processing approaches," *Pattern Recog. Lett.*, Vol. 29, No. 14, pp. 1966–1979, 2008.
15. Z. Guo, L. Zhang and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognition*, Vol. 43, No. 3, pp. 706-719, 2010.
16. D. Huang, Y. Wang and Y. Wang, "A robust method for near infrared face recognition based on extended local binary pattern," in *Proc. Int. Symp. Vis. Comput. Part2 LNCS 4842*, pp. 437–446, 2007.
17. Z. Guo, L. Zhang and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transaction on Image Process.*, Vol. 19, No. 6, pp. 1657–1663, 2010.
18. X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proc. Anal. Model. Faces Gestures, LNCS 4778*, pp. 168–182, 2007.
19. T. Ahonen and M. Pietikainen, "Soft histograms for local binary patterns," in *Proc. Fin. Signal Process. Symp.*, Oulu, Finland, 2007.
20. S. Liao, W. Max, K. Law, C. Albert and S. Chung, "Dominant local binary patterns for texture classification," *IEEE Transactions on Image Processing*, Vol. 18, No. 5, pp. 1107-1118, 2009.
21. H. Zhou, R. Wang and C. Wang, "A novel extended local binary pattern operator for texture analysis," *Information Sciences*, Vol. 178, No. 22, pp. 4314-4325, 2008.
21. L. Wolf, T. Hassner and Y. Taigman, "Descriptor based methods in

- the wild,” in *Proc. ECCV Workshop Faces ‘Real-Life’ Images: Detection, Alignment, Recog.*, Marseille, France, 2008.
22. L. Nanni, S. Brahmam and A. Lumini, “A simple method for improving local binary patterns by considering non-uniform patterns,” *Pattern Recognition*, Vol. 45, No. 10, pp. 3844-3852, 2012.
23. J. Fehr, “Rotational invariant uniform local binary patterns for full 3D volume texture analysis,” in *Proc. Fin. Signal Process. Symp.*, Oulu, Finland, 2007.
24. J. Fehr, “Rotational invariant uniform local binary patterns for full 3D volume texture analysis,” in *Proc. Fin. Signal Process. Symp.*, Oulu, Finland, 2007.
25. G. Zhao and M. Pietikainen, “Dynamic texture recognition using local binary patterns with an application to facial expressions,” *IEEE Transaction on Pattern Anal. Mach. Intell.*, Vol. 29, No. 6, pp. 915–928, 2007.
26. Y. K. Park and J. K. Kim, “Fast adaptive smoothing based on LBP for robust face recognition,” *Electron. Lett.*, Vol. 43, No. 24, pp. 1350–1351, Nov. 2007.
27. R. Singh, M. Vatsa and A. Noore, “Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition,” *Pattern Recog.*, Vol. 41, No. 3, pp. 880–893, 2008.
28. X. Tan and B. Triggs, “Fusing Gabor and LBP feature sets for kernelba face recognition,” in *Proc. Anal. Model. Faces Gestures, LNCS 4778*, pp. 235–249, 2007.
29. L. Liu, L. Zhao, Y. Long, G. Kuang and P. Fieguth, “Extended local binary patterns for texture classification,” *Image and Vision Computing*, Vol. 30, No. 2, pp. 86-99, 2012.
30. H. Lian and B. Lu, “Multi-view gender classification using local binary patterns and support vector machines,” in *Proc. Int. Symp. Neural Netw., LNCS 3972*, pp. 202–209, 2006.
31. M. Heikkila, M. Pietikainen and C. Schmid, “Description of interest regions with local binary patterns,” *Pattern Recog.*, Vol. 42, No. 3, pp. 425–436, 2009.
32. H. Ren and Z. N. Li, “Object Deection Using Boosted Local Binaries,” *Pattern Recognition*, 5 July 2016.
33. Y. Zhang and C. Hua, “Driver fatigue recognition based on facial expression analysis using local binary patterns,” *Optik International Journal for Light and Electron Optics*, Vol. 126, No. 23, PP. 4501–4505, 2015.
34. D. S. Kang, “Robust Face Detection Using Uniform Local Gradient Pattern (ULGP) and AdaBoost Algorithm ,” *4th International Conference on Modeling and Simulation (MAS)*, pp. 13–16, 25-28 Nov. 2015.
35. S. K. Singh and R. Kumar Singh, “Boosting local binary pattern with bag-of-filters for content based image retrieval,”

- retrieval,” *IEEE UP Section Conference on Electrical Computer and Electronics (UPCON)*, pp. 1–6, 4-6 Dec. 2015.
36. Y. Zhao, W. Jia, R. X. Hu and H. Min, “Completed robust local binary pattern for texture classification,” *Neurocomputing*, Vol. 106, pp. 68–76, 2013.
37. K. J. Dana, B. V. Ginneken, S. K. Nayar, J. J. Koenderink, “Reflectance and texture of real- world surfaces,” *ACM Trans. Graphics(TOG)*, Vol. 18, No. 1, pp. 1-34, 1999.
38. Z. Gu, L. Zhang, D. Zhang and S. Zhang, “Rotation Invariant Texture Classification Using Adaptive LBP With Directional Statistical Features,” *IEEE 17th International Conference on Image Processing (ICIP)*, pp. 285-288, September 26-29, 2010, Hong Kong.
39. J. T. Kwak, S. Xu and B. J. Wood, “Efficient data mining for local binary pattern in texture image analysis,” *Expert Systems with Applications*, Vol. 42, pp. 4529-4539, 2015.
40. L. Zhang, L. Zhang, Z. Guo and D. Zhang, “Monogenic-LBP: A New Approach For Rotation Invariant Texture Classification,” *IEEE 17th International Conference on Image Processing (ICIP)*, pp. 2677-2680, September 26-29, 2010, Hong Kong
41. R. Davarzani, S. Mozaffarin and Kh. Yaghmaie, “Scale-and rotation-invariant texture description with improved local binary pattern features,” *Signal Processing*, Vol. 111, pp. 274-293, 2015.
42. F. Bianconi, E. Gonzalez and A. Fernandez, “Dominant local binary patterns for texture classification: Labelled or unlabelled?,” *Pattern Recognition Letters*, 9 July 2015.

ارائه روشی مناسب برای دسته‌بندی نامه‌های الکترونیکی تبلیغاتی بر مبنای پروفایل کاربران

* محمد فتحیان
** رحیم حضرتقلی‌زاده

* استاد، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران
** کارشناسی ارشد، مهندسی فناوری اطلاعات، دانشگاه علم و صنعت ایران
تاریخ دریافت: ۹۲/۰۹/۲۵ تاریخ پذیرش: ۹۵/۰۳/۱۸

چکیده

به طور کلی، تعریف هرزنامه در ارتباط با رضایت یا عدم رضایت گیرنده است نه محتوای نامه الکترونیکی. بر طبق این تعریف، مشکلاتی در دسته‌بندی نامه‌های الکترونیکی در بازاریابی و تبلیغات مطرح می‌شود. برای مثال امکان دارد بعضی از نامه‌های الکترونیکی تبلیغاتی، برای عده‌ای از کاربران هرزنامه و برای عده‌ای دیگر هرزنامه نباشد. برای مقابله با این مشکل با توجه به پروفایل و رفتار کاربران، ضد هرزنامه‌های شخصی طراحی می‌شود. به طور عادی برای دسته‌بندی هرزنامه‌ها، روش‌های یادگیری ماشینی با دقت خوب به کار می‌رود. اما در هر حال یک روش منحصر به فرد موفق بر مبنای دیدگاه تجارت الکترونیک وجود ندارد. در این مقاله ابتدا پروفایل جدیدی برای شبیه‌سازی بهتر رفتار کاربران، تهیه می‌شود. سپس این پروفایل همراه با نامه‌های الکترونیکی به دانشجویان ارائه شده و پاسخ آنها جمع‌آوری می‌گردد. در ادامه برای دسته‌بندی نامه‌های الکترونیکی، روش‌های مشهور به ازای مجموعه داده‌های مختلف آزمایش می‌شود. سرانجام، با مقایسه معیارهای ارزیابی داده کاوی، شبکه عصبی به عنوان بهترین روش با دقت بالا، تعیین می‌گردد.

واژه‌های کلیدی: تجارت الکترونیکی، تبلیغات الکترونیکی، دسته‌بندی هرزنامه‌ها، داده کاوی، پروفایل

۱. مقدمه

الکترونیک^۲ مخصوصاً بازاریابی و تبلیغات اینترنتی است. این نوع بازاریابی و تبلیغات با عنوان "بازاریابی از طریق نامه‌های الکترونیکی"^۳ نیز مشهور است. همزمان با رشد استفاده از نامه‌الکترونیکی سوءاستفاده و فریبکاری نیز به تبع آن بالا می‌رود. یکی از نمونه‌های

امروزه نامه‌های الکترونیکی^۱ یکی از راه‌های عمومی، تاثیرگذار و با هزینه پایین در سطح اینترنت می‌باشد که با سرعت زیادی در حال رشد است. یکی از زمینه‌هایی که به وفور از نامه‌های الکترونیکی استفاده می‌شود، حوزه تجارت

^۲ Electronic Commerce

^۳ Email marketing

^۱ Electronic Mail

بگیریم، آنگاه خطای FP^۷ شامل نامه‌های الکترونیکی می‌شود که به اشتباه جزو هرنامه‌ها دسته‌بندی می‌گردند. خطای FN^۸ هم شامل نامه‌های الکترونیکی می‌شود که به اشتباه جزو نامه‌های الکترونیکی معتبر دسته‌بندی می‌شود. این خطاها در زمینه بازاریابی و تبلیغات از طریق نامه‌های الکترونیکی مشهودتر است. در مواجهه با این مشکلات بعضی از شرکتها اقدام به طراحی ضد هرنامه سازگار با زمینه تبلیغات می‌کنند [۵،۲۳].

از آنجاییکه بیشتر هرنامه‌ها در حوزه بازاریابی و نامه‌های الکترونیکی مطرح می‌شود لذا لازم است، که در طراحی ضد هرنامه‌ها^۹ دید صحیحی نسبت به حوزه تجارت الکترونیک داشته باشیم. در صورت نداشتن چنین دیدی در طراحی، ضد هرنامه‌ها با سرویس‌دهنده‌های نامه‌های الکترونیکی تبلیغاتی و بازاریابی هماهنگ نخواهند بود لذا درصد زیادی از نامه‌های الکترونیکی منتشر شده از این سرویس‌دهنده‌ها به جای هرنامه فیلتر^{۱۰} شده و هزینه زیادی را به این سرویس‌دهنده‌ها تحمیل می‌کنند. در صورتی که اگر این ضد هرنامه‌ها برای دیدگاه خاص مانند تبلیغات از طریق نامه‌های الکترونیکی طراحی شوند و اهداف مشخصی را دنبال کنند بهتر عمل خواهند کرد [۴،۷].

مشکل مهم دیگر که بیشتر در حیطه تجارت الکترونیک و تبلیغات از طریق نامه‌های الکترونیکی مطرح می‌شود، در نظر گرفتن مطلق یک نامه الکترونیکی خاص به عنوان هرنامه یا نامه معتبر است. این در حالی است که امکان دارد بعضی از نامه‌های الکترونیکی برای عده‌ای از کاربران هرنامه و برای عده‌ای دیگر هرنامه نباشد. برای مثال در تبلیغات از طریق نامه‌های الکترونیکی امکان دارد خرید اتومبیل برای کسی که قصد خرید اتومبیل دارد هرنامه حساب نشود در صورتیکه برای بعضی دیگر که قصد خرید ندارند هرنامه حساب شود. پس در این شرایط دسته‌بندی نامه‌های الکترونیکی دچار مشکل می‌شود که این عدم تخمین به صورت مطلق، در نامه‌های الکترونیکی تبلیغاتی فراوان وجود

سوءاستفاده از این روش ارتباطی، ارسال کورکورانه نامه‌های الکترونیکی ناخواسته و بی‌دعوت به نام هرنامه^۴ می‌باشد [۱،۲،۳،۴،۵،۶،۷،۸،۹،۱۰،۱۱،۱۲،۱۳،۱۴].

تعاریف زیادی برای اسپم یا هرنامه و چپستی و تفاوت آن با نامه‌های معتبر^۵ وجود دارد. کوتاهترین تعریف متداول از بین تعاریف موجود در مورد هرنامه، آنرا به عنوان یک نامه الکترونیکی ناخواسته^۶ بیان می‌کند. با این حال تعاریف مشابه زیادی نیز وجود دارد که بیان می‌کند، هرنامه یک نامه الکترونیکی ناخواسته است که به طور نا مشخص و مستقیم یا غیرمستقیم توسط فردی که نسبتی با گیرنده ندارد، فرستاده شده است. همان‌طور که می‌توان دید نقطه اشتراک برای تعریف هرنامه ناخواسته بودن آن است. بر طبق تعریف مورد توافق، هرنامه درباره رضایت یا عدم رضایت است نه محتوا [۱،۴،۱۷،۲۳]. هرنامه‌ها مشکلات متعددی را به بار می‌آورند که برخی از آنها مستقیماً باعث ضررهای اقتصادی می‌شوند، مانند ایجاد ترافیک و اتلاف پهنای باند و برخی دیگر زمان زیادی را تلف می‌کنند تا کاربران نامه‌های زاید را جداسازی کنند. علاوه بر موارد بیان شده، بعضی از هرنامه‌ها باعث آزار روحی و ایجاد عدم امنیت و اطمینان می‌شوند و سرانجام باعث ایجاد مشکلات قانونی مانند تبلیغات هرمی و کلاهبرداری‌های اقتصادی می‌گردند [۲،۳،۴،۷،۱۲،۲۴].

برای رفع این مشکلات، روشهای زیادی را در مقالات مختلف مطرح کرده‌اند، تا با بالا بردن درصد تخمین و دقت، باری از این هزینه‌ها کم کنند و آرامش و اطمینان را برای کاربران در تمامی حوزه‌ها بوجود بیاورند. با این همه به نظر می‌رسد هنوز هم مشکلاتی در این راه وجود داشته باشد. یکی از این مشکلات وجود خطای زیاد در روشهای مطرح شده می‌باشد، که می‌تواند عامل تاثیرگذار در تجارت الکترونیک باشد. اگر در اینجا دسته هرنامه‌ها را به عنوان کلاس مثبت و دسته نامه‌های الکترونیکی معتبر را به عنوان کلاس منفی در نظر

^۷ False Positive

^۸ False Negative

^۹ Anti-spam

^{۱۰} Filter

^۴ Spam

^۵ Legitimate, Ham

^۶ Unsolicited E-mail

$$f(m, \theta) = \begin{cases} C_{spam} & \text{if } spam \\ C_{leg} & \text{else} \end{cases} \quad (1)$$

در این تابع m نام الکترونیکی است که باید دسته‌بندی گردد. بردار پارامتر θ حاصل آموزش دسته بند با استفاده از یک مجموعه داده است که قبلاً جمع‌آوری شده است که می‌توان آنرا به صورت فرمول (۲) بیان کرد:

$$\theta = \Theta(M), \quad (2)$$

$$M = \{(m_1, y_1), \dots, (m_n, y_n)\}, y_i \in \{C_{spam}, C_{leg}\}$$

$$\forall i, i = 1, 2, \dots, n$$

در این فرمول m ها نام‌های الکترونیکی هستند که قبلاً جمع‌آوری شده‌اند و y ها نیز برچسب متناظر آنها می‌باشد [۱]. عمده کارهای انجام شده در این زمینه را می‌توان به صورت جدول ۱ دسته‌بندی کرد. در این جدول بعضی از روشهای کلی توضیح داده شده است. روشهای بیان شده در این جدول مخصوصاً روشهای یادگیری ماشینی^{۱۵} از نظر تخمین و دقت پیش‌بینی نتیجه مطلوبی را در برداشته‌اند. با این همه، برای مقابله با مشکلات اصلی بیان شده لازم است که ضد هرزنامه‌های شخصی در حوزه بازاریابی و تبلیغات و سازگار با این حوزه تولید شود. در زمینه تولید ضد هرزنامه‌های شخصی شده بر مبنای پروفایل و رفتار کاربران، بعضی از کارهای انجام شده را می‌توان به صورت زیر بیان کرد هرچند این تحقیقات نیز به صورت تخصصی به حوزه بازاریابی و تبلیغات از طریق نام‌های الکترونیکی نپرداخته است.

سوسا و همکاران^{۱۶} یک روش تولید ضد هرزنامه شخصی شده همکارانه^{۱۷} را بررسی کرده‌اند. در این روش ابتدا پروفایل کاربران دسته‌بندی می‌شود سپس بر مبنای گزارشات رسیده از هر گروه دسته‌بندی نام‌های الکترونیکی انجام می‌پذیرد. بدیهی است، در این روش نیاز به انتقال اطلاعات مابین سرویس‌دهنده‌های مختلف می‌باشد که این انتقال از

دارد، که به آنها نام‌های الکترونیکی خاکستری^{۱۱} نیز می‌گویند. برای مقابله با این مشکل با توجه به رفتار کاربران اقدام به ساخت ضدهرزنامه‌های شخصی^{۱۲} می‌کنند که در مقالات متعددی در مورد روشهای مختلف بحث شده است [۲، ۴، ۵، ۲۰، ۲۱]

در این مقاله ما ابتدا پروفایل^{۱۳} جدیدی را که می‌تواند به شبیه‌سازی بهتر رفتار کاربران منجر شود تهیه می‌کنیم. سپس این پروفایل را همراه با نام‌های الکترونیکی تبلیغاتی ساختگی در حیطه کتابفروشی بر خط به دانشجویان ارائه کرده و پاسخ آنها را جمع‌آوری می‌کنیم. در ادامه روشهای موجود و مشهور برای دسته‌بندی نام‌های الکترونیکی را مورد آزمایش و مقایسه قرار می‌دهیم. اجرا و پیاده‌سازی روشهای انتخاب شده در نرم‌افزار کلمنتاین^{۱۴} انجام می‌پذیرد. در پایان به تجزیه و تحلیل هر یک از این روشها پرداخته و روش مناسب را برای زمینه بازاریابی و شخصی‌سازی انتخاب می‌کنیم.

سازماندهی بخشهای بعدی به این شکل می‌باشد: در بخش ۲ مقاله، به بررسی کارهای مرتبط و دسته‌بندی آنها می‌پردازیم سپس در بخش ۳ به بیان طراحی پروفایل و تولید داده‌ها و مشخصات جامعه آماری آن می‌پردازیم. در ادامه در بخش ۴ روش پیشنهادی خود را مطرح می‌کنیم. در بخشهای باقیمانده به ارزیابی نتایج و جمع‌بندی و ارائه پیشنهادات برای کارهای آتی می‌پردازیم.

۲. مروری بر ادبیات موضوع

به طور کلی برای دسته بندی و پیش بینی هرزنامه ها، تکنیک ها و روشهای زیادی مطرح شده است. دسته بندی نام‌های الکترونیکی، یک برنامه کاربردی است که بر اساس تابع (۱) پیاده‌سازی می‌شود:

^{۱۵} Machine learning

^{۱۶} Sousa and et al.

^{۱۷} Collaborative

^{۱۱} Gray, Grey

^{۱۲} Personalized Anti-spam

^{۱۳} Profile

^{۱۴} Clementine ۱۲،۰

طریق معماری 18 p2p انجام می‌پذیرد. این ضد هرزنامه به صورت غیر متمرکز بوده که در سرویس دهنده‌های نامه‌های الکترونیکی اجرا می‌شود. از معایب این روش می‌توان به نیاز به امنیت در حین انتقال اطلاعات، نیاز به پهنای باند بیشتر برای انتقال اطلاعات و سختی مدیریت غیر متمرکز را نام برد [۳،۷].

در مقابل راوی و همکاران 19 ، روش متمرکز دیگری را بیان کردند که در این روش در دو مرحله با استفاده از روش‌های شبکه عصبی و در سرور 20 نامه‌های الکترونیکی طرف فرستنده اجرا می‌شود. این عمل از اتلاف پهنای باند در ازای پیشگیری از انتقال هرزنامه‌ها جلوگیری می‌کند و به شکل کاملاً منطبق با افکار انسان شکل گرفته است. اما باز هم دارای معایبی از جمله متکی بر رفتار اشخاص خاص به خاطر عدم دستیابی به کل جامعه آماری دارد [۲۲].

از دیگر روش‌های اجرا شده استفاده از درخت تصمیم مانند $C_{4,5}$ می‌باشد. در این روش ییح و همکاران 21 ، پروفایل کاربران را همراه با نامه‌های الکترونیکی پاسخ داده شده جمع‌آوری کرده‌اند. سپس از طریق روش فرکانس معکوس سند ($TF-IDF^{22}$) به استخراج و انتخاب ویژگی‌های مناسب پرداخته است. در حقیقت این مرحله توکن کردن نام دارد که در آن متن نامه‌های الکترونیکی به ریشه کلمات اصلی تبدیل شده و کلمات پرکاربرد به صورت یک مقدار باینری در نظر گرفته می‌شود. در صورت وجود این کلمه در یک متن مقدار آن یک و در غیر اینصورت مقدار آن برابر صفر در نظر گرفته می‌شود. بعد از این مرحله با استفاده از درخت تصمیم به تولید قوانین می‌پردازند. این قوانین تولید شده از جهت دقت 23 پیش‌بینی، مورد بررسی قرار گرفته و قوانین با دقت بالا انتخاب می‌شود [۱۳].

یکی دیگر از روش‌های مشابه در این زمینه استفاده از تولید داده‌های ساختگی برای نامه‌های الکترونیکی توسط کیم و

همکاران 24 می‌باشد. در این تحقیق به خاطر نیاز به پروفایل کاربران همراه با پاسخ نامه‌های الکترونیکی تبلیغاتی، لازم بود به تولید ساختگی این موارد در قالب پرسشنامه پرداخته شود. در این تحقیق از درخت تصمیم همراه با روش‌های معنایی مانند روش قبلی استفاده شده است. تفاوت این تحقیق با روش قبلی در مرحله استخراج و انتخاب ویژگی‌ها می‌باشد. این روش در مقایسه با روش قبلی به خاطر در نظر گرفتن شخصی‌سازی و خصوصیات نامه‌های الکترونیکی خاکستری می‌تواند نمونه مناسبی برای حوزه تجارت الکترونیک باشد. از جمله معایب این روش را می‌توان عدم تطابق بین پروفایل و محتوای نامه‌های الکترونیکی ساختگی با جامعه آماری پاسخ‌دهندگان بیان کرد. در این تحقیق تنها به برچسب زدن نامه‌های الکترونیکی تبلیغاتی در ۱۰ دسته توسط کارشناس انسانی 25 و ارائه آن به دانشجویان برای پاسخگویی پرداخته‌اند. به نظر می‌رسد که این دسته‌بندی جزئی بوده و جامعه آماری پاسخ‌دهنده نتواند جواب دقیق و واقعی را ارائه کند [۵].

تغییرات نتایج پیش‌بینی در بین روش‌های مرسوم تا حدودی زیاد است. این نتیجه به خاطر انتخاب ویژگی‌های متفاوت روش‌ها اتفاق می‌افتد. هر چند عمل مقایسه با دیگر کارهای مشابه به خاطر متفاوت بودن مجموعه داده‌ها و روش پیشنهادی نمی‌تواند مبنایی برای ارزیابی دقیق باشد اما در جدول ۲ به مقایسه اجمالی می‌پردازیم. اگر چه استفاده از تمامی محتوای نامه‌های الکترونیکی می‌تواند دقت نتایج را افزایش دهد اما به خاطر درگیر بودن با مجموعه بزرگی از ویژگی‌های استخراج شده، معمولاً نیاز به انتخاب ویژگی‌های مناسب می‌باشد. این حالت به ازای زمان و فضای مصرفی زیاد می‌تواند دقت پیش‌بینی را تا حدودی بهبود بخشد.

در هر حال روش‌های مختلفی برای شخصی‌سازی ضد هرزنامه‌ها وجود دارد که در آن از الگوریتم‌های یادگیری ماشینی و داده‌کاوی مانند شبکه بیزین 26 ، شبکه‌های

 18 Peer-to-peer 19 Ravi and et al. 20 Server 21 Yih and et al. 22 Term Frequency- Inverse Document Frequency 23 Accuracy 24 Kim and et al. 25 Human Expert 26 Bayesian network

کرد [۱۲]:

۳. استخراج ویژگی‌ها و آماده‌سازی داده‌ها

برای انجام این تحقیق لازم است که در ابتدا نامه‌های الکترونیکی همراه با برچسب (هرز یا معتبر بودن که از طرف کاربران نسبت داده شده است) و پروفایل کاربران جمع‌آوری شود. اگر چه مجموعه نامه‌های الکترونیکی با برچسب قابل اطمینان برای آزمون و آزمایش در شرکت‌های تحقیقاتی موجود می‌باشد، اما این مجموعه نامه‌های الکترونیکی یا فاقد پروفایل کاربران می‌باشند و یا در زمینه تبلیغات الکترونیکی نمی‌باشند. لذا از آنجایی که این اطلاعات در اکثر اوقات در دسترس نمی‌باشد باید مانند مقالات موجود در این زمینه به تولید آن پردازیم.

عصبی^{۲۷}، درخت تصمیم^{۲۸} و SVM^{۲۹} استفاده می‌کنند [۱،۵،۶،۸،۱۲،۱۶،۱۷،۱۹،۲۴]. با این همه هیچ یک از این موارد به طور تخصصی به امر تولید ضد هرزنامه‌های شخصی‌سازی شده با نگرش نامه‌های الکترونیکی خاکستری، بازاریابی و تبلیغات نپرداخته است. به خاطر اهمیت این حوزه و اینکه اکثر نامه‌های الکترونیکی امروزه را نامه‌های الکترونیکی خاکستری مخصوصا تبلیغاتی تشکیل می‌دهد نیاز به تولید ضد هرزنامه سازگار با این حوزه احساس می‌شود تا از این رهگذر نامه‌های الکترونیکی بازاریابی در سرویس‌دهنده‌های پست الکترونیکی هدف به عنوان هرزنامه فیلتر نشوند. از بررسی روش‌های بیان شده به این نتیجه می‌رسیم که برای طراحی و تولید ضد هرزنامه‌ها ۴ گام اصلی ضروری می‌باشد که در جدول ۳ می‌توان مشاهده

^{۲۷} neural network

^{۲۸} decision tree

^{۲۹} support vector machines

جدول ۱. انواع روشهای مهم برخورد با هرزنامه‌ها

هر شخص برای فرستادن ایمیل بایستی هزینه‌ای را بپردازد که برای هرزنامه‌نویسان این هزینه قابل توجه است [۱].	روش اقتصادی	روشهای غیر فیلتری (پیشگیری از ایجاد و انتقال هرزنامه)
بعضی از قانونگذاران برای حفظ امنیت و آرامش در جامعه مجازی اقدام به وضع قوانینی برای جلوگیری از تولید و انتشار هرزنامه‌ها کرده‌اند [۲۳].	روش قانونگذاری	
برای رفع نقص پروتوکلهای موجود، یک گام برای شناسایی هویت ارسال‌کننده نامه‌های الکترونیکی اضافه می‌شود [۱۸].	روش تغییر پروتوکلهای	
در این روش لیستی از آدرسهایی که به عنوان مبدا انتشار هرزنامه شناخته شده‌اند، تهیه می‌شود. هر نامه الکترونیکی که از این آدرسها فرستاده شود توقیف خواهد شد یا در نسخه دیگر فقط از آدرسهای مشخص شده موجود در لیست سفید، نامه الکترونیکی قبول می‌کند [۹].	لیست سیاه و سفید	روشهای فیلتری (بعد از انتقال هرزنامه سعی دارند که در سرویس دهنده‌های پست الکترونیکی به دسته بندی نامه های الکترونیکی به دو دسته هرزنامه و نامه معتبر پردازند هرچند هنوز بعضی از مشکلات بیان شده همچنان وجود دارد [۱،۳،۶،۱۲])
در این روش فیلدهای To, From و Cc و Bcc را از سرآیند نامه‌های الکترونیکی کاربران استخراج و بررسی می‌کنند سپس با استفاده از آنها گراف روابط اجتماعی کاربران را می‌سازند در نهایت با استفاده از این گراف روابط اجتماعی اقدام به دسته‌بندی نامه‌های الکترونیکی کاربران می‌کنند.	روشهای غیر محتوایی شبکه اجتماعی فرستندگان	
دانش راجع به رفتار که در پشت یک پیغام یا مجموعه‌ای از پیغامها قرار دارد را از دل ویژگی‌های غیرمتمنی استخراج می‌کند و سپس آن را با دانش از پیش تعریف شده‌ی (یا استخراج شده) مربوط به کاربرهای طبیعی و یا خرابکار، مقایسه می‌کند	رفتار فرستندگان	
اولین فیلترها به صورت سطحی فقط وجود یا عدم وجود یک سری توکنهای از پیش تعریف شده را در بدنه پیغام بررسی می‌کردند و بر مبنای یک سری قواعد ثابت عمل می‌کردند که به روش کلمات کلیدی و آماری معروف بودند که امروزه با روشهای یادگیری جایگزین شده اند.	توکن کردن (در این روش با استفاده از ابزار خاص اقدام به جداسازی کلمات مهم کرده و فقط از مهمترین آنها استفاده می‌کنند [۱،۱۶])	روشهای محتوایی (در این روشها از محتوای نامه های الکترونیکی مانند متن و عنوان آن استفاده می‌کنند [۱۰،۱۱])
در این روش برای پیش بینی از روشهای داده کاوی و یادگیری ماشینی بهره می‌برند به طوریکه این دسته بیشترین کاربرد را امروزه دارد و از تمامی ابزار داده کاوی که جهت دسته بندی و پیش بینی به کار می‌رود می‌توان استفاده کرد.	روشهای اولیه یادگیری ماشینی	
این روش ها بر این فرض استوار هستند که بدنه پیغامها به زبان طبیعی می‌باشند و روش هایی که مبتنی بر مدل‌های مقایسه ای - مانند عمل مقایسه ای مارکف و نیز پیش بینی هستند را استفاده می‌کنند.	آنالیز زبانی	

جدول ۲. کارهای مشابه انجام شده

مقاله	روش پیشنهادی	مجموعه داده	دقت	توضیحات
بینگ و همکاران [۱۱]	روش ترکیبی	خصوصیات کلی نامه‌ها همراه با فیلد های سرآیند	۹۱,۷۸٪	در این مقاله استخراج ویژگیها با نظر خبرگان و غیر محتوایی می باشد
کیم و همکاران [۵]	روش تولید قوانین از درخت تصمیم همراه با روش معنایی	تولید ساختگی نامه های الکترونیکی پرسشنامه ای برای صنایع مختلف	٪۸۵,۰	دقت به ازای هر قانون متفاوت است ما بیشترین را در نظر گرفته ایم.
سوسا و همکاران [۳]	روش همکارانه	تولید ساختگی داده ها همراه با داده واقعی	٪۹۳,۵	از محتوای نامه ها استفاده کرده است
یاون و همکاران [۱۵]	روش درخت تصمیم همراه با آنتولوژی	داده های واقعی همراه با تکرار چند باره	٪۹۷	از کلیه محتوا استفاده کرده و فیلتر کردن بر مبنای کلمات کلیدی خاص

جدول ۳. گام‌های طراحی ضد هزینه‌نامه

گامهای طراحی	توضیح	روش کار
گام اول : استخراج ویژگیها	در این روش ویژگیهای و کلمات پرکاربرد از متن نامه ها استخراج می شود. در شبیه سازی(تولید مجموعه داده) دیگر نیازی به مرحله استخراج نیست بلکه به صورت پیش فرض و با نظر خبرگان این ویژگیها مستقیماً پرسش می شود.	ابزار مختلفی مانند متن کاوی و TF-IDF و مشابه آن به کار برده می شود.
گام دوم : انتخاب ویژگیها	در این مرحله از بین ویژگیهای مختلف بعضی از ویژگیهای تاثیر گذار و با اهمیت انتخاب می شود.	برای اینکار از ابزاری مانند IG استفاده می کنند که این ابزار از مهمترین روشها بوده و در نرم افزار کلمتاین ابزاری بر این مبنا وجود دارد.
گام سوم : چارچوب پیشنهادی	در این مرحله به ارائه روشهای خود با استفاده از روشهای مختلف موجود در داده کاوی و یادگیری ماشینی در قالب چارچوب می پردازند.	طراحان سعی می کنند تا از طریق مقایسه یا ترکیب روشها به دقت بیشتری دست پیدا کنند.
گام چهارم : ارزیابی و اعتبار سنجی	دقت در طراحی وابسته به سه مرحله قبلی می باشد لذا ارزیابی تا حدودی از طریق معیارهای ارزیابی مشهور داده کاوی صورت می گیرد.	ارزیابی از طرق معیارها، ارزیابی از طریق مقایسه مولفه های نوآوری شده،

۳-۱- استخراج ویژگی‌های موثر نامه‌های الکترونیکی

بعد از انتخاب حوزه کاری و جامعه آماری اکنون لازم است که محتوایی برای هر یک از ویژگی‌های بیان شده استخراج کنیم. برای اینکار تعدادی از نامه‌های الکترونیکی تبلیغاتی در حوزه کتابفروشی برخط را انتخاب می‌کنیم. سپس با نظر خبرگان در امر تبلیغات و همچنین تعدادی از جامعه آماری اقدام به تهیه محتوا برای هر یک از ویژگی‌های مطرح شده می‌پردازیم. جدول ۴ ویژگی‌های استخراج شده نامه‌های الکترونیکی همراه با محتوا برای حوزه کاری مطرح شده را نشان می‌دهد.

در ویژگی سرآیند دو گزینه بیان شده است: اگر فرستنده نامه الکترونیکی تبلیغاتی برای گیرنده آشنا باشد (یعنی در دفترچه آدرس شخص موجود باشد که این نوعی تبلیغات از طریق مشتریان سازمانها می‌باشد که می‌توانند کالا یا خدماتی را به دوستان خود سفارش کنند).

اگر آدرس فرستنده برای گیرنده آشنا نباشد (در حقیقت تبلیغات از طریق شرکتهای موجود و با آدرسهای مختلف انجام شود). در ویژگی عنوان و متن گزینه‌های مختلف و جذابی که در نامه‌های الکترونیکی تبلیغاتی می‌تواند وجود داشته باشد بیان شده است. ویژگی بعدی ویژگی‌های کلی نامه‌های الکترونیکی می‌باشد. محتوای نسبت داده شده به این ویژگیها دارای حالت عمومی هستند و می‌توانند در دیگر حوزه‌ها نیز مطرح شوند. به خاطر وجود ویژگیهای زیاد ما بر مبنای نظر خبرگان و اصول بازاریابی مطرح شده روزیتر و بلمن^{۳۴} [۱۴]، دو دسته اصلی برای این ویژگیها را در نظر گرفتیم [۱۶].

جدول ۴: ویژگی استخراج شده بر مبنای حوزه کاری

ویژگی کلی	ویژگی انتخاب شده بر مبنای حوزه کاری	گزینه های موجود
سرآیند	فرستنده نامه الکترونیکی	۱، از طرف آدرس فرستنده آشنا
		۲، از طرف آدرس فرستنده ناآشنا
موضوع	عنوان نامه الکترونیکی	۱، کتابهای رایگان

تولید این مجموعه نامه‌های الکترونیکی از طریق بررسی مجموعه مقالات موجود و نظر خبرگان همراه با بررسی نامه‌های الکترونیکی تبلیغاتی صورت می‌پذیرد. روش کار برای تولید نامه‌های الکترونیکی تبلیغاتی در قالب پرسشنامه به شرح زیر می‌باشد. ابتدا طبق تمامی تحقیقات موجود در این زمینه به این مسئله می‌پردازیم که چه ویژگی‌هایی می‌تواند از نامه‌های الکترونیکی استخراج شود.

سه دسته ویژگی اساسی، شامل موارد زیر را می‌توان نام برد [۱]:

- ویژگی‌هایی از سرآیند^{۳۰}
- ویژگی‌هایی از متن^{۳۱} نامه که شامل موضوع^{۳۲} هم می‌باشد

- ویژگی‌هایی از کل ساختار^{۳۳} نامه الکترونیکی

از آنجاییکه هدف ما پالایش و دسته‌بندی نامه‌های الکترونیکی تبلیغاتی می‌باشد لذا لازم است که محتوای نامه‌های الکترونیکی نیز در همین راستا باشد. در کارهای مشابه انجام شده بدلیل اینکه چنین محدودیتی وجود نداشته، نامه‌های الکترونیکی از صنایع مختلف را برای محتوای نامه‌های الکترونیکی در نظر گرفته‌اند. [۵]

در چنین مواردی بدون در نظر گرفتن ویژگی‌های جامعه آماری پاسخ‌دهندگان اقدام به تهیه محتوای نامه‌های الکترونیکی کرده‌اند. ما برای دقت در این تحقیق ابتدا جامعه آماری پاسخ‌دهندگان خود را در نظر گرفته و سپس اقدام به تهیه محتوای نامه‌های الکترونیکی می‌کنیم.

جامعه آماری پاسخ‌دهندگان ما را جامعه دانشگاهی و دانشجویان تشکیل می‌دهند لذا لازم است حوزه‌ای را برگزینیم که افراد آگاهی و تمایل نسبت به این حوزه داشته باشند. در نتیجه ما نامه‌های الکترونیکی تبلیغاتی کتابفروشی برخط را به عنوان نمونه مطالعه موردی انتخاب می‌کنیم.

^{۳۰} Header

^{۳۱} Bodey

^{۳۲} Subject

^{۳۳} General structure

^{۳۴} Rossiter and Bellman. (۲۰۰۵)

متفاوت ویژگیهای رفتاری مختلفی دارند. بعضی از افراد عنوان می‌کنند که باید هیچ یک از نامه‌های الکترونیکی معتبر آنها به اشتباه فیلتر نشود و در مقابل دریافت چند هرزنامه روزانه را قبول می‌کنند، مخصوصاً وقتی که این هرزنامه‌ها، نامه‌های الکترونیکی تبلیغاتی باشد. در مقابل بعضی از افراد راضی به دریافت هیچ هرزنامه‌ای نیستند، هر چند بعضی از نامه‌های الکترونیکی معتبر آنها به اشتباه فیلتر شود. در حقیقت با این گزینه دو گروه افراد متفاوت را از لحاظ رفتاری می‌توان تشخیص داد.

این بخش مطابق با گام اول یعنی گام استخراج ویژگیها از نامه‌های الکترونیکی می‌باشد. در اینجا ما نیاز به ابزاری خاص برای توکن کردن و غیرساختاری کردن متن نیاز نداریم. در حقیقت با استخراج ویژگیها و محتوا با روش مطالعه مقالات مشابه، مطالعه نمونه موردی، نظر خبرگان و رتبه‌بندی آنها توانستیم به ویژگیهای مورد نیاز خود دست یابیم.

همانطور که ملاحظه می‌شود یک نمونه نامه الکترونیکی از ضرب دکارتی موارد بیان شده از جدول ۴ حاصل می‌شود که تعداد $2 * 3 * 5 = 30$ قالب نامه الکترونیکی بدست می‌آید. تعداد سوالات پروفایل نیز برابر با ۱۰ عدد می‌باشد که هر یک از آنها مقادیر مختلفی می‌تواند داشته باشد. در نتیجه هر پرسشنامه شامل ۱۰ سوال برای پروفایل کاربران و تعداد ۶۰ نامه الکترونیکی ساختگی و پرچسب پاسخگویی (هرزنامه یا معتبر) می‌باشد.

این پرسشنامه بعد از طراحی از طریق وب و در برخی موارد به صورت رودرو توسط ۷۰ نفر از دانشجویان پاسخ داده شد، که پس از اعمال پاکسازی تعداد ۶۶ عدد از آنها مورد استفاده قرار گرفت که از این تعداد ۳۰ نفر را زن و ۳۶ را مرد تشکیل می‌دهد. در ادامه ما داده‌های جمع‌آوری شده را به صورت تصادفی درهم کردیم سپس این داده‌ها را به دو قسمت مساوی تقسیم کردیم. در ادامه برای ارزیابی چارچوب از دو نوع مجموعه داده زیر استفاده کردیم. (هر رکورد شامل ۱۰ فیلد پروفایل و ۴ فیلد ویژگی نامه الکترونیکی و یک برچسب پاسخگویی می‌باشد).

- مجموعه داده نوع اول: تعداد ۱۸۴۳ رکورد که شامل ۱۱۷۲ رکورد هرزنامه می‌باشد.

۲، تازه ترین کتاب	زمینه و محتوای نامه الکترونیکی	متن
۳، پرفروشترین کتاب		
۱، مهندسی و علوم پایه		
۲، پزشکی		
۳، علوم انسانی		
۴، هنر	ویژگیهای دسته بندی شده نامه الکترونیکی	ساختار کلی
۵، سایر(غیر تخصصی و متفرقه)		
۱، ساینز حافظه زیاد (شامل گرافیک یا ضمیمه و...)	ویژگیهای دسته بندی شده نامه الکترونیکی	ساختار کلی
۲، ساینزحافظه کم (شامل فقط متن و...)		

۲-۳- استخراج پروفایل کاربران

مرحله بعدی استخراج ویژگیها که مهمترین مرحله نیز می‌باشد پروفایل کاربران می‌باشد. در اکثر تحقیقات از پروفایل‌های استاندارد موجود در اکثر سایتها مانند شغل، جنسیت، تحصیلات، رشته تحصیلی، سن، علاقه‌مندی و غیره استفاده شده است. از آنجاییکه شخصی‌سازی بر مبنای پروفایل کاربران شکل می‌گیرد، لذا لازم است گزینه‌های دیگری نیز برای بالا بردن دقت در نظر گرفته شود. برای اینکار ما دو گزینه را از مقالات مختلف جمع‌آوری کرده و در پروفایل خود قرار می‌دهیم [۵، ۱].

گزینه اول تعداد دفعاتی است که یک شخص بعد از دریافت یک نامه الکترونیکی آن را هرزنامه اعلام می‌کند. این گزینه برای افراد مختلف متفاوت است به صورتیکه امکان دارد یک شخص در مرحله اول یک نامه الکترونیکی را هرزنامه اعلام کند در صورتیکه امکان دارد شخص دیگری در دفعات تکرار زیاد نامه الکترونیکی مذکور را هرزنامه اعلام کند. گزینه‌ای مشابه این گزینه با نام قدرت مورد انتظار برای ضد هرزنامه شخصی شده وجود دارد [۵]. علت اصلی قرار دادن چنین گزینه‌ای در پروفایل اشخاص به خاطر نامه‌های الکترونیکی خاکستری می‌باشد.

گزینه دومی که در پروفایل اشخاص قرار داده شده و مورد پرسش قرار می‌گیرد، نسبت خطاهای مورد تحمل شخص در فیلتر کردن است که می‌تواند قبول کند. در حقیقت افراد

کنیم که در بخش آموزش چارچوب ایجاد می‌شود و در بخش آزمایش به ارزیابی چارچوب می‌پردازیم.

بخش اول پایگاه داده‌ها در کل شامل موارد زیر می‌باشد: پروفایل کاربران: این پروفایلها از پاسخ‌دهندگان به نامه‌های الکترونیکی ساختگی جمع‌آوری شده است. نامه‌های الکترونیکی: این پایگاه داده نتیجه مطالعات، نظر خبرگان، بررسی نامه‌های الکترونیکی مختلف در حیطه تبلیغات می‌باشد.

پاسخهای جمع‌آوری شده: این بخش برچسب نسبت داده شده از طرف پاسخ‌دهندگان به نامه‌های الکترونیکی شبیه‌سازی شده می‌باشد. در حقیقت شامل دو گزینه نامه معتبر یا هرزنامه می‌باشد.

بخش دوم پایگاه داده‌ها شامل پیش‌بینی‌های چارچوب می‌باشد این قسمت شامل دو پایگاه داده به صورت زیر می‌باشد:

پایگاه داده معتبر: این پایگاه داده در حقیقت شامل پیش‌بینی نامه‌های معتبر چارچوب می‌باشد. پایگاه داده هرزنامه: این پایگاه داده شامل نامه‌های الکترونیکی است که چارچوب آنها را به عنوان هرزنامه شناخته و برای استفاده آتی در اینجا ذخیره کرده است.

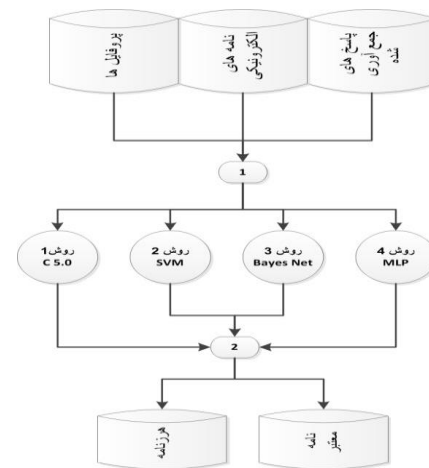
مسیر ۱ از هر سه پایگاه داده موجود در چارچوب استفاده می‌کند. این مسیر خود به تنهایی به ۴ مسیر فرعی منشعب می‌شود. در هر مسیر فرعی چارچوب از روشهای داده کاوی و یادگیری ماشینی برای پیش‌بینی استفاده شده است. هر روش شامل دو قسمت انتخاب ویژگیها از میان ویژگیهای استخراج شده و انجام اعمال پیش‌بینی را شامل می‌شود. در مسیر ۲ عمل مقایسه ما بین نتایج بدست آمده صورت می‌گیرد. چهار روش یا الگوریتم استفاده شده در این مسیر به شرح زیر می‌باشد [۶،۱۰،۱۲،۱۶]:

مسیر ۱ از هر سه پایگاه داده موجود در چارچوب استفاده می‌کند. این مسیر خود به تنهایی به ۴ مسیر فرعی منشعب می‌شود. در هر مسیر فرعی چارچوب از روشهای داده کاوی و یادگیری ماشینی برای پیش‌بینی استفاده شده است. هر روش شامل دو قسمت انتخاب ویژگیها از میان ویژگیهای استخراج شده و انجام اعمال پیش‌بینی را شامل می‌شود. در مسیر ۲ عمل مقایسه ما بین نتایج بدست آمده صورت

• مجموعه داده نوع دوم: تعداد ۱۸۴۳ رکورد که شامل ۹۵۹ رکورد هرزنامه می‌باشد.

۴. انتخاب ویژگی‌ها و چارچوب پیشنهادی

بعد از گام اول نوبت به گام دوم می‌رسد. در این گام لازم است که از بین ویژگیهای موجود در پروفایل و ویژگیهای نامه‌های الکترونیکی ساختگی بهترین آنها را برای چارچوب پیشنهادی خود انتخاب کنیم. بدیهی است که تمامی ویژگیهای بیان شده نمی‌تواند برای چارچوب مفید واقع شود. در بعضی موارد حتی مشاهده شده است که وجود بعضی از ویژگیها باعث کاهش دقت شده است. ما برای انتخاب ویژگیهای مناسب از ابزار انتخاب ویژگی موجود در نرم‌افزار کلمنتاین که از روش IG بهره می‌برد، استفاده می‌کنیم. این روش در اکثر مقالات موجود در این زمینه به کار گرفته شده و نتیجه مطلوبی را به همراه داشته است [۵]. در گام سوم روش پیشنهادی خود را در قالب یک چارچوب ارائه می‌دهیم. در این چارچوب سعی می‌کنیم از اکثر روشهای مشهور و زیاد استفاده شده از حوزه داده کاوی، یادگیری ماشینی و آمار استفاده کنیم. شکل ۱ چارچوب پیشنهادی ما را نمایش می‌دهد.



شکل ۱. چارچوب ارائه شده

بخش اول پایگاه داده‌ها می‌تواند هم پایگاه داده مربوط به آموزش و هم آزمایش چارچوب را شامل شود. برای راحتی کار ما در این چارچوب فقط یکی از این دو را نمایش دادیم. اما بدیهی است که بعد از تقسیم مجموعه داده‌های موجود به دو بخش آموزش (۷۰٪ داده‌ها بر اساس اکثر تحقیقات) و آزمایش می‌توانیم هر کدام از آنها را به چارچوب اعمال

تفسیر نتایج را از اجرا و پیاده‌سازی چارچوب در نرم‌افزار کلمنتاین بدست می‌آوریم. نمودار کلی برای اجرای این مرحله را، می‌توان به صورت شکل ۲ در نرم‌افزار کلمنتاین نمایش داد. این ساختار برای هر دو مجموعه داده موجود یکسان بوده و مبنایی برای اجرا و پیاده‌سازی این روشها می‌باشد. در این نمودار ابتدا مجموعه داده وارد نرم‌افزار می‌شود. سپس اگر نیازی به فیلتر کردن بعضی از ویژگیها باشد اقدام به فیلتر کردن ویژگی مورد نظر می‌کنیم. در گره بعدی نوع داده‌های ویژگیها را برای نرم‌افزار مشخص می‌کنیم. مهمترین ویژگی که نوع داده آن باید به درستی مشخص شود، ویژگی مورد پیش‌بینی می‌باشد. گره بعدی گره پارتیشن می‌باشد. این گره وظیفه انتخاب تصادفی مجموعه آموزش (در اینجا ۷۰٪ مجموعه) و مجموعه آزمایش را دارد. گره‌های بعدی چهار روش عنوان شده در چارچوب می‌باشد. بعد از این مرحله پیاده‌سازی خود را در این نرم‌افزار اجرا می‌کنیم. هر یک از روشها بعد از اجرا دارای نتایجی می‌باشند که ما فقط بعضی از نتایج که برای ارزیابی چارچوب لازم است را ارائه می‌کنیم. چنانچه قبلاً نیز بیان شد در هر یک از روشها قبل از اجرا به انتخاب ویژگیهای با اهمیت می‌پردازیم، نمودار شکل ۳ نمونه‌ای از این انتخاب ویژگیهای با اهمیت را، برای شبکه عصبی نشان می‌دهد. در مقایسه این نمودارها برای هر چهار روش می‌توان موارد زیر را بیان کرد:

هر یک از روشها مجموعه‌ای از انتخاب ویژگیهای منحصر به خود را دارد. در این نمودارها سوال ۹ (tr۹) و سوال ۱۰ (df۱۰) که همان موارد اضافه شده در پروفایل کاربران در این تحقیق می‌باشد دارای جایگاه خوبی می‌باشند. دقت و نتیجه به دست آمده از هر روشی تا حدودی وابسته به ویژگیهای مورد استفاده در روش می‌باشد.

می‌گیرد. چهار روش یا الگوریتم استفاده شده در این مسیر به شرح زیر می‌باشد [۶،۱۰،۱۲،۱۶]:

C۵.۰: روش یا الگوریتم اول که همان درخت تصمیم نیز می‌باشد به وفور و در منابع مختلف برای اعمال پیش‌بینی استفاده می‌شود. برای ایجاد این درخت روشهای زیادی وجود دارد که امروزه با نرم‌افزارهای موجود و در دسترس به سادگی می‌توان از C۴.۵ و یا C۵.۰ استفاده کرد. در اینجا ما از C۵.۰ که در نرم‌افزار کلمنتاین وجود دارد استفاده کردیم.

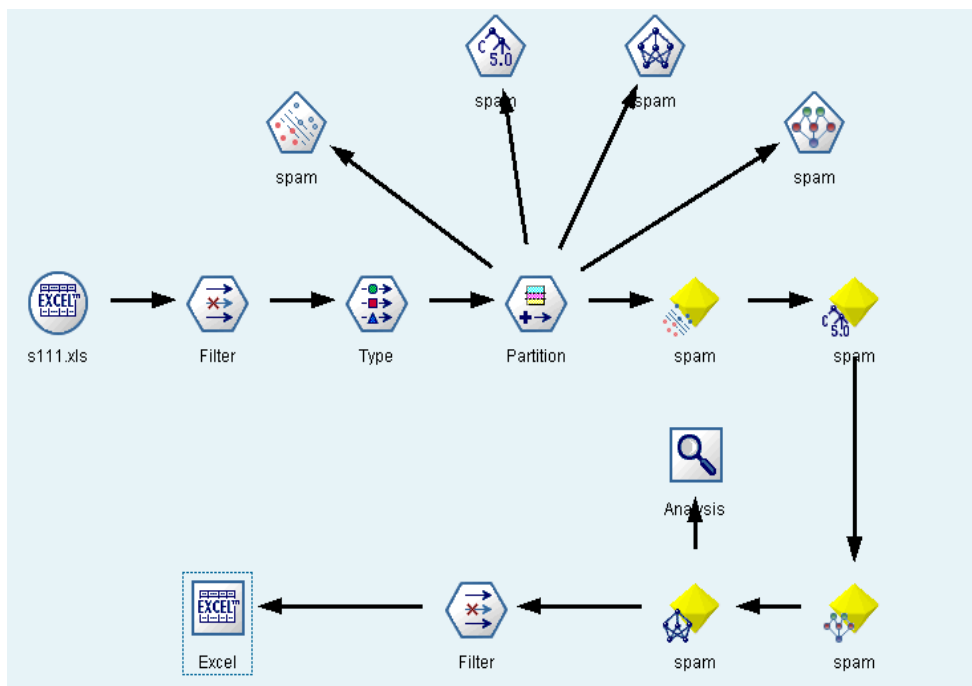
SVM: روش یا الگوریتم دوم که در اینجا استفاده می‌شود SVM می‌باشد. این روش برای پیش‌بینی بعضی از ویژگیها مخصوصاً در حیطه تصاویر کاربرد دارد.

BN: روش یا الگوریتم سوم همان شبکه بیزین می‌باشد. این الگوریتم در روشهای محتوایی به فراوانی و کرات مورد استفاده قرار گرفته است. این روش بیشتر مطابق با روشهای آماری و یادگیری ماشینی می‌باشد.

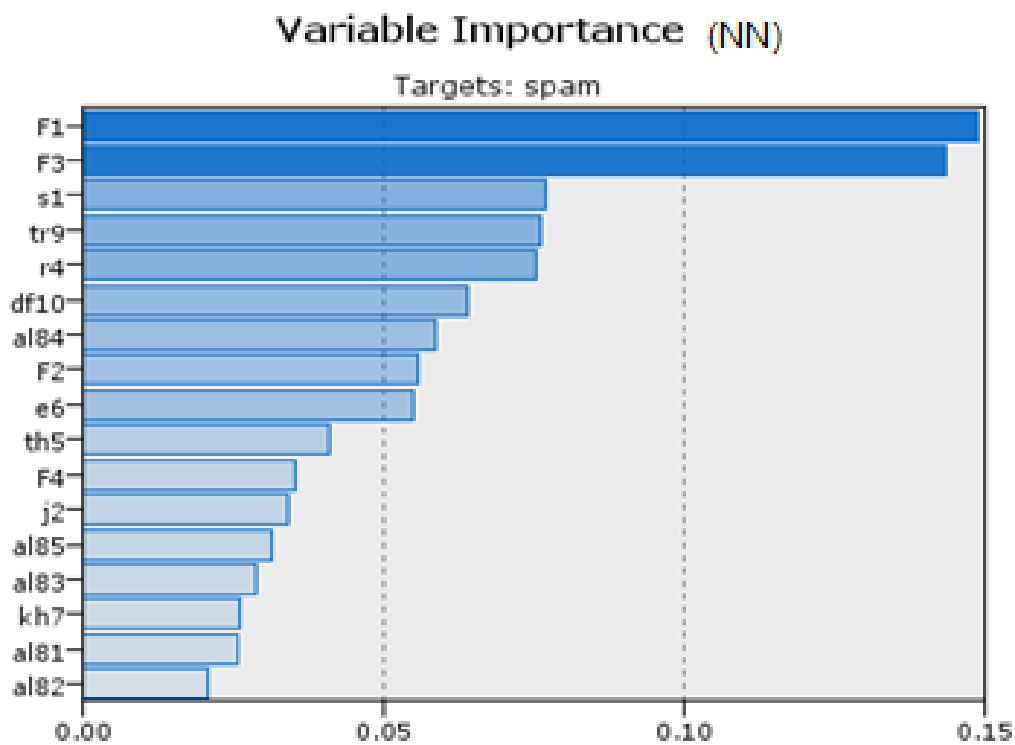
MLP: این روش یا الگوریتم همان روش شبکه‌های عصبی برای پیش‌بینی می‌باشد. انواع روشهای مختلفی برای پیش‌بینی با استفاده از شبکه‌های عصبی موجود می‌باشد. اما از میان روشهای مختلف موجود و ابزار موجود در نرم‌افزار کلمنتاین روشی را برمی‌گزینیم که نسبت به روشهای دیگر از نظر زمانی همخوانی داشته باشد. ما بیشتر از روش شبکه‌های عصبی چند لایه استفاده می‌کنیم که MLP^{۳۵} می‌تواند بهترین مورد هم از نظر زمان و هم از نظر پیش‌بینی باشد.

۵. ارزیابی نتایج و تفسیر آنها

بعد از اینکه داده‌های مورد نیاز خود را جمع‌آوری و پردازش کردیم اکنون نوبت به اجرا و پیاده‌سازی چارچوب ارائه شده می‌رسد. ما اطلاعات مورد نیاز برای ارزیابی، نتیجه‌گیری و



شکل ۲. اجرای چهار روش چارچوب در نرم‌افزار کلمنتاین



شکل ۳. متغیرهای با اهمیت در انتخاب ویژگی برای شبکه عصبی

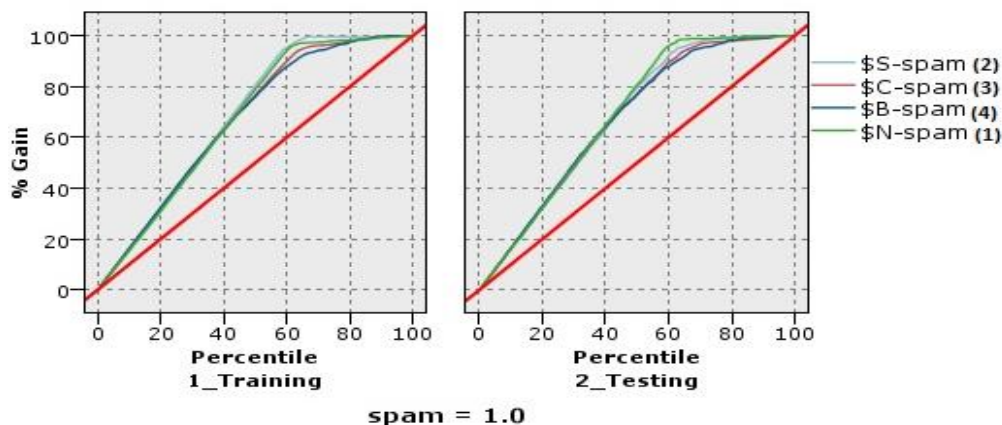
جدول ۵: معیارهای ارزیابی برای هر دو مجموعه داده: **a**:
 هرزنامه‌ای که به عنوان هرزنامه پیش‌بینی شده، **d**: نام
 معتبری که به عنوان نام معتبر پیش‌بینی شده، **b**: هرزنامه
 که به عنوان نام معتبر پیش‌بینی شده (FN)، **c**: نام معتبر
 که به عنوان هرزنامه پیش‌بینی شده (FP)

مجموعه داده روش معیار		مجموعه اول		مجموعه دوم	
		C0,0	۸۷,۸۷	۸۷,۱۷	۹۰,۵۱
Accuracy	$\frac{a+d}{a+d+b+c}$	SVM	۹۱,۳۹	۹۰,۵۱	
		BN	۸۵,۵۹	۷۷,۸۶	
		NN	۹۵,۴۳	۹۵,۰۸	
		C0,0	۱۲,۱۳	۱۲,۸۳	
Error Rate	1 - Accuracy	SVM	۸,۶۱	۹,۴۹	
		BN	۱۴,۴۱	۲۲,۱۴	
		NN	۴,۵۷	۴,۹۲	
		C0,0	۲۱,۶	۱۴,۵	
FP Rate	$\frac{c}{d+c}$	SVM	۱۴,۶	۸,۸	
		BN	۲۲,۵	۲۰,۹۹	
		NN	۷,۰	۴,۶	
		C0,0	۹۴,۱	۸۸,۸۸	
Spam Recall	$\frac{a}{b+a}$	SVM	۹۵,۳	۸۹,۹۳	
		BN	۹۰,۹	۷۶,۷۳	
		NN	۹۷,۰	۹۴,۷۹	
		C0,0	۸۶,۸۲	۸۶,۱۹	
Spam Precision	$\frac{a}{c+a}$	SVM	۹۰,۰۸	۹۱,۱۹	
		BN	۸۵,۹۵	۷۸,۹۲	
		NN	۹۵,۲۵	۹۵,۴۵	
		C0,0	۸۶,۸۲	۸۶,۱۹	

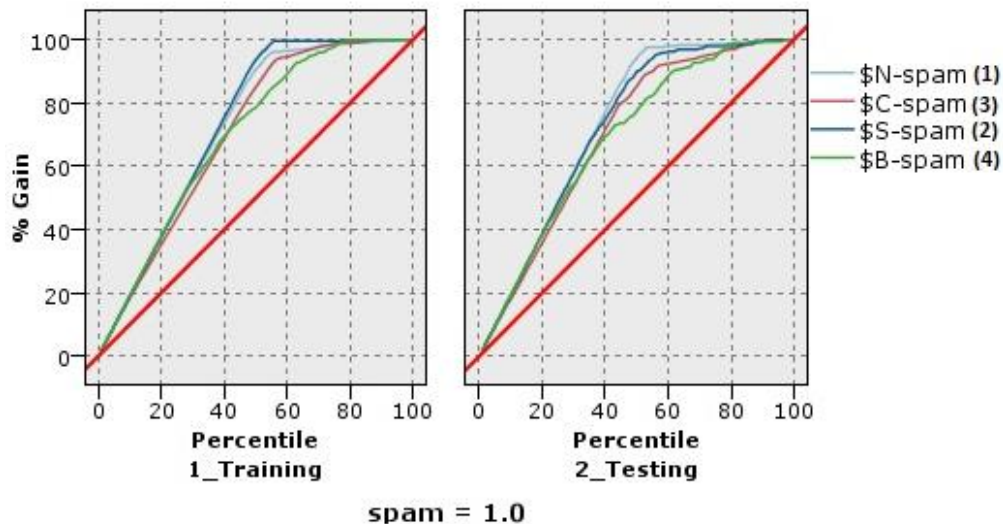
حال نتایج به دست آمده برای هر دو مجموعه داده اول و دوم را ارائه می‌کنیم. در جدول ۵ معیارهای ارزیابی [۱۸،۱۲،۱۶،۱۹] مورد مقایسه برای هر دو مجموعه بر مبنای چهار روش اجرا شده را می‌توان مشاهده کرد.

در هر یک از مجموعه داده‌ها به ازای هر روش تقریباً نتایج یکسانی حاصل می‌شود. در نتایج هر مجموعه داده شبکه عصبی بهترین نتیجه و پیش‌بینی را نسبت به دیگر روشها دارا می‌باشد. در نتیجه شبکه عصبی به عنوان پایدارترین و بهترین روش می‌تواند مورد توجه قرار گیرد.

در پایان نمودار بهره (Gain) برای مجموعه داده اول و مجموعه داده دوم به ترتیب در قالب نمودارهای شکل ۴ و ۵ برای هر چهار روش نمایش داده می‌شود. این نمودارها در حقیقت یک روش بصری و آماری برای کمک به درک کارایی روشهای مطرح شده می‌باشد. در این نمودارها ابتدا مجموعه داده‌ها به صد قسمت تقسیم شده و به صورت درصدی در محور افقی نمایش داده می‌شود، سپس به صورت تجمعی تعداد کل پیش‌بینی‌های درست هرزنامه به ازای کل پیش‌بینی‌ها در قالب نمودار در محور عمودی به صورت درصد نمایش داده می‌شود. این نمودارها گویای آن است که روش شبکه عصبی بهترین بهره را دارا می‌باشد. هر چند می‌توان این نمودار را به صورت نقطه‌ای و غیرتجمعی و با تقسیمات مختلف داده نمایش داد که از آن صرف‌نظر می‌کنیم.



شکل ۴. نمودار بهره Gain برای داده اول



شکل ۵. نمودار بهره Gain برای داده دوم

۶. نتیجه‌گیری و پیشنهاد کارهای آتی

دقت می‌باشد. این دقت و ثبات در نتایج بدست آمده از شبکه عصبی به خاطر نوع خاص روش شبکه عصبی می‌باشد که بر مبنای افکار انسان عمل می‌کند. به طور کلی می‌توان در هر چهار مرحله بیان شده طراحی نوآوری‌هایی دیگری در نظر گرفت. در مرحله اول می‌توان دسته‌بندی‌های دیگر و بر مبنای جامعه آماری دیگر شبیه‌سازی کرد. در قسمت چارچوب می‌توان از ترکیب روشهای موجود مانند رای‌گیری برای پیش‌بینی استفاده کرد. هر روش در حین انتخاب ویژگیها به صورت منحصر به فرد عمل می‌کند لذا هر یک بسته به ویژگیهای انتخابی دقت محدودی را در همان بازه کسب می‌کند. با ترکیب کردن نتایج روشهای مختلف می‌توان از حداکثر ویژگیهای استخراج شده بهره برد. برای دقت بیشتر می‌توان از روشهای معنایی، آنتولوژی همراه با روشهای همکارانه بهره برد. برای مثال می‌توان به خوشه‌بندی پروفایل‌های کاربران پرداخت. سپس با استفاده از این خوشه‌بندی در مراحل مختلف اقدام به بهبود دقت چارچوب کرد.

به طور کلی در ایران در زمینه دسته‌بندی و پالایش نامه‌های الکترونیکی در امر بازاریابی و تبلیغات پژوهش زیادی انجام نشده است، لذا در این طرح سعی بر ایجاد یک ضد هرزنامه شخصی شده برای تخمین اهمیت و دسته‌بندی نامه‌های الکترونیکی تبلیغاتی کاربران با توجه به رفتار. پروفایل آنها شده است. در واقع ما از سه منبع مقالات، مطالعه و بررسی نامه‌های الکترونیکی تبلیغاتی و نظر خبرگان برای انجام این تحقیق استفاده کرده‌ایم.

برای طراحی بهتر لازم بود که حوزه کاری و بعدی از تجارت الکترونیک که طراحی ضد هرزنامه برای آن صورت می‌پذیرد، مشخص شود. طراحی ضد هرزنامه برای نامه‌های الکترونیکی تبلیغاتی که بیشتر در حوزه نامه‌های الکترونیکی خاکستری قرار می‌گیرد، صورت پذیرفته است. در بعد تجارت الکترونیک، برای بازاریابی از طریق نامه‌های الکترونیکی تبلیغاتی و بیشتر برای بازاریابی B2C سازگار شده است. بعد از طراحی پرسشنامه، جمع‌آوری پاسخها، مقایسه روشها بر مبنای معیارهای ارزیابی مطرح شده و دو مجموعه داده مجزا با یکدیگر مشخص شد که شبکه عصبی دارای

منابع

۱۲. Saad O., Darwish A., Faraj R.,(۲۰۱۲) A survey of machine learning techniques for Spam filtering., International Journal of Computer Science and Network security, VOL.۱۲ No.۲, February.
۱۳. Yih W., McCann R., Kołcz A.,(۲۰۰۷) Improving Spam Filtering by Detecting Gray Mail, In Proceedings of the ۴rd Conference on Email and Anti-Spam.
۱۴. Rossiter J. R., Bellman S.,(۲۰۰۵) “Marketing Communications” Prentice Hall, English.
۱۵. Youn S., McLeod D.,(۲۰۰۹) Spam Decisions on Gray E-mail using Personalized Ontologies, Proceedings of the ۲۰۰۹ ACM Symposium on Applied Computing (SAC), Honolulu, Hawaii, USA, pp. ۱۲۶۲-۱۲۶۶.
۱۶. Guzella T.S., Caminhas W.M.,(۲۰۰۹) A review of machine learning approaches to Spam filtering , Expert Systems with Applications,vol. ۳۶,pp.۱۰۲۰۶-۱۰۲۲.
۱۷. Saad O., Darwish A., Faraj R.,(۲۰۱۲) A survey of machine learning techniques for Spam filtering., International Journal of Computer Science and Network security, VOL.۱۲ No.۲, February.
۱۸. Dwork C., Naor M.,(۱۹۹۲) Pricing via processing or combatting junk mail, In Advances in Cryptology - Crypto ۹۲ Proceedings, Springer Verlag, pp ۱۳۹-۱۴۷.
۱۹. SHI L., WANG Q., MA X., WENG M., QIAO H.,(۲۰۱۲) Spam Email Classification Using Decision Tree Ensemble, Journal of Computational Information Systems,vol. ۱: ۳,pp. ۹۴۹-۹۵۶.
۲۰. Spam definition.(۲۰۱۲) Available at [http://en.wikipedia.org/wiki/Spam_\(electronic\)](http://en.wikipedia.org/wiki/Spam_(electronic)).
۲۱. GrayEmail definition,(۲۰۱۲) Available at [http://en.wikipedia.org/wiki/Graymail_\(email\)](http://en.wikipedia.org/wiki/Graymail_(email)).
۲۲. Ravi J., Shi W., Xu C., (۲۰۰۵) Personalized Email Management at Network Edges, IEEE Internet Computing, Vol.۹(۲), pp.۵۴-۶۰.
۲۳. Nicola L.,(۲۰۰۴) European union vs. spam: A legal response, In Proceedings of the First Conference on Email and Anti-Spam, CEAS'۲۰۰۴.
۲۴. Rafiqul I., Jemal A.,(۲۰۱۳) A multi-tier phishing detection and filtering approach, Journal of Network and Computer Applications, Volume 36, Issue 1, January 2013, Pages 324–335.
۱. Blanzieri E., Bryl A. ,(۲۰۰۸) A survey of learning-based techniques of email spam filtering, Artif Intell Rev, vol.۲۹,pp.۶۳-۹۲.
۲. Cukier W. L., Cody S., Nesselroth E. J., (۲۰۰۶) Genres of Spam: Expectations and Deceptions, Proceedings of the ۳۹th Hawaii International Conference on System Sciences, .
۳. Sousa p., et al,(۲۰۱۰) A Collaborative Approach for Spam Detection ,Second International Conference on Evolving Internet, IEEE.
۴. Raad M.,et al,(۲۰۱۰) Impact of spam advertisement through e-mail: A study to assess the influence of the anti-spam on the e-mail marketing, African Journal of Business Management, Vol. ۴(۱۱), pp. ۲۳۶۲-۲۳۶۷.
۵. Kim J., Dou D., Liu H., Kwak D., (۲۰۰۷) Constructing a User Preference Ontology for Anti-spam Mail Systems, Canadian AI ۲۰۰۷, LNAI ۴۵۰۹, pp. ۲۷۲ – ۲۸۳.
۶. Kakade A.G., Kharat P.K., Gupta A.K.,(۲۰۱۳), Survey of Spam Filtering Techniques and Tools, and Map Reduce with SVM, IJCSMC, Vol. ۲, Issue. ۱۱, November ۲۰۱۳, pg. ۹۱ – ۹۸.
۷. Wenxuan S., Maoqiang X.,(۲۰۱۳) A Reputation-based Collaborative Approach for Spam Filtering, ۲۰۱۳ AASRI Conference on Parallel and Distributed Computing and Systems, Volume ۵, ۲۰۱۳, Pages ۲۲۰-۲۲۷
۸. Almeida, T. A., Yamakami, A.,(۲۰۱۲) Facing the spammers: A very effective approach to avoid junk e-mails, Expert Systems with Applications,vol. ۳۹, pp.۶۵۵۷-۶۵۶۱.
۹. Cook D., Hartnett J., Manderson K., scanlan J., (۲۰۰۶) catching Spam Before it Arrives: Domain Specific Dynamic Blacklists , ACM International Conference Proceeding Series; Vol.۱۶۷,pp.۱۹۳-۲۰۲.
۱۰. Almeida T.A., Yamakami A.,(۲۰۱۰) Content-Based Spam Filtering, The ۲۰۱۰ International Joint Conference on Neural Networks (IJCNN), IEEE.
۱۱. Ying K.C., et al,(۲۰۱۰) An ensemble approach applied to classify spam e-mails, Expert Systems with Applications.vol ۳۷,pp. ۲۱۹۷-۲۲۰۱.

استخراج معماری فناوری اطلاعات بیمارستان با رویکرد قابلیت پیاده سازی در ایران

*عاطفه السادات حقیقت حسینی** حسین بوبرشاد*** فاطمه ثقفی* هادی زارع*
*کارشناسی ارشد مهندسی مهندسی فناوری اطلاعات پزشکی، دانشکده علوم و فنون نوین، دانشگاه تهران
**استادیار دانشکده علوم و فنون نوین، دانشگاه تهران
***استادیار دانشکده مدیریت، دانشگاه تهران
تاریخ دریافت: ۹۳/۰۹/۰۲ تاریخ پذیرش: ۹۵/۰۴/۱۵

چکیده

امروزه ارائه خدمات هوشمند و سریع به بیماران و حرکت به سمت بیمارستان‌های نسل آینده، از ضروریات حوزه سلامت است. تهیه معماری اطلاعات برای بیمارستان‌ها مقدمه دستیابی به خدمات هوشمند و به تبع آن ارائه خدمات با سرعت و کیفیت بسیار بالاتر نسبت به سیستم‌های سنتی است. در این مقاله هدف ارائه یک معماری اطلاعات بومی مبتنی بر معیارها و شاخص‌های مهم در بیمارستان‌های ایران بوده است. در این مقاله از معماری فناوری اطلاعات TOGAF استفاده و سپس این معماری متناسب با شرایط بومی ایران برای بیمارستان شریعتی بومی سازی شده است. برای این کار از نظر خبرگان و ابزار پرسشنامه ۱۳۴ سوالی بهره گرفته شده و بر اساس آزمون‌های مناسب آماری مورد تحلیل قرار گرفته است. در این مقاله معماری فناوری اطلاعات در قالب یک مدل مفهومی دارای چهار ورودی و چهار لایه زیرساختی طراحی شد. نتایج نشان داد از میان ۱۳۴ جزء معماری فناوری اطلاعات توگف، تعداد ۱۰۴ جزء، توسط خبرگان برای کاربرد در بیمارستان مورد تایید قرار گرفت. این چارچوب اختصاصی سازی شده، معماری فناوری اطلاعات بیمارستان نامیده شد. معماری پیشنهادی که در هشت لایه و یازده جزء ترسیم شد و می‌تواند به عنوان یک معماری مرجع بومی برای اجرای معماری اطلاعات در بیمارستان‌های ایران مورد استفاده قرار گیرد.

واژه‌های کلیدی: بیمارستان‌های نسل آینده، معماری فناوری اطلاعات، معماری توگف

دسترسی خواهد شد و برای نگهداری از این حجم بزرگ اطلاعات فناوری داده‌های حجیم مورد استفاده واقع می‌شود. در این شرایط بیمارستانی که توان استفاده از این فناوری را به دلیل نبود عماری مناسب نداشته باشد توان ارائه خدمات در سطح بین المللی را نخواهد داشت. محیط متغیر و ناپایدار امروزی سازمان‌ها، نتیجه افزایش اعجاب انگیز سهولت تبادل اطلاعات و پیدایش جامعه‌ی اطلاعاتی است. در چنین شرایطی درک دقیق اجزای متنوع سازمان از

مقدمه

فناوری‌های عصر آینده در قالب سه فناوری اصلی داده‌های حجیم (Big Data)، اینترنت اشیا (Internet of Things) و رایانش ابری (Computing Cloud) خواهد بود. لذا در بیمارستان‌های نسل آینده، همه اجزا اعم از بیماران، دستگاه‌ها و اتاق‌ها با شماره‌های شناسایی و سنسورهای خاص مشخص شده و به اینترنت متصل خواهند بود، اطلاعات آنها از طریق رایانش ابری در سراسر دنیا قابل

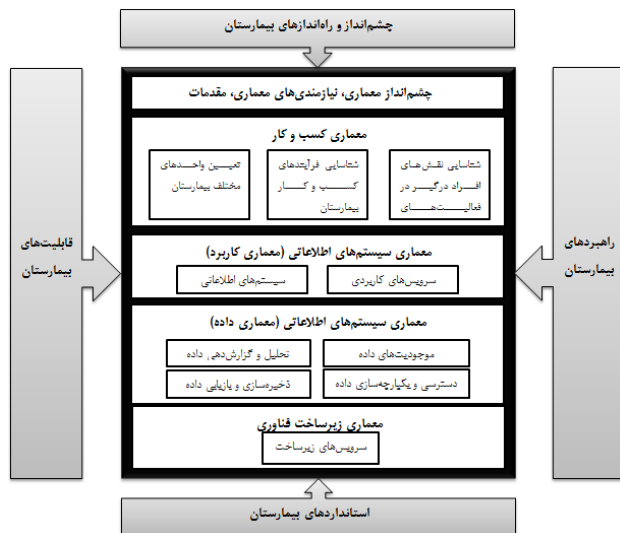
چین، چالش ارائه خدمات را در تنوع و ناهمگونی سیستمهای بیمارستانی دانسته است. و به طراحی مجدد سیستمها و ساختارهای پزشکی پرداخته است [۱۱]. در پژوهش دیگری در کشور انگلستان، با توجه به پیچیدگی طراحی نظام مراقبت‌های پزشکی به علت پیچیدگی فعالیت‌ها که ناشی از چالش‌های مدیریت منابع و بهبود خدمات ارائه می‌باشد، به مدلسازی مراقبت‌های پزشکی از دید فرآیندها با استفاده از معماری اطلاعات پرداخته شده است. نتایج این تحقیق نشان می‌دهد که معماری اطلاعات می‌تواند کمک شایانی در همگرایی فرآیندهای مراقبت پزشکی و فناوری اطلاعات نماید و مدیریت مراقبت پزشکی را بهبود بخشد. [۱۲]. جستجو در پایگاههای داده های مقالات الزویر با دو کلید واژه Hospital and architecture نشان داد که تنها ۲۲ مقاله در این زمینه وجود دارد که کلمات فوق در عنوان یا چکیده مقاله وجود داشته باشد. با بررسی این مقالات تنها دو مقاله مرتبط با معماری یافت شد. از این دو مقاله Rijo و دیگران [۱۳] برای تامین نیاز یک بیمارستان ۷۰۰ تخت در پرتقال به داشتن دیدگاه مدیریتی برای استفاده از فناوری‌های اطلاعاتی پیچیده روز، یک مدل مفهومی برای معماری بیمارستان ارائه کرده بودند. در این مقاله تنها به ۴ چارچوب معماری سازمانی زکمن، توگف، چارچوب فدرال و چارچوب گارنتر اشاره شده و سه مورد اول را به دلیل پیچیدگی چارچوب برای بیمارستان مناسب ندانسته است و چارچوب چهارم را نیز نامناسب دانسته است فقط به دلیل عملیاتی بودن بیان کرده که از این چارچوب می‌توان الگو گرفت. در عنوان Walsh و دیگران [۱۴] نیز کلمات معماری و بیمارستان وجود دارد ولی این مقاله در مورد مدل معماری سازمانی نیست بلکه در خصوص معماری مدل رهبری برای تواناسازی سیستم‌های سلامت است. با جستجو در گوگل نیز تنها یک مقاله یافت شد. در این مقاله Lu و دیگران [۱۱] در زمینه معماری سازمانی بیمارستان کار کرده‌اند. ولی مقاله آنها تاکید بر داشتن راهبرد برای خرید تجهیزات بیمارستانی است ولی معماری سازمانی انجام نشده است.

هدف اصلی این مقاله ارائه مدل معماری فناوری اطلاعات برای توانمند سازی بیمارستان‌های ایرانی برای پاسخگویی به

دیدگاه‌ها سیستمی برای انطباق محیط سازمانی با محیط متغیر و ناپایدار ضروری است [۱]. اهمیت تدوین معماری اطلاعات امروزه به قدری است که عدم تدوین آن به منزله‌ی در دسترس نبودن رویه‌های رسمی انجام کار و در نتیجه سازمان رسمی است. عدم دسترسی به سازمان رسمی فهم تعاملات بین سازمان رسمی و غیر رسمی را نیز برای مدیران ناممکن یا دشوار می‌سازد و در پی آن اصلاحات سازمانی در حد تغییرات صوری و جابجایی‌های سلیقه‌ای و مصرف تزئینی یا غیرکارای فناوری تنزل خواهد کرد [۲].

چارچوب معماری سازمانی یک ساختاری منطقی برای رده‌بندی، سازماندهی و ارائه‌های توصیفی از سازمان است که برای مدیریت و توسعه سیستم‌های آن سازمان حائز اهمیت است [۳]. این چارچوب تعریف مشخصی از جنبه‌ها و دیدگاه‌های مختلف سازمان را ارائه‌داده و توصیف‌های لازم را برای هر یک از آنها پیشنهاد می‌کند [۴]. زکمن چارچوبی را برای معماری سیستم‌های اطلاعاتی ارائه داد [۵]. به دنبال معرفی این چارچوب، برخی چارچوب‌های معماری سازمانی دیگر با بهره‌گیری از آن مطرح شدند. چارچوب معماری سازمانی فدرال [۶]، چارچوب معماری سازمانی TOGAF [۷] و چارچوب معماری سازمانی C4ISR [۸] نمونه‌های آن هستند. در پژوهشی ضمن ذکر گسترده‌ی مراقبت‌های درمانی، بیمارستان را یک سیستم پیچیده‌ی فنی و اجتماعی با ذینفعان زیاد و دارای خواسته‌ها و اهداف ناهمگون معرفی کرده است. از سوی دیگر بیمارستان را نیازمند ایجاد و کشف قابلیت‌هایی برای ارائه خدمات مراقبتی مناسب، در زمان مناسب، در مکان مناسب و با هزینه کافی به بیماران می‌داند. همچنین چالش اصلی در دستیابی به این اهداف نبود دیدگاه سیستماتیک و جامع به تمام حوزه‌های سازمان چه در داخل و چه در خارج از سازمان عنوان شده است. برای حل معضلات فوق در چند بیمارستان و مرکز درمانی در ایالات متحده آمریکا، چگونگی همراستا نمودن معماری سازمانی با مرزهای کنترل داخل و خارج سازمان بررسی شده است [۹]. در مقاله دیگری در یکی از بیمارستان‌های کشور ژاپن، معماری اطلاعات این بیمارستان تحلیل شده و مبتنی بر اهداف راهبردی بیمارستان، وضعیت مطلوب سازمان طراحی شده است [۱۰]. در تحقیق دیگری در کشور

اطلاعات توگف، بیمارستان را به عنوان مجموعه‌ای از واحدها که مجموعه اهداف مشترکی دارند تعریف می‌کند، یک ساختار منطقی برای طبقه‌بندی و سازماندهی اطلاعات پیچیده ایجاد می‌کند و دیدها و افق‌های گوناگون سیستم‌اطلاعاتی را تشریح می‌کند. در این تحقیق با مطالعه محتوی معماری توگف ۹،۱ [۱۹] و انطباق آن بر فازهای روش توسعه معماری [۲۰]، مدل مفهومی معماری فناوری اطلاعات توگف برای بیمارستان مطابق شکل (۱) بدست آمد.



شکل ۱: محتوی کلی مدل مفهومی معماری فناوری اطلاعات در بیمارستان

در این مدل مفهومی، بیمارستان دارای چشم‌انداز، راه‌انداز و راهبرد، استاندارد و قابلیت‌های خاص خود است. این عوامل ورودی چارچوب توگف هستند. مولفه‌های مورد نیاز برای هر کدام از ورودی‌های در نظر گرفته شده در این مدل، با بررسی مستندات و مطالعات وضعیت موجود مرکز آموزشی، پژوهشی و درمانی دکتر شریعتی و سایر بیمارستان‌های ایران و تحلیل وضعیت مطلوب بیمارستان‌های خارج از ایران [۲۱][۲۲] استخراج شدند و سپس با انطباق آن بر وضعیت کنونی بیمارستان شریعتی تهران سفارشی شدند که در جدول ۱ نشان داده شده است.

بیماران و تامین رضایت آنها است. برای این کار حقیقت حسینی و همکاران [۱۵] در مقاله ای ۱۷ چارچوب معماری سازمانی را مطالعه نموده و برای انتخاب چارچوب مناسب برای بیمارستان از روش تحلیل سلسله مراتبی AHP استفاده کردند و با نظر خبرگان و صاحب‌نظران در این حوزه، معیارها و شاخص‌های مهم با استفاده از مطالعات تنگ و همکاران [۱۶][۱۷]، شناسایی شده و سپس چارچوب توگف Togaf با بالاترین امتیاز به عنوان چارچوب مناسب برای معماری فناوری اطلاعات بیمارستان انتخاب شده است. از آنجا که پیاده سازی موفق این معماری به شرایط بومی هر کشور وابسته است. در این مقاله قصد بر آن است که با در نظر گرفتن موارد استفاده معماری، حیطه و خصوصیات معماری توسعه داده شود و از تجربیات موجود روی دیگر چارچوب‌ها، محدودیت منابع و زمان برای تولید محصولات، سیاست سازمان و نیاز به سازگاری با سازمان‌های دیگر نیز استفاده شود و چارچوب معماری برای طراحی سامانه‌های مورد نیاز بیمارستان شریعتی تعیین شود. لذا ابتدا عناصر و اجزا چارچوب معماری سازمانی از ادبیات موضوع و با استفاده از مصاحبه با خبرگان استخراج شد و سپس در قالب یک مدل مفهومی با استفاده از روش پرسشنامه و آزمون تحلیل واریانس یک طرفه (آنووا) و آزمون دو جمله‌ای، ارزیابی شده و برای استفاده در بیمارستان مشخص شد.

یافته‌ها

مدل مفهومی معماری توگف در این تحقیق، از شناخت وضعیت فعلی کلیه مولفه‌های اصلی بیمارستان‌ها آغاز می‌شود و با طرح‌ریزی برنامه‌ای بلند مدت که چگونگی دستیابی به تمامی اهداف و مأموریت‌های بیمارستان‌ها در آن گنجانده شده است، ادامه می‌یابد تا در نهایت معماری سازمانی وضعیت مطلوب بوجود آید. معماری فناوری

جدول ۱: عوامل ورودی تاثیرگذار در مدل مفهومی چارچوب توگف

بدهای ورودی	فرضیه‌های مورد قبول با اطمینان ۹۵٪	فرضیه‌های رد شده
چشم‌اندازها (Vision)	<p>۱. دارا بودن بالاترین سطح ارائه کیفیت خدمات براساس ارتقای عملکرد بالینی بیمارستانی</p> <p>۲. افزایش درصد رضایتمندی افراد در بهره‌مندی از خدمات واحدهای بیمارستانی</p> <p>۳. برترین بیمارستان در زمینه ارائه خدمات آموزشی، درمانی و پژوهشی</p> <p>۴. کاهش حوادث و نرخ مرگ و میر در بخش‌های بیمارستان</p>	<p>۱. درک بهتر از وضعیت سلامت جامعه</p> <p>۲. کاهش زمان اولین ویزیت پزشک بیمارستان در فوریت‌های پزشکی</p> <p>۳. افزایش عادلانه دریافت خدمات سلامت</p>
راه‌اندازها (Driver)	<p>۱. الزام به راه‌اندازی پرونده الکترونیک بیماران</p> <p>۲. الزام به توسعه و کاربری فناوری ارتباطات و اطلاعات در بخش بهداشت و درمان</p> <p>۳. الزام به تامین، حفظ و ارتقای سطح سلامت جامعه و بهبود کیفیت زندگی افراد</p> <p>۴. وجود قوانین مدون در مورد نظارت و کنترل امور آزمایشگاهی، دارویی و بهداشت محیط و کار</p> <p>۵. وجود قوانین نظام ارجاع در بیمارستان با محوریت پزشک خانواده</p> <p>۶. وجود قوانین برای کاهش تعرفه خدمات درمانی به نسبت سایر هزینه‌های زندگی</p>	<p>۱. افزایش انتظارات مردم برای امنیت اطلاعات سلامت</p> <p>۲. افزایش انتظارات افراد برای برخورداری از یک زندگی سالم و با کیفیت توأم با طول عمر قابل قبول و عاری از بیماری و ناتوانی</p>
راهبردها (Strategy)	<p>۱. به کارگیری فناوری اطلاعات و ارتباطات به‌عنوان پیشران و پشتیبان عملیات اصلی سازمان</p> <p>۲. ایجاد زیرساخت‌های الکترونیک لازم برای تدوین و اجرای کد ملی سلامت</p> <p>۳. استقرار نظام حاکمیت بالینی و نظارت عالی بر اجرای قوانین و مقررات مرتبط با کیفیت</p> <p>۴. ایجاد زیرساخت‌های لازم و توسعه نهادهای حمایتی برای توانا ساختن سرویس‌های حوزه IT</p> <p>۵. ایجاد بستر لازم برای ارائه نظام بیمه‌ای سلامت نگر</p> <p>۶. استقرار نظام مدیریت راهبردی در بخش حوادث و فوریت‌های پزشکی و رسیدگی به شکایات</p> <p>۷. استقرار نظام جامع مدیریت دارو، لوازم و تجهیزات پزشکی، دندانپزشکی و آزمایشگاهی کشور</p> <p>۸. استقرار نظام جامع سلامت در راستای ایجاد یک نظام پاسخگو</p> <p>۹. استقرار نظام اطلاع‌رسانی و ارتباطات بخش سلامت و آموزش پزشکی و نظام جامع آمار</p> <p>۱۰. استقرار نظام جامع مدیریت اطلاعات سلامت</p> <p>۱۱. استقرار نظام جامع مدیریت خدمات سلامت کشور با تاکید بر تخصیص بهینه و عادلانه منابع</p> <p>۱۲. استقرار نظام مستندسازی و گزارش‌گیری بخش سلامت و حسابداری ملی سلامت</p>	<p>تمامی فرضیه‌ها پذیرفته شدند.</p>
قابلیت‌ها (Capability)	<p>۱. ارائه خدمات پایه سلامت در قالب یک نظام جامع و یکپارچه به‌صورت شبانه‌روزی</p> <p>۲. مدیریت و کاهش عوارض ناشی از اشتباهات پرسنل پزشکی در طول ارائه خدمات</p> <p>۳. ارائه خدمات فوریت‌های پزشکی به مردم</p> <p>۴. تعامل بهتر با بیماران برای تصمیم‌گیری آگاهانه درمان</p> <p>۵. مدیریت حجم، تنوع و ناهمگونی اطلاعات پزشکی و درمانی بیماران</p> <p>۶. حمایت از استانداردهای تبادل اطلاعات پزشکی و بیمارستانی</p> <p>۷. توسعه و مدیریت بهتر دانش پزشکی و زیستی در بیمارستان از طریق ایجاد امکان به اشتراک‌گذاری منابع اطلاعاتی و دانش</p> <p>۸. مدیریت عملیات و فرآیندهای پیچیده بیمارستان</p>	<p>تمامی فرضیه‌ها پذیرفته شدند.</p>
استانداردها	<p>۱. به‌کارگیری استاندارد تبادل اطلاعات بیمارستانی HL7</p> <p>۲. کنترل رعایت الزامات و استانداردهای واحد فناوری اطلاعات بیمارستانی</p> <p>۳. دارا بودن استانداردهای بالینی در بیمارستان بر اساس شاخص‌های کیفیت ارائه خدمات</p> <p>۴. به‌کارگیری استاندارد جامع اصطلاحات بهداشتی و درمانی SNOMED</p> <p>۵. به‌کارگیری استاندارد LOINC برای ارسال نتایج الکترونیکی آزمایشگاه</p> <p>۶. به‌کارگیری تجهیزات تصویربرداری بیمارستان با خروجی دیجیتال و مطابق استاندارد DICOM</p> <p>۷. به‌کارگیری استاندارد کدگذاری ICD</p> <p>۸. دارا بودن استانداردهای بین‌المللی لوازم و تجهیزات پزشکی</p> <p>۹. اخذ گواهینامه ارزیابی عملکردی برای سیستم اطلاعات بیمارستانی</p> <p>۱۰. استفاده از واژه‌ها و تعاریف استاندارد در بیمارستان</p> <p>۱۱. اخذ گواهینامه‌هایی مطابق با استانداردهای سپاس (سامانه پرونده الکترونیک سلامت ایران) و استانداردهای اعتباربخشی بیمارستانی در ایران</p>	<p>تمامی فرضیه‌ها پذیرفته شدند.</p>

۳. لایه معماری سیستم‌های اطلاعاتی (داده): توصیف سرفصل‌های اطلاعاتی، مدل‌های منطقی داده‌ها و مدل‌های فیزیکی داده‌ها را بر عهده دارد.

۴. لایه معماری زیرساخت فناوری: مدل‌های مرجع فنی و استانداردهای فنی که باید در سطح سازمان رعایت شود را پوشش می‌دهد.

علاوه بر چهار لایه اصلی معماری سازمانی، نیازمندی‌های مقدماتی معماری برای مدل مفهومی در جدول ۲ نشان داده شده است.

معماری فناوری اطلاعات، سازمان‌هایی که دارای چندین لایه باشند را توصیف می‌کند [۲۳]، از طرفی مدل مفهومی معماری توگف، توسعه چهار لایه اصلی معماری سازمانی را به شرح زیر پوشش می‌دهد [۲۴]:

۱. لایه معماری کسب و کار: بالاترین سطح معماری سازمانی به حساب می‌آید. هدف این معماری، شناسایی و توصیف حوزه‌ها و خطوط مأموریتی و وظایف سازمانی است.

۲. لایه معماری سیستم‌های اطلاعاتی (کاربرد): فرآیندهای کاری، سیستم‌های اطلاعاتی، برنامه‌های کاربردی و روش‌های تعامل سیستم‌ها را در بر می‌گیرد.

جدول ۲: لایه‌های داخلی در مدل مفهومی معماری فناوری اطلاعات توگف

فرضیه‌های رد شده	فرضیه‌های مورد قبول با اطمینان ۹۵٪	لایه‌های داخلی	
۱. انتخاب و پیاده‌سازی ابزارهای پشتیبانی و سایر زیرساخت‌ها، جهت حمایت از عملکردهای معماری ۲. تعریف جایگاه معماری سازمانی توگف در بیمارستان ۳. تعریف دامنه و اولویت‌بندی کارهای معماری سازمانی توگف ۴. بیان مفاهیم، اصول و اهداف معماری سازمانی توگف ۵. اطمینان از تعهد و همکاری مستمر ذینفعان بر سر حوزه‌های معماری	۱. تعریف استراتژی‌های کسب و کار بیمارستان و بیان محدودیت‌ها ۲. شناسایی ذینفعان بیمارستان و پاسخ به نگرانی‌های آنها ۳. شناسایی و حوزه‌بندی عناصر تحت تاثیر بیمارستان، تعریف محدودیت‌ها و مفروضات ۴. ایجاد ساختار کنترل و فعالیت‌هایی جهت نظارت بر کنترل امور معماری ۵. اخذ تصویب رسمی و تعهد مدیریت برای انجام معماری توگف ۶. تعریف چارچوب توگف و روش‌های دقیق توسعه معماری سازمانی ۷. شناسایی نیازمندی‌های معماری سازمانی و اولویت‌های آنها شامل الزامات، فرضیات و شکاف‌ها برای پاسخ به نیازها و محدودیت‌ها ۸. اعتبارسنجی اصول کسب و کار، اهداف، پیشران‌ها/گردانندگان و شاخص‌های عملکرد کلیدی ۹. بررسی زمینه‌های گوناگون بیمارستان جهت انجام معماری سازمانی	نیازمندی‌های مقدماتی معماری	معماری کسب و کار
۱. افراد پشتیبان ۲. پژوهشگران	۱. پزشکان ۲. پرستاران ۳. بیماران ۴. مدیر بیمارستان ۵. کارکنان بخش‌های مختلف بیمارستان ۶. مسئول فناوری اطلاعات ۷. مسئول درمانگاه ۸. مسئول آزمایشگاه ۹. مسئول تصویربرداری ۱۰. مسئول داروخانه	شناسایی نقش‌ها و تعریف افراد درگیر در فعالیت‌های بیمارستان	معماری کسب و کار
فرآیندهای مربوط به واحدهای پشتیبانی و اداری (شامل: فرآیندهای بایگانی پرونده‌ها- فرآیند جمع‌آوری و ارسال آمار بیمارستان- فرآیند پرداخت حقوق پرسنلی-فرآیند نحوه محاسبه و ارسال صورت حساب	۱. فرآیندهای مربوط به واحدهای درمانی بیمارستان ۲. فرآیندهای مربوط به تامین سلامت ۳. فرآیندهای مربوط به ثبت و دسترسی به اطلاعات سلامت افراد ۴. فرآیندهای مربوط به مدیریت بیمار ۵. فرآیندهای مربوط به مدیریت بیمارستان ۶. فرآیند مربوط به مدیریت بیماری	شناسایی فرآیندهای کسب و کار بیمارستان	معماری کسب و کار

بیمه و نسخ سرپایی به سازمان‌های بیمه‌گر و ...)	۷. فرآیندهای مربوط به هشدارهای سلامت عمومی ۸. فرآیندهای مربوط به انتخاب پزشک بیمارستان		
۱. کمیته‌های مرکز (شامل: کمیته اعتباربخشی و حاکمیت بالینی، کمیته آمار و اسناد پزشکی، کمیته اخلاق پزشکی و ...) ۲. پژوهشکده‌ها و مراکز تحقیقات (شامل: مرکز تحقیقات خون و انکولوژی، پژوهشکده بیماری‌های گوارش و کبد، پژوهشکده علوم غدد و متابولیسم و ...)	۱. بخش‌های ویژه ۲. بخش‌های بالینی ۳. درمانگاه‌های موجود در بیمارستان ۴. بخش‌ها و واحدهای پاراکلینیک ۵. داروخانه‌های موجود در بیمارستان ۶. واحدهای ستادی ۷. واحدهای اداری و خدماتی	تعیین واحدهای مختلف بیمارستان	معماری کسب و کار
۱. سیستم تغذیه بیمارستان	۱. سیستم اطلاعات پذیرش ۲. سیستم اطلاعات مدارک پزشکی ۳. سیستم اطلاعات ترخیص ۴. سیستم اطلاعات بیماران سرپایی ۵. سیستم اطلاعات بخش‌های بیمارستانی ۶. سیستم اطلاعات اتاق عمل ۷. سیستم اطلاعات آزمایشگاه ۸. سیستم اطلاعات رادیولوژی ۹. سیستم اطلاعات داروخانه	سیستم‌های اطلاعات بیمارستانی	معماری سیستم‌های اطلاعاتی (کاربرد)
۱. سرویس امنیتی ۲. سیستم‌های پشتیبان تصمیم‌گیری ۳. سرویس ارتباطی ۴. سرویس واژه‌شناسی	۱. پرونده الکترونیکی پزشکی ۲. پرونده الکترونیکی پرستاری ۳. سیستم اطلاعات پرسنلی و زمان‌بندی ۴. سیستم اطلاعات مدیریت تخت ۵. سرویس دورپزشکی ۶. سیستم اطلاعات مدیریت منابع	سرویس‌های کاربردی	
۱. اطلاعات بیمارستان و مراکز درمانی (مثل مکان جغرافیایی، تجهیزات، سوابق و ...)	۱. پرونده بیماران ۲. پرونده پزشکان ۳. اطلاعات روش‌های درمان هر نوع بیماری ۴. اطلاعات علائم و نشانه‌های بیماری‌ها ۵. اطلاعات توصیفی پیرامون بیماری‌ها ۶. اطلاعات و منابع علمی سلامت	موجودیت‌های داده	معماری سیستم‌های اطلاعاتی (داده)
۱. توصیف موجودیت‌های شبکه‌ی داده ۲. توصیف تکنولوژیکی داده‌ها	۱. تعریف و توصیف انواع داده و اطلاعات بیمارستانی ۲. توصیف اطلاعات مهم فرآیندهای درمانی بیمارستان	تحلیل و گزارش‌دهی داده	
	۱. ثبت صحیح اطلاعات بیماران در مرحله پذیرش و در طول درمان ۲. ثبت داده‌ها در سیستم آمار و اطلاعات بیمارستانی (آواب) ۳. ثبت سابقه سلامت فرد در طول عمر در سیستم‌های اطلاعاتی	ذخیره‌سازی و بازیابی داده	
۱. وجود یک متد سریع و ساده برای ورود داده در سیستم‌های اطلاعات بیمارستانی	۱. امکان دسترسی سریع، آسان و درست به داده‌های سیستم‌های اطلاعات بیمارستان	دسترسی و یکپارچه‌سازی داده	
۱. زیرساخت شبکه ۲. زیرساخت سخت‌افزاری ۳. سیستم عامل ۴. زیرساخت ذخیره‌سازی ۵. زیرساخت ارتباطی (Communication Infrastructure) ۶. زیرساخت مخابراتی	۱. زیرساخت نرم‌افزاری (شامل: نرم‌افزارهای مالی و اداری بیمارستان؛ نرم‌افزارها و تجهیزات نرم‌افزاری لازم در بیمارستان مانند نرم‌افزارهای پایه، نرم‌افزارهای کمکی، نرم‌افزارهای کاربردی عمومی) ۲. زیرساخت واسط کاربرد (Interface Application Platform) (شامل: توصیف نرم‌افزارهای لازم برای ارتباط بین نرم‌افزارهای کاربردی و سرویس‌ها برای دسترسی به خدمات بیمارستانی) ۳. زیرساخت مدیریت محتوی	-	معماری زیرساخت فناوری

استفاده شد. برای سفارشی‌سازی آن، پرسشنامه ۱۳۴ سوالی طراحی شد که مولفه‌های مورد نیاز معماری یک بیمارستان در بخش‌های ۸ گانه معماری توگف را تعیین کرده و مورد سوال قرار می‌داد. سوالات این پرسشنامه برای بیمارستان اختصاصی سازی شدند. در نهایت معماری با ۱۰۴ عنصر بدست آمد که طبق نظر خبرگان برای بیمارستانهای دانشگاهی مناسب است.

تحلیل نتایج به دست آمده از داده‌های جمع‌آوری شده توسط آزمون تحلیل واریانس یک طرفه (آنووا) تمامی فرضیه‌ها را تایید کرد به عبارتی نشان داد که اختلاف معناداری بین حداقل دو فرضیه وجود دارد. نتایج نشان داد:

بخش ۱- یعنی چشم‌انداز بیمارستان شامل دو زیر حوزه چشم‌انداز، راه‌انداز (پیش‌ران) است. در بخش چشم‌انداز از ۷ فرضیه ۴ فرضیه تایید شدند. این چهار فرضیه بر بالاترین سطح خدمات، رضایتمندی و کاهش آسیب‌پذیری تاکید دارند. به نظر می‌رسد که سه مورد حذف شده یعنی درک بهتر از وضعیت جامعه و عدالت بیشتر با نظام سلامت و نه بیمارستان ارتباط دارد. ضمناً کاهش زمان اولین ویزیت در بخش اورژانس نیز در ۴ فرضیه اولیه قابل دستیابی است. در زیر بخش راه‌انداز ۶ فرضیه تایید و دو فرضیه رد شدند. ۲ فرضیه رد شد. در این بخش انتظارات مردم برای امنیت اطلاعات سلامت و افزایش کیفیت زندگی بطور مستقیم به عنوان راه‌انداز بیمارستان محسوب نمی‌شوند.

بخش ۲- بخش راهبردهای بیمارستان ۱۲ راهبرد برای بخش‌های مختلف تایید شدند. در این بخش بکارگیری فناوری اطلاعات و ارتباطات و تمام زیرسیستم‌های معماری سازمانی به عنوان راهبردهای سازمان پذیرفته شدند.

بخش ۳- قابلیت‌های بیمارستان ۸ فرضیه داشت که همه تایید شدند. این قابلیت‌ها در راستای چشم‌انداز و راهبرد هستند و برای ارائه خدمات با کیفیت ضروری هستند.

بخش ۴- استانداردهای بیمارستان: در این بخش ضرورت تمام ۱۱ فرضیه، به عبارتی تمام استانداردهای که مرتبط با بخش بیمارستانی هم اختصاصی شده بودند تایید شدند.

ملاحظه می‌شود معماری فناوری اطلاعات توگف طبق مدل مفهومی ارائه شده دارای ۴ ورودی و ۴ لایه داخلی جمعا ۸ بخش است. ضمناً هر لایه نیز مشتمل بر چند زیر لایه است. برای تعیین عوامل مورد نیاز در هر لایه از مولفه‌های بیان شده در مستندات معماری فناوری اطلاعات در داخل و خارج از ایران استفاده شد و پرسشنامه‌ای مشتمل بر ۱۳۴ سوال در مقیاس لیکرت و پنج گزینه‌ای تهیه شد که با احتساب نمراتی از ۱ تا ۵ برای گزینه‌ها طراحی شد. پس از جمع‌آوری و تدوین پرسشنامه و تایید روایی آن توسط ۵ نفر از اساتید دانشگاه، پرسشنامه برای آزمون اولیه بین ۲۰ نفر از مخاطبین توزیع شد و مقدار آلفای کرونباخ ۰,۹۱ تعیین شد. سپس پرسشنامه در حد وسیع بین افراد توزیع شد. در مرحله نهایی مقدار آلفای کرونباخ معادل ۰,۹۷ به دست آمد که نشان دهنده پایایی نتایج است. برای تحلیل نتایج به دست آمده از آزمون تحلیل واریانس یک طرفه (آنووا) و سپس آزمون دوجمله‌ای برای هر یک از فرضیه‌های موجود با نرم‌افزار SPSS استفاده شد. نتایج نشان داد در آزمون تمام فرضیه‌های پرسشنامه با آنووا توسط نرم‌افزار SPSS، تمامی فرضیه‌های بیمارستان دارای مقدار p.value برابر صفر است که مقداری کمتر از ۰,۰۱ می‌باشد. بنابراین فرض صفر مبنی بر تساوی میانگین‌ها رد شد و فرض یک با سطح معنی‌دار بودن ۹۹٪، تایید شد. یعنی می‌پذیریم هیچکدام از فرضیه‌های موجود در بعدهای مختلف بیمارستان، دارای ارزش و اثر یکسان نیست. همچنین با استفاده از آزمون دوجمله‌ای و تعیین مقدار p.value، تاثیر یا عدم تاثیر تک به تک فرضیه‌ها در هر بعد از مدل مفهومی چارچوب توگف بررسی شد.

بحث

هدف از این مقاله ارائه معماری فناوری اطلاعات بیمارستان بود. نتایج این تحقیق نشان داد علیرغم وجود معماری سازمانی برای سازمان‌های دولتی و بنگاه‌های تجاری و خدماتی، هیچکدام از آنها نیازمندی‌های بیمارستان را برآورده نمی‌کند. لذا طبق تحقیق انجام شده، از معماری فناوری اطلاعات TOGAF که قابلیت روز آمد سازی و قابلیت سفارشی‌سازی در حوزه بیمارستانی را دارد [۱۵]

نیستند و هزینه بر هم هستند. هارن [۱۹] نیز به این نکته اشاره کرده است که در معماری سیستم‌های اطلاعاتی ممکن است سیستم‌ها بطور پی در پی و یا به طور همزمان توسعه یابند. لذا می‌توان در مراحل بعدی در صورت نیاز این سیستم‌ها را توسعه داد.

بخش ۷- معماری سیستم‌های اطلاعاتی (داده) بیمارستان است. این بخش شامل ۴ زیر بخش است. ولد [۲۶] و وگر و دیگران [۲۷] نیز بر اهمیت این چهار بخش تاکید کرده اند. زیر بخش موجودیت داده بیمارستان ۷ فرضیه داشته که تنها یکی از آنها تایید نشد. اطلاعات بیمارستان و مراکز درمانی گزینه‌ای است که تایید نشده است. زیرا برای معماری سازمانی درون بیمارستان اولویت ندارد و می‌تواند بعداً در صورت نیاز توسعه یابد. زیربخش دوم، تحلیل و گزارش‌دهی داده است که شامل ۴ فرضیه است که دو فرضیه توصیف موجودیت‌های شبکه‌ی داده و توصیف تکنولوژیکی داده‌های استفاده شده در بیمارستان تایید نشده است. اولی به دلیل آنکه افراد سازمان با این تعاریف آشنا هستند و دومی بدلیل آنکه علاقه‌ای به بخش تکنولوژیکی آن ندارند. در زیربخش سوم ذخیره‌سازی و بازیابی داده همه فرضیه‌ها تایید شدند. اهمیت این مورد در ولد [۲۶] نیز مورد تایید قرار گرفته است. در زیربخش چهارم، دسترسی و یکپارچه‌سازی داده‌های سیستم‌های اطلاعات بیمارستان آسان و درست به داده‌های سیستم‌های اطلاعات بیمارستان " تایید و فرضیه دوم " وجود یک متد سریع و ساده برای ورود داده در سیستم‌های اطلاعات بیمارستانی " رد شده است. زیرا اولویت اول صحت اطلاعات بوده و نه سرعت.

بخش ۸- معماری زیرساخت فناوری بیمارستان است. این بخش شامل ۹ زیرساخت است که سه مورد آن مورد تایید قرار گرفت و ۶ مورد آن که به بخش پشتیبانی خدمات برمی‌گردد برون سپاری شده لذا در اولویت قرار نگرفت.

نتیجه‌گیری

در این تحقیق معماری فناوری اطلاعات توگف مشتمل بر ۸ لایه و ۱۰۴ جزء برای بیمارستان دانشگاهی مطابق زیست بوم ایران و شرایط محیطی آن بومی سازی شد. چارچوب حاضر برای یک بیمارستان دانشگاهی با محدودیت‌های قانونی کشور ایران که در آن ارائه خدمات و پشتیبانی از آن

بخش ۵- معماری کسب و کار بیمارستان شامل ۴ حوزه است. در زیر حوزه اول ۱۴ فعالیت در قالب فرضیه‌ها بیان شده بودند که ۹ فرضیه قبول و ۵ فرضیه رد شدند. نتایج مصاحبه با ذینفعان برای تحلیل این نتایج نشان داد از آنجا که در ایران قوانین نانوشته در اذهان مدیران جای دارد و مدیریت ارشد بیمارستان خود را متعهد به حمایت از کار و خرید بهترین تجهیزات می‌داند. لذا فرضیه‌های حذف شده با وجود اهمیت آنها جزو فعالیت‌های اصلی محسوب نمی‌شوند. به نظر می‌رسد در موارد مشابه در سایر کشورها هم اگر این تعهد وجود داشته باشد نتایج قابل تعمیم باشد. در زیر حوزه دوم نقش افراد پشتیبان و پژوهشگران مورد تایید قرار نگرفت و بقیه تایید شدند. یعنی از ۱۲ گروه ذینفع، دو گروه تایید نشدند. موارد حذف شده به این دلیل توجیه می‌شود که بیمارستان شریعتی برخی خدماتش را برون سپاری کرده و نیاز به استخدام نیروی پشتیبانی ندارد. ضمناً چون بیمارستان زیرمجموعه دانشگاه است و پزشکان و پرسنل خودشان پژوهشگر هستند لذا ظرفیت برای سایر پژوهشگران ندارد. در زیر حوزه سوم از ۹ فرآیند، ۸ فرآیند تایید و یک فرآیند رد شد. رد شدن فرآیند پشتیبانی به دلیل برون‌سپاری خدمات قابل توجیه است. در زیر حوزه چهارم از ۹ بخش در واحد بیمارستانی، ضرورت وجود ۷ مورد تایید شد. عدم تایید "کمیته‌های مرکز" و "پژوهشکده‌ها و مراکز تحقیقات" به آن دلیل بوده است که این بخش‌ها زیر نظر دانشگاه هستند و نه بیمارستان.

بخش ۶- معماری سیستم‌های اطلاعاتی (کاربردی) بیمارستان شامل دو زیر بخش سیستم‌های اطلاعات بیمارستانی و سرویس‌های کاربردی سیستم اطلاعات بیمارستان است. زیر بخش اول یعنی سیستم‌های اطلاعات بیمارستانی شامل ۱۰ فرضیه بود که ۹ مورد آن تایید شد و تنها سیستم تغذیه بیمارستانی به دلیل برون‌سپاری خدمات مورد تایید واقع نشد. در مقاله ولد [۲۵] بر اهمیت این سرویس‌ها تاکید شده است. در زمینه سرویس‌های کاربردی ۱۰ سرویس در قالب سوالات شناسایی شده بودند که ۶ سرویس آن تایید شدند. سرویس‌های امنیتی، سیستم‌های پشتیبان تصمیم‌گیری، سرویس ارتباطی و سرویس واژه‌شناسی رد شدند. زیرا این بخش‌ها از اولویت برخوردار

۸. Levis, A. H. and Wagenhals, L. W., "C⁴ISR architectures: I. Developing a process for C⁴ISR architecture design," Syst. Eng., ۲۰۰۰, Vol. ۳, No. ۴, pp. ۲۲۵-۲۴۷.

۹. Fradinho, Jorge Miguel dos Santos. Towards high performing hospital enterprise architectures: elevating hospitals to lean enterprise thinking. Diss. Massachusetts Institute of Technology, ۲۰۱۱.

۱۰. Akiyama, M. "Migration of the Japanese healthcare enterprise from a financial to integrated management: strategy and architecture," Stud. Health Technol. Inform., ۲۰۰۱, No. ۱, pp. ۷۱۵-۷۱۸.

۱۱. Lu, X.; Duan, H., Li, H., Zhao, C., and An, J., "The architecture of enterprise hospital information system," in Engineering in Medicine and Biology Society, ۲۰۰۵. IEEE-EMBS ۲۰۰۵. ۲۷th Annual International Conference of the, ۲۰۰۶, pp. ۶۹۵۷-۶۹۶۰.

۱۲. Ahsan, Kamran, Hanifa Shah, and Paul Kingston. "Healthcare Modelling through Enterprise Architecture: A Hospital Case." Information Technology: New Generations (ITNG), ۲۰۱۰. Seventh International Conference on. IEEE, ۲۰۱۰, pp. ۴۶۰-۴۶۵.

۱۳. Rijo, R., Martinho, R., and Ermida, D., "Developing an Enterprise Architecture Proof of Concept in a Portuguese Hospital," Procedia Comput. Sci., ۲۰۱۵, Vol. ۶۴, pp. ۱۲۱۷-۱۲۲۵.

۱۴. Walsh, M. T., Dublin, B. A., Ishigami, E. M., and Shebaro, I. A., "The architecture of a shared leadership model for health systems strengthening initiatives led by a US-based academic hospital," Ann. Glob. Health, ۲۰۱۵, Vol. ۸۱, No. ۱, p. ۱۸۶.

۱۵. Haghithathoseini, A., Boubarsad, H., Saghafi, F. "Identifying the best organizational architecture framework for hospitals for implementation in Iran," Journal of Medical Council of Islamic

باید برون‌سپاری شوند نوشته شده است و می تواند برای موارد مشابه در ایران و سایر کشورهای در حال توسعه ای که شرایط مشابه دارند، مفید باشد. در تحقیقات آتی می توان معماری فناوری اطلاعات را برای بیمارستان های خصوصی با گسترش مولفه های مالی و سودآوری سفارشی سازی کرد. همچنین می توان این چارچوب را برای بیمارستان های دولتی بدون محدودیت برون‌سپاری خدمات به روزرسانی کرد. ضمناً می توان این چارچوب را برای بیمارستان های غیر دانشگاهی سفارشی سازی نمود.

سیاسگزاری

از آقای مهندس مهدی زرگرنجاج بخاطر نظرات راهگشا در طول انجام این پژوهش سپاسگزاری می شود.

منابع

۱. Minoli, D., Enterprise architecture A to Z: frameworks, business process modeling, SOA, and infrastructure technology. CRC Press, ۲۰۰۸.

۲. Spewak, S. H. and Hill, S. C., Enterprise Architecture Planning: Developing a Blueprint for Data, Applications and Technology. Wellesley, MA, USA: QED Information Sciences, Inc., ۱۹۹۳.

۳. Inmon, W. H., Zachman, J. A. and Geiger, J. G.; Data Stores, Data Warehousing and the Zachman Framework: Managing Enterprise Knowledge, ۱st ed. New York, NY, USA: McGraw-Hill, Inc., ۱۹۹۷.

۴. Zachman, J. A., "A framework for information systems architecture," IBM Syst. J., ۱۹۸۷, Vol. ۲۶, No. ۳, pp. ۲۷۶-۲۹۲.

۵. Sowa and J. F. Zachman, J. A., "Extending and formalizing the framework for information systems architecture," IBM Syst. J., ۱۹۹۲, Vol. ۳۱, No. ۳, pp. ۵۹۰-۶۱۶.

۶. Ji, W. and Xia, A., "Federal enterprise architecture framework," Comput. Integr. Manuf. Syst.-BEIJING-, ۲۰۰۷, Vol. ۱۳, No. ۱, p. ۵۷.

۷. Blevins, T. J., Spencer, J., and Waskiewicz, F., "TOGAF ADM and MDA," Open Group OMG, ۲۰۰۴.

- New York: McGraw-Hill Medical, ۲۰۱۲.
۲۲. U. S. N. and W. Report, Best Hospitals ۲۰۱۵. U.S. News and World Report, ۲۰۱۴.
۲۳. Jonkers H, van Burren R, Arbab F, De Boer F, Bonsangue M, Bosma H, Ter Doest H, Groenewegen L, Scholten JG, Hoppenbrouwers S, Iacob ME. Towards a language for coherent enterprise architecture descriptions. In Enterprise Distributed Object Computing Conference, ۲۰۰۳. Proceedings. Seventh IEEE International ۲۰۰۳ Sep ۱۶ (pp. ۲۸-۳۷). IEEE.
۲۴. R. Harrison, TOGAF® ۹ Foundation Study Guide. Van Haren, ۲۰۱۳.
۲۵. Van de Velde, R., Hospital Information Systems - The Next Generation, Softcover reprint of the original ۱st ed. ۱۹۹۲ edition. Place of publication not identified: Springer, ۲۰۱۳.
۲۶. Van. de Velde, R. and Degoulet, P., Clinical Information Systems: A Component-Based Approach, ۲۰۰۳ edition. New York: Springer, ۲۰۰۳.
۲۷. Wager, K. A., Lee, F. W., and Glaser, J. P., Health Care Information Systems: A Practical Approach for Health Care Management, ۲ edition. Jossey-Bass, ۲۰۰۹.
- Republic of Iran, ۲۰۱۶, Vol. ۳۴, No. ۱, pp. xx. under publication, [in Persian]
۱۶. Tang A, Han J, Chen P. A comparative analysis of architecture frameworks. In Software Engineering Conference, ۲۰۰۴. ۱۱th Asia-Pacific ۲۰۰۴ Nov ۳۰ (pp. ۶۴۰-۶۴۷). IEEE.
۱۷. Lim N, Lee TG, Park SG. A comparative analysis of enterprise architecture frameworks based on EA quality attributes. In Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, ۲۰۰۹. SNPD'۰۹. ۱۰th ACIS International Conference on ۲۰۰۹ May ۲۷ (pp. ۲۸۳-۲۸۸). IEEE.
۱۸. A. Ishizaka and A. Labib, "Analytic Hierarchy Process and Expert Choice: Benefits and limitations," *Insight*, ۲۰۰۹, Vol. ۲۲, No. ۴, pp. ۲۰۱-۲۲۰.
۱۹. Haren, V., TOGAF Version ۹, ۱, ۱۰th ed. Van Haren Publishing, ۲۰۱۱.
۲۰. Weisman, R., "An Overview of TOGAF® Version ۹, ۱," ۲۰۱۱.
۲۱. McKean, S.C., Ross, J.J., Dressler, D.D., Brotman, J.J. and Ginsberg, J.S., Principles and Practice of Hospital Medicine, ۱ edition.

ارائه مدلی جهت استفاده از عاملهای متحرک در سیستم‌های تشخیص نفوذ توزیع شده مبتنی بر تئوری بازی

* امین نظارات
** مهدی رجا
*** غلامحسین دستغیبی‌فرد

* استادیار، گروه کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، ایران
** بخش مهندسی و علوم کامپیوتر، دانشگاه شیراز
*** دانشیار، بخش مهندسی و علوم کامپیوتر، دانشگاه شیراز

تاریخ دریافت: ۹۲/۰۸/۰۱ تاریخ پذیرش: ۹۵/۰۸/۲۳

چکیده

سیستم‌های تشخیص نفوذ در شبکه، ابزارهایی هستند که به منظور محافظت از منابع شبکه در مقابل حملات استفاده می‌شوند. با توجه به گستردگی حملات در فضای اینترنت و تغییر در شکل و نوع حملات از حالت متمرکز به توزیع شده، معماری اینگونه سیستم‌ها نیز به سمت توزیع‌شدگی حرکت می‌کند. در این مقاله روشی مبتنی بر عاملهای متحرک که به عنوان سنسورهای تشخیص‌دهنده حرکات غیر معتبر عمل می‌کنند پیشنهاد شده است. عاملهای متحرک تشخیص‌دهنده حمله^۱ به صورت پراکنده در شبکه در حال جابجایی از یک گره به گره دیگر می‌باشند و در هر زمان یک شبکه فوقانی^۲ امنیتی را ساخته و با استفاده از نوعی بازی همکارانه و برقراری ارتباط با یکدیگر، پس از رسیدن به مقدار شیلی می‌توانند میزان و منشاء حمله را تشخیص و گزارش دهند. در این مقاله روشی پیشنهاد شده است که WGA در یک بازی غیرهمکارانه با عنصر مهاجم سعی در برقراری یک ارتباط مکاشفای جهت محاسبه مقدار نش و رسیدن به حداکثر سودمندی را دارد تا بتواند ضمن تفکیک حملات و یا درخواستهای واقعی، میزان و شدت حمله را با کمک سایر WGA بدست آورد.

واژه‌های کلیدی: عاملهای متحرک، IDS، تئوری بازی، سیستم چند عاملی، تعادل نش، مقدار شیلی، امنیت شبکه

^۱ White Globule Agents
^۲ Overlay Network

۱- مقدمه

حملات و آسیب‌پذیری شبکه‌های کامپیوتری روز به روز چه از بعد پیچیدگی و چه تکنولوژی در حال گسترش می‌باشند. تا جایی که بعضی از این حملات به راحتی شرکت‌های تجارت الکترونیکی را از گردونه تجارت خارج می‌کنند. سیستم‌های تشخیص نفوذ می‌بایستی بتوانند در کاهش آسیب‌ها و حملات به شبکه‌ها پیشنهادات مناسبی را ارائه نمایند. برای رسیدن به این هدف سازمانها و موسسات تجاری و دولتی اقدام به تاسیس مراکز عملیات امنیت جهت آنالیز و رصد گزارشات و حوادث نفوذ نموده‌اند.

به عنوان مثال یکی از ابزارهای امنیتی شرکت Symantec با عنوان Threat con warning Sys با اندازه‌گیری میزان آسیب‌پذیری شبکه درجه آنرا به مدیر شبکه گزارش می‌دهد [۱] و مدیر شبکه می‌تواند براساس آسیب‌پذیری، تمهیدات لازم را در نظر بگیرد. البته این سیستم از یک ساختار اندازه‌گیری دقیق استفاده می‌کند و معیارهای آن مبتنی بر روشهای ریاضی و محاسباتی نیست، که این امر می‌تواند مدیر شبکه را به اشتباه اندازد. علاوه بر این اندازه‌گیری مربوط مبتنی بر پایش داده‌های رد و بدل شده بین حمله‌کننده و دفاع‌کننده نیست. سیستم‌های تشخیص نفوذ اخیر را می‌توان به دو گروه تقسیم بندی کرد [۲]:

۱- واکنشی (Reactive) (مبتنی بر تشخیص تمضا یا مجوز جهت اجازه ورود یا خروج)

۲- کنشی (Proactive) (سرویسهای فوقانی امن، شبکه proxy و ...)

سرویسهای فوقانی امن^۳ دارای معماری هستند که برای جلوگیری از حملات و ایجاد آمادگی قبلی در مقابل حملات توزیع شده DOS استفاده می‌شود. اما روشی که در این معماری برای برقراری ارتباط با گره‌ها استفاده می‌شود مبتنی بر ارتباط دائمی است که سربار زیادی را بر روی شبکه تحمیل می‌نماید و این موضوع مانعی است در

مقابل گسترش بیشتر معماری [۳]. اگرچه روش کنشی بسیار کارتر از مدل واکنشی است اما هنوز مسائل زیادی در پیاده‌سازی کارایی در این معماری باقی مانده است. سیستمهای IDS متمرکز بسیار مستعد خطای تک نقطه^۴ هستند و می‌توانند توسط مهاجمین کشف و مورد تهاجم قرار گیرند. برای حل این مشکل می‌توان از تعداد بیشتری IDS استفاده کرد تا بتوان میزان حملات تشخیص داده نشده را کاهش داد. که البته این مسئله مستلزم پرداخت هزینه بیشتر می‌باشد.

قدرت یک سیستم تشخیص نفوذ در ایجاد یک تعادل بین تعداد مدافعان و میزان تشخیص‌های خطا یا عدم تشخیص‌ها می‌باشد. به همین منظور جهت مدل کردن حملات برای اندازه‌گیری درجه ریسک امنیتی (تهدید یک حمله) و کمک به تصمیم‌گیری در پاسخ مناسب به حملات ضروری به نظر می‌رسد که از یک روش توزیع شده تشخیص نفوذ بهره گرفته شود. عاملهای متحرک می‌توانند به عنوان مدافعانی در گره‌های شبکه در نظر گرفته شوند که به صورت خود مختار بین گره‌ها جابجا شوند و مجموعه‌ای از آنها تشکیل یک شبکه فوقانی از یک سیستم چند عامله را می‌دهند. هر عامل متحرک می‌تواند به منظور ایجاد درک بهتر از یک حمله امنیتی و تشخیص میزان خطر آن وارد یک تعامل با حمله‌کننده شود.

به منظور افزایش دقت در تصمیم‌گیری و کشف حمله می‌توان از تئوری بازیها در تعامل عامل تشخیص‌دهنده و مهاجم استفاده کرد. سپس هر یک عاملهای تشخیص‌دهنده می‌توانند در یک بازی دیگر به منظور بررسی میزان دقت نتیجه حاصله وارد یک مذاکره^۵ با سایر عاملهای درون شبکه فوقانی شوند. این بازی از نوع بازیهای همکارانه جهت رسیدن به یک توافق در خصوص میزان درجه ریسک امنیتی^۶ حمله مربوطه می‌باشد.

در بازی اول که بین عامل متحرک WGA به عنوان مدافع و یک مهاجم صورت می‌گیرد، یک تعامل دوطرفه

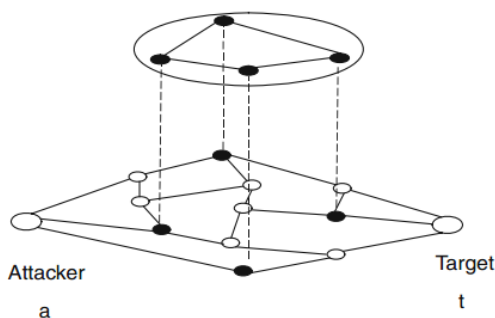
^۴ Single Point of Failure

^۵ Negotiate

^۶ Security Risk value

^۳ Secure Overlay Network

هر یک از این گروه‌ها میزبان یک عامل سیار است. این عاملها بایستی در هر لحظه به نحوی در گره‌های شبکه قرار گیرند که حداکثر پوشش امنیتی برقرار باشد و با توجه به سیار بودن آنها، حیطة حرکت عامل محافظ می-بایست به نحوی باشد که تجمع عاملها و یا خلا در شبکه بوجود نیاید. به همین دلیل و به منظور کاهش نرخ تشخیص خطا، افزایش نرخ تشخیص و کاهش مصرف پهنای باند شبکه، بایستی عواملی از جمله تعداد عاملهای محافظ، محیط تحت پوشش هر یک و درجه حساسیت امنیتی آنها مشخص و در طی زمان اصلاح و اعلام گردد. در این تحقیق به منظور افزایش مراقبت از شبکه و مقابله با مهاجمان معماری جهت شبکه فوقانی و مدیریت حرکت عاملها و همچنین بازی امنیتی جهت رسیدن به درجه حساسیت عامل (مقدار شیپلی) مورد بررسی و پیشنهاد قرار گرفته شده است [۴].



شکل ۱: شبکه فوقانی

۳- معماری مدل پیشنهادی سیستم چند عاملی محافظ

نظر به گستردگی شبکه‌های کامپیوتری و افزایش حجم مبادلات و ارتباطات درون شبکه‌ها و افزایش بار آنها، ایجاد ارتباطات دائمی و نگهداری آنها می‌تواند باعث افزایش سربار شبکه و ایجاد اختلال در کار اصلی نرم‌افزارهای موسسات گردد. سیستم‌های امنیتی و نظارتی بایستی بتوانند با کمترین اختلال در شبکه و حداقل سربارکاری به وظایف خود عمل نمایند.

یکی از انواع عاملها، نوع سیار آنها می‌باشد. یک عامل سیار، نوعی نرم‌افزار خود مختار، کنشی و متحرک می‌باشد که می‌تواند با جابجایی درون شبکه و بدون نیاز به کنترل مستمر توسط سرور اصلی، اقدام به انجام وظایف خود اعم از جمع‌آوری اطلاعات، مقابله با حملات، خرید و فروش یا مذاکره با سایر عاملها نماید و سپس نتایج حاصله را به

برقرار می‌شود که یک بازی از نوع غیرهمکارانه دونفره می‌باشد. هر دوطرف بازی به دنبال بیشینه کردن منافع و سودمندی خود می‌باشند. حمله‌کننده قصد نفوذ به شبکه و تخریب یا جمع‌آوری اطلاعات را دارد و مدافع نیز قصد مقابله و کشف تعدا بیشتری حملات به شبکه را دارد. هر دو بازیکن دارای یک تابع سودمندی^۷ هستند که در یک نقطه به تعادل کنش خواهند رسید که در آن نقطه هر دو بازیکن، نوعی از بازی خودرا که در آن تمایل به تغییر بازی ندارند را انجام می‌دهند. در این نقطه عامل متحرک مدافع به مقداری خواهد رسید که آن مقدار درجه امنیتی حمله را مشخص می‌نماید.

برای اطمینان از اینکه حمله‌ای صورت گرفته است یا خیر و جلوگیری از ارسال هشدارهای اشتباه در بازی دیگری از نوع همکارانه با سایر عاملهای مدافع در محیط چند عاملی اقدام به مذاکره می‌نماید. در این بازی عاملها می‌توانند جهت جلوگیری از ایجاد سربار در شبکه به صورت گروههای چند نفره تشکیل ائتلاف دهند. در هر ائتلافکه با موضوع یک حمله تشکیل می‌شود عاملها پس از مذاکره به یک مقدار تفاهمی مشترک به نام مقدار شیپلی^۸ می‌رسند که این مقدار درجه نهایی امنیت شبکه و نوع تصمیم در مقابل حمله صورت گرفته را مشخص می‌نماید. با توجه به عدد بدست آمده می‌توان شدت و میزان برخورد را مشخص نمود.

۲- شبکه فوقانی چند عامله^۹

نوع توزیع عاملهای متحرک و نحوه قرارگیری آنها در هر زمان برای تحت پوشش قرار دادن بخش عمده‌ای از شبکه، یکی از مسائل مهم در تشکیل شبکه فوقانی و از عوامل تاثیرگذار در تصمیم هر عامل برای جابجایی در شبکه است.

به عنوان مثال: در شکل ۱ فرض شده که یک مهاجم قصد ارسال یک بسته از میزبان α (مهاجم) به میزبان t (مقصد) را دارد و یک شبکه فوقانی چند عاملی از گره‌های $MA = \{MA_1, MA_2, \dots, MA_n\}$ تشکیل شده است.

^۷ Utility function

^۸ Shaply value

^۹ Multi Agent overlay network

اطلاع سرور فرستنده عامل برساند بدون اینکه نیاز به بازگشت به سرور و اشغال پهنای باند داشته باشد و یا حتی نیاز به ارسال کلیه اطلاعات داشته باشد [۵].

امروزه کاربردهای فراوانی از سیستمهای مبتنی بر عاملها وجود دارد. عامل ها می‌توانند با تشکیل اجتماعات چند عاملی و برقراری ارتباطات هم‌افزا با یکدیگر به نتایج بسیار سریع و با کمترین دخالت انسانی برسند. عاملها همچنین می‌توانند یاد گرفته و تغییر شکل دهند و استراتژی اولیه را بهبود بخشند. در این مقاله از خصوصیت مذاکره عاملها استفاده شده و معماری پیشنهادی ارائه شده است.

به منظور جلوگیری از بزرگ شدن عاملها و درگیر شدن آنها در بررسی بسته‌های شبکه‌های کامپیوتری در این تحقیق از یکی از IDS های شناخته شده و Open source به نام SNORT استفاده شده است [۶].

SNORT یک IDS خوب و پرکاربرد می‌باشد که با بررسی بسته‌های شبکه و مطابقت آنها با سناریوهای حمله موجود در بانک دانش خود می‌تواند حملات را تشخیص دهد و پیام هشدار مناسب را ارائه نماید. اما همانگونه که قبلاً نیز اشاره شد این سیستم یک IDS مرکزی است و برای اینکه بتواند مجموعه‌ای از حملات به گره‌های مختلف شبکه را تشخیص دهد می‌بایستی در نقاط مختلف شبکه نصب شده و همچنین امکان برقراری ارتباط با یکدیگر را نیز داشته باشند. در حال حاضر این سیستم از چنین قابلیت‌هایی برخوردار نمی‌باشد و برای تحقق این امر بایستی تمامی log های حملات به سایر IDS مرکزی منتقل و ادغام کرده تا بتوان کل حمله را به درستی تشخیص داد. که در حال حاضر چنین امکانی وجود ندارد.

در مدل پیشنهادی که ترکیبی است از عاملهای متحرک، از یک سرور مرکزی به منظور تولید عاملها استفاده می‌شود. در اینجا از مدل دفاعی بدون الهام گرفته شده است. تعدادی عامل متحرک به عنوان گلوبولهای سفید در گره‌های شبکه مستقر شده و در زمانهای خاصی جابجا می‌شوند. ترکیب سرور شامل اجزاء زیر است:

- کارخانه تولید عاملهای گلوبولهای سفید^{۱۰}

- کارخانه تولید عاملهای خنثی کننده^{۱۱}

- SNORT

- پایگاه دانش حملات^{۱۲}

- کارخانه تولید عاملهای گلوبول سفید

این عامل وظیفه تولید عاملهای جدید مدافع را برعهده دارد. در ابتدای شروع به کار سیستم تعدادی محل WGAهایی که بایستی تولید شوند از قبل توسط مدیر شبکه تعیین می‌شوند و این عامل پس از تولید هر WGA آنها به گره‌ای از شبکه ارسال می‌کند. نوع حرکت عامل WGA به گره‌های شبکه از نوع جابجایی‌پذیری قوی^{۱۳} می‌باشد و در هر زمان وضعیت کلیه وقایع قبلی و نتیجه حاصله از آنها را به همراه دارد. هر عامل WGA به صورت دوره‌ای با این عامل ارتباط برقرار می‌کند و گزارشی از مشاهدات خود را اعلام می‌نماید. در هر لحظه این عامل می‌داند که هر WGA کجاست. تفاوت WGAهای تولید شده در این عامل این است که یک عامل WGA به همراه خود خلاصه‌ای از سناریوهای انواع حملات در شبکه‌ها را دارد و با کمک ابزار تشخیص نفوذ

SNORT، رفتارهای مشکوک در شبکه را شناسایی می‌کند. پس از شناسایی یک رفتار مشکوک با سایر عاملهای WGA مذاکره کرده و دانش خود را تکمیل یا تصحیح می‌نماید. این عامل از چندین عامل دیگر تشکیل شده است که هر یک وظیفه‌ای را بر عهده دارند:

عامل AMS^{۱۴}: مدیر اصلی عاملهاست و به هر عامل یک شماره اختصاصی انتساب می‌دهد. وظیفه ایجاد، حذف، جابجایی و ... عاملها را بر عهده دارد.

عامل DF^{۱۵}: این عامل اطلاعات کاملی از کلیه عاملهای ایجاد شده و در حال اجرا را در اختیار دارد. نام عامل، محل عامل، مشخصات و تاریخچه آن را می‌داند. زمانی که یکی از عاملهای WGA قصد برقراری ارتباط با سایر

^{۱۱} Lymphocit Agent Factory

^{۱۲} IDS knowledge

^{۱۳} Strong mobility

^{۱۴} Agent Management System

^{۱۵} Directory Facilitator

^{۱۰} White Global Agent Factory

شبکه امکان اعمال قوانینی را به کاربر می‌دهد. در این ابزار از تعدادی قانون جهت شناسایی حملات استفاده می‌شود و زمانی که سناریوی موجود در قوانین محقق می‌شود عکس‌العمل مناسب را نشان می‌دهد. در این مقاله از SNORT جهت مطابقت سناریو با بسته‌های شبکه استفاده شده است.

پایگاه دانش حملات: در این پایگاه دانش سناریوهای مربوط به حملات امنیتی در شبکه درج شده است تا هر عامل WGA بدانند که در زمان بروز یک تهدید امنیتی (که با خواندن Log فایل SNORT قابل تشخیص است) چه اقدامی را انجام دهد. از آنجا که یک عامل WGA بایستی در زمانی که یک تهدید را حس می‌کند اقدام به تعامل با حمله‌کننده نماید، با استفاده از این دانش درون این پایگاه می‌داند که در مرحله بعدی چه عملی را بایستی انجام دهد تا حمله‌کننده قسمت بعدی حمله را به اجرا گذارد، لذا با شبیه‌سازی یک آسیب‌پذیری می‌تواند تهدید را تشخیص دهد.

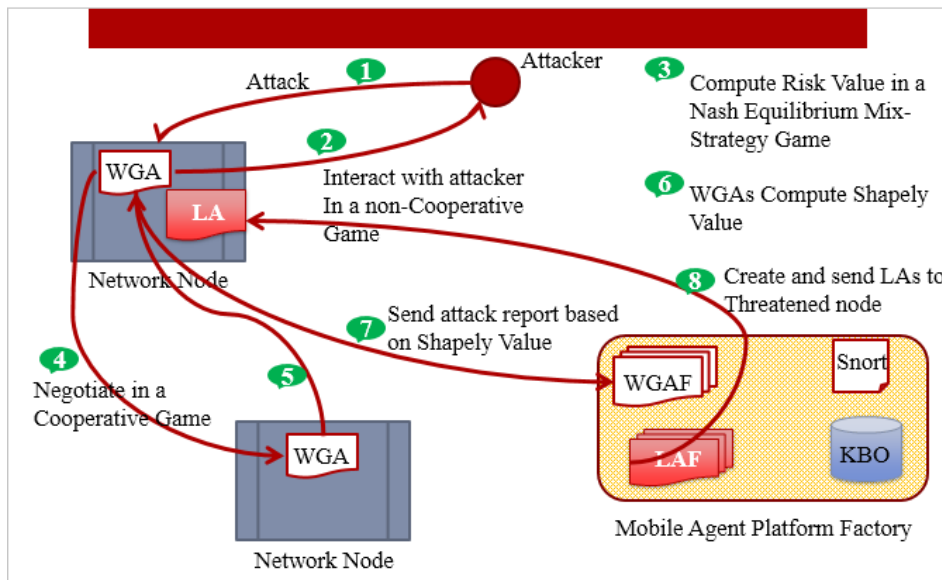
عاملها را داشته باشد، درخواست خود را به این عامل ارسال می‌نماید.

- کارخانه تولید عاملهای خنثی‌کننده (LAF)

یکی از وظایف سیستم‌های تشخیص نفوذ، جمع‌آوری اطلاعات و شناسایی حملات جهت گزارش‌دهی به سیستم‌های مقابله با حمله یا IDSها می‌باشد. در این مدل نوعی عامل در نظر گرفته شده است که وظیفه مقابله با حملات و از بین بردن حمله‌کننده را بر عهده دارد. پس از شناسایی و تایید یک حمله توسط ائتلافی از WGA گزارشی به WGAF ارسال می‌شود که درجه ریسک امنیتی این حمله را مشخص می‌نماید.

WGAF نیز با استفاده از درجه امنیتی، تقاضایی را برای این عامل ارسال می‌نماید و عامل مذکور بر همین اساس تعداد عاملهای LA و محلهایی که بایستی ارسال شوند را تشخیص می‌دهد.

- SNORT: یک ابزار متن باز IDS می‌باشد که با بررسی بسته‌های رد و بدل شده در شبکه و Sniff کردن



شکل ۲: مدل پیشنهادی سیستم توزیع شده تشخیص نفوذ

۴- روش کار مدل پیشنهادی

همانگونه که اشاره شد این مدل امنیتی از اجزاء مختلفی تشکیل شده است. در این بخش به نحوه کار مدل و روش اجرای آن پرداخته می‌شود. عامل WGAF تعدادی عامل WGA را ایجاد کرده و به گره‌های شبکه ارسال می‌کند. در صورت حمله یک عامل خارجی، WGA مستقر در آن گره شروع به تعامل با attacker می‌نماید. این عامل از نوع یک بازی غیر همکارانه می‌باشد. عامل WGA با کمک KB و اطلاعات حاصل شده از تعامل با مهاجم یک استراتژی آمیخته را پیش می‌گیرد تا بتواند به یک نقطه تعادلی نش برسد. این نقطه تعادلی نش به عنوان مقدار ریسک آن حمله در نظر گرفته می‌شود.

برای افزایش دقت مقدار ریسک و تشخیص این موضوع که آیا مشابه با این حمله در سایر گره‌های شبکه نیز شناسایی شده است یا خیر، عامل WGA اقدام به تشکیل یک ائتلاف با تعدادی از عامل‌های WGA دیگر می‌کند و در بازی همکارانه با آنها مذاکره می‌نماید تا در نهایت از مقادیر ریسک امنیتی بدست آمده در سایر WGAها به یک مقدار مورد توافق به نام مقدار شیپلی برسند. پس از محاسبه مقدار شیپلی، عامل WGA این مقدار را برای WGAF ارسال می‌نماید. این عدد به عنوان مقدار ریسک امنیتی شبکه که مورد توافق ائتلافی از WGAها می‌باشد در نظر گرفته شده و WGAF آن را به LAF گزارش می‌دهد و همچنین می‌تواند با توجه به این درجه امنیتی تصمیم به افزایش یا کاهش WGAها و یا تغییر مکان آنها بگیرد.

عامل LAF با دریافت درجه امنیتی جدید شبکه عامل-های خنثی‌کننده LA را ساخته و به گره‌های تهدید شده ارسال می‌کند. به هر گره تهدید شده یک عامل LA ارسال می‌شود و به وی تعداد تکثیرهای بعدی از خود را اعلام می‌کند. برای جلوگیری از اشغال پهنای باند شبکه، هر عامل LA پس از رسیدن به مقصد، خود را به میزان لازم تکثیر می‌کند تا بتواند با توجه به درجه امنیتی گزارش شده مقابله مناسب با حمله را داشته باشد.

با توجه به درجه امنیتی جدید بدست آمده برای شبکه، تعداد و محل قرارگیری WGAها توسط WGAF مورد بازنگری قرار گرفته و تعداد و محل‌های جدید اعلام می‌شود و در نهایت یک شبکه فوقانی از عامل‌های سیار و یک

محیط چند عاملی جدید ساخته می‌شود. ملاک‌های این تصمیم‌گیری بر مبنای پارامترهای: گره‌های تهدید شده، درجه تهدید هر گره، مسیر طی شده توسط یک مهاجم، تغییر درجه امنیتی نسبت به قبل و ... می‌باشد.

۵- بازی ریسک امنیتی

در حالت کلی، ریسک امنیتی به تعداد نفوذهایی که احتمال رخداد آنها وجود دارد گفته می‌شود (نرخ نقض امنیت). نفوذ امنیتی زمانی رخ می‌دهد که یک مهاجم بتواند از نواقص امنیتی استفاده کرده و وارد سیستم شود [۷] [۸]:

$$\text{Security risk} = \text{امنیت نقض تعداد نرخ}$$

$$\text{نفوذ هر هزینه میانگین} *$$

در این مقاله هدف تعیین میزان ریسک امنیتی است تا بتوان با استفاده از آن سیستم تشخیص نفوذ توزیع شده^{۱۶} را تنظیم کرد. در بازی امنیتی پیشنهادی نقش‌های زیر وجود دارد.

- مهاجم

- WGAgent

مهاجم: بطور دائم یک شبکه به کاربران مختلفی در حال سرویس‌دهی می‌باشد که بعضی از آنها می‌توانند به عنوان یک مهاجم باشند. هدف یک مهاجم نفوذ به درون شبکه و اجرای اهداف از قبل تعیین شده می‌باشد [۹]. اگر حمله یک مهاجم تشخیص داده نشود معادل b_1 هزینه پرداختی آن مهاجم می‌باشد. در صورت تشخیص حمله این هزینه معادل b_2 می‌باشد. همچنین اگر فرض می‌شود که $b_1 < b_2$ است یعنی مهاجم در صورتی که تشخیص داده شود یک سودمندی مثبت بدست می‌آورد و در صورتی که $b_1 > b_2$ باشد، سودمندی مهاجم منفی خواهد بود. در صورتی که یک مهاجم به WGA با شماره i که بر روی یکی از گره‌های شبکه قرار دارد حمله کند. دو پارامتر \hat{I}_1 , \hat{I}_2 را در نظر می‌گیریم که اولی میزان گسترش مهاجم از و دومی مربوط به پهنای باندی است که مهاجم برای حمله به عامل i ام اشغال می‌کند. این پارامتر توسط [۱۰] پیشنهاد

^{۱۶} Distributed IDS

جدول ۲: هزینه عایدی مهاجم

Attacker	نتیجه	هزینه
	نفوذ	$-b_1$
	عدم نفوذ	b_2

در نهایت هریک از دو بازیگر (WGA_i , attacker) یک مقدار سودمندی $payoff$ را متصور هستند که به صورت زیر برای هر یک محاسبه می‌شود. برای attacker میزان نتیجه بدین صورت است:

$$payoff_{att} = r_1 [p_a b_2 - b_1 (1 - p_a)]$$

که r_1 احتمال اینکه یک attacker رفتار خرابکارانه داشته و قصد نفوذ به سیستم را داشته باشد. مقدار بازده مورد انتظار WGA در IDS نیز به صورت زیر محاسبه می‌شود:

$$payoff_{WGA} = r_1 c_2 + p_a c_2 - r_1 p_d (c_1 + c_2 + c_3)$$

حال به بازی غیر همکارانه با استراتژی آمیخته بین دو بازیگر (WGA , attacker) می‌پردازیم مجموعه استراتژی‌هایی که هر یک از بازیگران می‌توانند داشته باشند بدین صورت است:

$$S_{attacker} = \{u_1, u_2, u_3\}$$

$$S_{WGA} = \{d_1, d_2\}$$

u_1 نشان‌دهنده یک حمله کامل‌تر توسط مهاجم با احتمال r_1 می‌باشد که قبلاً "عنوان گردید. u_2 را به عنوان استراتژی تکثیری یک حمله توسط مهاجم به نحوی که با احتمال r_2 به صورت گسترده اقدام به تکثیر خود نماید در نظر می‌گیریم.

u_3 نشان می‌دهد که با احتمال $1-r_1-r_2$ حمله‌ای صورت نمی‌گیرد. برای عامل نیز استراتژی d_1 نشان‌دهنده تشخیص یک نفوذ با احتمال q و اعلام هشدار لازم می‌باشد. استراتژی d_2 نیز حالتی را نشان می‌دهد که عامل هیچ عکس‌العملی را در مقابل حمله انجام نمی‌دهد. احتمال d_2 را با $1-q$ نشان می‌دهیم. با توجه به استراتژی‌های فرض شده میزان بازده مورد انتظار $payoff$ مهاجم و عامل را به صورت زیر بسط می‌دهیم. دلیل بسط بازده نوع بازی انتخابی است که از نوع استراتژی آمیخته در نظر گرفته شده است و برای اینکه بتوان احتمال کلیه استراتژی‌های فرض شده را در بازده نهایی لحاظ کرد جدول زیر را q -mix بازده انتظاری

شده‌اند. می‌توان نتیجه گرفت که یک حمله گسترش یافته λf_i از پهنای باند را اشغال می‌کند.

WGAgents: مجموعه‌ای از عامل‌های تشخیص‌دهنده WGA به صورت $N = \{WGA_1, WGA_2, \dots, WGA_n\}$ نمایش داده می‌شوند می‌باشد. این عاملها بر روی گره‌های شبکه مستقر شده و یک شبکه فوقانی چند عامله را برای تشخیص نفوذ و تشکیل یک IDS توزیع شده را می‌دهند. وظیفه این عاملها این است که با بررسی بسته‌های انتقالی درون شبکه کاربران عادی را از مهاجمین تشخیص دهند. فرض کنید که m_i نشان‌دهنده نرخ تحرک^{۱۷} عامل WGA_i در حرکت از یک میزبان به میزبان دیگر می‌باشد. عامل WGA_i بطور معمول می‌تواند دچار دو دسته خطا شود: دسته‌بندی مهاجم به عنوان کاربر عادی یا دسته‌بندی کاربر عادی به عنوان مهاجم. در این مدل به دنبال برقراری یک تعادل بین اینگونه هشدارها و هشدارهای صحیح می‌باشیم.

برای عامل WGA_i نرخ تشخیص نفوذ را با p_d (احتمال تشخیص) و نرخ عدم تشخیص را با $(1-p_d)$ نشان می‌دهیم و p_f را به عنوان احتمال تشخیص اشتباه حمله در نظر می‌گیریم. هزینه c_1 برای تشخیص یک حمله توسط WGA_2 و C_2 و C_3 به ترتیب هزینه‌هایی که IDS بابت تشخیص اشتباه و عدم تشخیص بایستی بپردازد فرض می‌کنیم.

جدول ۱: احتمالات شناسایی نفوذ توسط عامل.

WGA_i	P_d	احتمال تشخیص نفوذ	هزینه IDS
	$1-P_d$	احتمال عدم تشخیص نفوذ	$-C_1$
	P_f	احتمال تشخیص نفوذ اشتباه	C_2

از طرف دیگر مهاجم در صورت نفوذ به سیستم هزینه b_1 را بدست می‌آورد و در صورت عدم موفقیت در نفوذ هزینه b_2 را بایستی بدهد. هشدار اشتباه هزینه صفر را برای مهاجم در بردارد.

^{۱۷} Mobility

عامل) و r-mix (بازده انتظاری مهاجم) نشان می‌دهد در

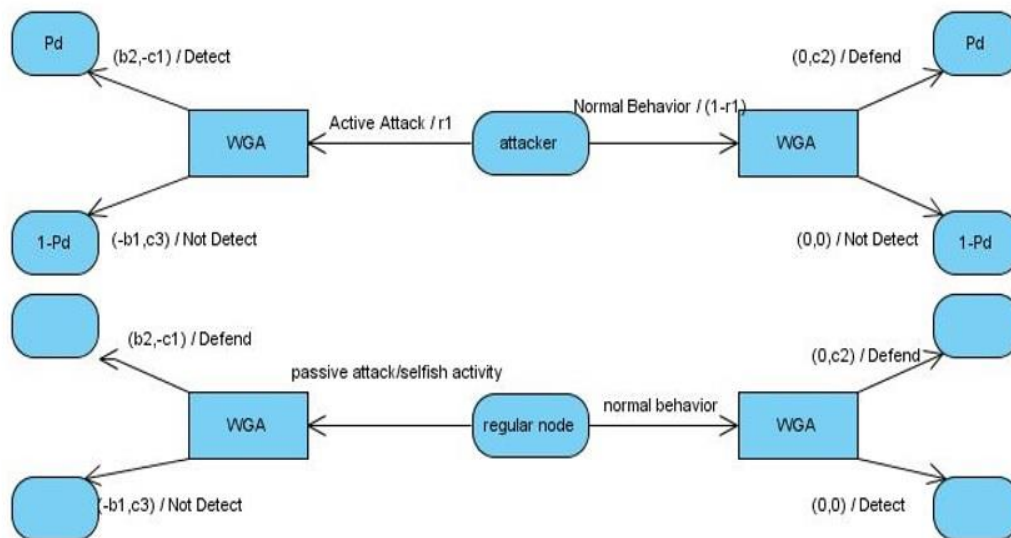
نظر می‌گیریم [۱۱].

جدول ۳: ماتریس payoff

attack	WGA _i		
	d ₁	d _r	q - mix
u ₁	$b_1 f_i, -c_1(1 + p_d + m_i)$	$-b_r f_i, c_r(1 + (1 - p_d))$	$b_1 f_i q - b_r f_i(1 - q)$
u _r	$b_1(1 + \lambda f_i + l_i), -c_1(1 + p_d + l_i + m_i)$	$b_r(1 + \lambda f_i + l_i), -c_r(1 + l_i + (1 - p_d))$	$b_1(1 + \lambda f_i + l_i) q - b_r(1 + \lambda f_i + l_i)(1 - q)$
u _r	$\cdot, c_r(1 + p_f + m_i)$	$\cdot \cdot$	\cdot
r - mix	$-c_1(1 + p_d + m_i)r_1 - c_1(1 + p_d + l_i + m_i)(r_r) + c_r(1 + p_f + m_i)(1 - r_1 - r_r)$	$c_r(1 + (1 - p_d))r_1 + c_r(1 + l_i + (1 - p_d))r_r$	

حرکت بازیگر دیگر هستند. عبارتی عامل مدافع دارای یک باور از حرکت خود و پاسخ متقابل مهاجم می‌باشد. در شکل ۴ نحوه این تعامل نشان داده شده است.

ماتریس بازده دو بازیگر در حالات مختلف انتخاب استراتژی را نشان می‌دهد. فرض بر این است که هر دو بازیگر چنین ماتریسی را داشته و قادر به محاسبه حالت



شکل ۴: وضعیت های مختلف تشخیص

نش) برسد. این نقطه که با (r^*, q^*) نمایش داده می‌شود از ماتریس ۳*۲ جدول ۳ بدست می‌آید.

در این مرحله هر بازیکن بایستی بازی را انجام دهد که در آن حالت هیچ انگیزه ای برای تغییر بازی خود نداشته باشد و بازیکن مقابل نیز به همین نقطه (تعادل

$$q^*[-c_1(1+p_d+m_i)r_1^* - c_1(1+p_d+l_i+m_i)(r_2^*) + c_2(1+p_f+m_i)(1-r_2^*-r_1^*)] + (1 - q^*)[c_3(1+(1-p_d))r_1^* + c_3(1+l_i+(1-p_d))r_2^*] \\ \leq q[-c_1(1+p_d+m_i)r_1^* - c_1(1+p_d+l_i+m_i)(r_2^*) + c_2(1+p_f+m_i)(1-r_2^*-r_1^*)] + (1 - q^*)[c_3(1+(1-p_d))r_1^* + c_3(1+l_i+(1-p_d))r_2^*]$$

بعد از حل نامساوی، اگر فرض کنیم که $r_2 = 1 - r_1$ باشد خواهیم داشت (برای مهاجم دو حالت کلی در نظر گرفته شود)

عبارت بالا نشان می‌دهد که بهترین استراتژی که بازیکن مقابل انتخاب می‌کند بایستی در هر حالتی از استراتژی انتخابی بازیکن کمتر باشد. q, r هر دو بین صفر و یک هستند $0 \leq q, r \leq 1$.

$$\text{if } -c_1(1+p_d+m_i)r - c_1(1+p_d+l_i+m_i)(1-r) = c_3(1+(1-p_d))r + c_3(1+l_i+(1-p_d))(1-r) \text{ and } b_1f_iq - b_2f_i(1-q) = b_1(1+\lambda f_i+l_i)q - b_2(1+\lambda f_i+l_i)(1-q) \text{ then}$$

$$r^* = \frac{c_1(1+p_d+l_i+m_i) + c_3(1+l_i+(1-p_d))}{-c_1(1+p_d+m_i) + c_1(1+p_d+l_i+m_i) - c_3(1+l_i+(1-p_d)) + c_3(1+(1-p_d))} \\ q^* = \frac{-b_2(1+\lambda f_i+l_i) + b_2f_i}{b_1(1+\lambda f_i+l_i) + b_2(1+\lambda f_i+l_i) + (b_1+b_2)f_i}$$

$$\text{if } -c_1(1+p_d+m_i)r + c_2(1+p_f+m_i)(1-r) = c_3(1+l_i+(1-p_d))(r) \text{ and } b_1f_iq - b_2f_i(1-q) = 0 \text{ then}$$

$$r^* = \frac{c_2(1+p_f+m_i)}{c_1(1+p_d+m_i) + c_2(1+p_f+m_i) + c_3(1+l_i+(1-p_d))} \\ q^* = \frac{b_2f_i}{(b_1+b_2)f_i}$$

$$\text{if } c_2(1+p_f+m_i)(1-r) - c_1(1+p_d+l_i+m_i)(r) = c_3(1+l_i+(1-p_d))(r) \text{ and } b_1(1+\lambda f_i+l_i)q - b_2(1+\lambda f_i+l_i)(1-q) = 0 \text{ then}$$

$$r^* = \frac{c_2(1+p_f+m_i)}{c_2(1+p_f+m_i) + c_1(1+p_d+l_i+m_i) + c_3(1+l_i+(1-p_d))} \\ q^* = \frac{b_2(1+\lambda f_i+l_i)}{b_1(1+\lambda f_i+l_i) + b_2(1+\lambda f_i+l_i)}$$

$\{d_1, d_2\}$ عامل مدافع بدست می‌آید.

فرض می‌کنیم v_i مقدار ریسک امنیتی عامل i ام (WGA_i) باشد آنگاه:

$$v_i = \frac{q_i^*(d_1)}{q_i^*(d_1)} + \frac{r_i^*(u_1) + r_i^*(u_2)}{r_i^*(u_r)} \quad i \in N$$

پس از محاسبه استراتژی‌های بهینه و تعادل نش بازی بردار احتمالی $r^* = \{r^*(u_1), r^*(u_2), r^*(u_r)\}$ برای مهاجم زمانی که استراتژی‌های $\{u_1, u_2, u_3\}$ را انتخاب می‌کند بدست می‌آید و بردار احتمالی $q^* = \{q^*(d_1), q^*(d_2), q^*(d_3)\}$ برای استراتژی‌های

۶- بازی همکارانه عاملها با یکدیگر

در این بخش بازی همکارانه عاملهای WGA برای تشکیل ائتلاف و تشخیص مقدار شیپلی با استفاده از نقطه تعادل نش یا همان SRV محاسبه شده در بخش قبلی توضیح داده خواهد شد. مقدار شیپلی یک اشاره‌گر قدرتمند برای مسئله انتساب هزینه است [۱۲]. بازی همکارانه که در آن همه بازیگران ($WGAS$) طی یک سری مذاکره با یکدیگر به دنبال افزایش میزان سودمندی گروه و ائتلاف تشکیل شده می‌باشند، راه حل مناسبی برای تشخیص میزان درجه امنیتی شبکه فوقانی است.

ابتدا تابع $R: V \rightarrow R^+$ به عنوان یک تابع یک به یک از اعداد حقیقی را به صورتی که هر عنصر از V ، I به صورت $V = \{v_1, v_2, \dots, v_j\}$ تعریف می‌شود. سطح امنیتی IDS پیشنهادی برای استفاده از L تعریف می‌کنیم. L به نحوی تعریف می‌شود که

$$L = \{l_1, \dots, l_L\} \text{ when } 0 < k_1 < k_2 < \dots < k_L$$

به عنوان مقادیر حد آستانه در نظر گرفته می‌شوند. در رابطه زیر حالات مختلف سطح امنیتی با کمک بردار خروجی عاملها V نشان داده شده است.

$$SL = \begin{cases} l_1 & \text{if } \sum_{i=1}^N R(v_i) \geq k_1 \\ l_j & \text{if } \sum_{i=1}^N R(v_i) \geq k_j \\ l_{j+1} & \text{if } \sum_{i=1}^N R(v_i) \geq k_{j+1} \\ l_L & \text{if } \sum_{i=1}^N R(v_i) \geq k_L \end{cases}$$

زمانی که

$$k_1 = v_{min} + k_{in} \quad , \quad k_j = v_{min} + j k_{in} \quad , \quad k_{j+1} = v_{min} + (j+1)k_{in} \quad , \dots \quad , \quad k_H = v_{min} + Lk_{in}$$

$$k_{in} = \frac{v_{max} + v_{min}}{L + 1}$$

عاملهای $WGAS$ را با استفاده از سطح امنیتی که توسط هر یک بدست آمده (k_j) در یکی از گروهها دسته‌بندی می‌کنیم SRV عامل را می‌توان بوسیله یک بازی N نفره با $X = \{1, 2, \dots, n\}$ مدل‌سازی کرد، که در آن مجموعه‌ای از بازیگران می‌باشد و می‌توان هر زیر مجموعه‌ای از آن را به صورت $V \subset N$ بدست آورد، به

نحوی که $\forall j \in V$ و $v_j \neq 0$ به عنوان یک ائتلاف شناخته می‌شود [۱۳] [۱۴]. ائتلاف X عامل در یک گروه با حد آستانه K از سطوح امنیتی، نشان‌دهنده یک الگوی حمله و سطح امنیتی L در گروه می‌باشد. مقدار مجموع ائتلاف^{۱۸} از جمع مقادیر SRV اعضا ائتلاف به صورت $R(c) = \sum_{i \in c} R(v_i)$ محاسبه می‌شود و به آن تابع ائتلاف گفته می‌شود. فرض کنید $R(c) = \sum_{i \in c} R(v_i)$ ، $v_i \in V$ ، $c \subset X$ با تعداد اعضا C می‌باشد. آنگاه مقدار شیپلی از I بین عضو بردار عاملها به صورت زیر تعریف می‌شود.

عاملهای $WGAS$ را با استفاده از سطح امنیتی که توسط هر یک بدست آمده (k_j) در یکی از گروهها دسته‌بندی می‌کنیم SRV عامل را می‌توان بوسیله یک بازی N نفره با $X = \{1, 2, \dots, n\}$ مدل‌سازی کرد، که در آن مجموعه‌ای از بازیگران می‌باشد و می‌توان هر زیر مجموعه‌ای از آن را به صورت $V \subset N$ بدست آورد، به نحوی که $\forall j \in V$ و $v_j \neq 0$ به عنوان یک ائتلاف شناخته می‌شود [۱۳] [۱۴]. ائتلاف X عامل در یک گروه با حد آستانه K از سطوح امنیتی، نشان‌دهنده یک الگوی حمله و سطح امنیتی L در گروه می‌باشد. مقدار مجموع ائتلاف^{۱۹} از جمع مقادیر SRV اعضا ائتلاف به صورت $R(c) = \sum_{i \in c} R(v_i)$ محاسبه می‌شود و به آن تابع ائتلاف گفته می‌شود. فرض کنید $R(c) = \sum_{i \in c} R(v_i)$ ، $v_i \in V$ ، $c \subset X$ با تعداد اعضا C می‌باشد. آنگاه مقدار شیپلی از I بین عضو بردار عاملها به صورت زیر تعریف می‌شود.

$$SP(i) = \sum_{c \subset X, i \in c} \frac{(c-1)! (n-c)!}{n!} [R(c) - R(c - \{i\}w)]$$

پس از محاسبه $sp(i)$ برای هر یک از عاملها که با استفاده از عضویت آنها در ائتلاف بدست آمده است می‌توان سطح ریسک امنیتی حمله صورت گرفته به آن عامل را با استفاده از مقادیر مختلف k_j در گروههای مختلف

^{۱۸} aggregate

^{۱۹} aggregate

امنیتی L قرار دهیم. بدین منظور ۴ حد آستانه برای L در نظر گرفته و عاملها را در هر یک از این دسته‌ها (ائتلاف) قرار می‌دهیم. از محاسبه مقادیر NE بردار v_i برای ۲۰ عامل $WGAS$ بدست می‌آید. سپس با استفاده از k_{in} حد آستانه ۴ گروه ائتلافی را محاسبه کرده و با استفاده از آن و مقادیر v_i مقدار دقیق شیپلی را بدست می‌آوریم. پس از محاسبه مقدار SP برای هر عامل می‌توان با کمک k_{in} های محاسبه شده برای ۴ سطح مورد نظر امنیتی، سطح تهدید برای هر عامل را در یکی از این ۵ گروه دسته‌بندی کرده و میزان خطر هر حمله را تشخیص داد. پس از دسته‌بندی عاملها در هر یک از گروهها مدیر سیستم می‌تواند با توجه به اطلاعات بدست آمده یا تصمیم‌گیری لازم را داشته باشد. مقادیر آستانه مربوط به سطح امنیتی می‌تواند به مرور زمان توسط مدیر سیستم تغییر کرده به نحوی که دسته‌بندی واقعی آن ارائه نماید.

امنیتی قرار دارد و مطابق با قرارگیری در هر گروه $l_1, l_2, \dots, l_{j+1}, l_j, l_1, l_L$ برای هر عامل یک درجه امنیتی تعیین نمود.

۷- شبیه‌سازی

به منظور شبیه‌سازی کارایی مدل پیشنهادی نیاز به یک مثال عددی وجود دارد ابتدا ۲۰ عدد فرضی به صورت تصادفی در یک ماتریس برای ۲۰ عامل تولید می‌کنیم. در این ماتریس به ازاء هر یک از عاملها اعدادی را نیز به عنوان پارامترهای حمله‌کننده برای پارامترهای مختلف آن از قبیل b_1, b_2, \dots و به صورت تصادفی تولید می‌نماییم. در ادامه با استفاده از نرم‌افزار شبیه‌سازی بازی به نام $GAMBIT$ [۱۵][۱۶] نقطه تعادل نش را برای هر یک از عاملها و حمله‌کننده بدست آورده و در جدول درج می‌کنیم. سپس برای محاسبه مقدار شیپلی از نرم‌افزار $MATLAB$ استفاده کرده و سعی می‌کنیم که مقدار SRV هر عامل را مشخص کرده و در یکی از

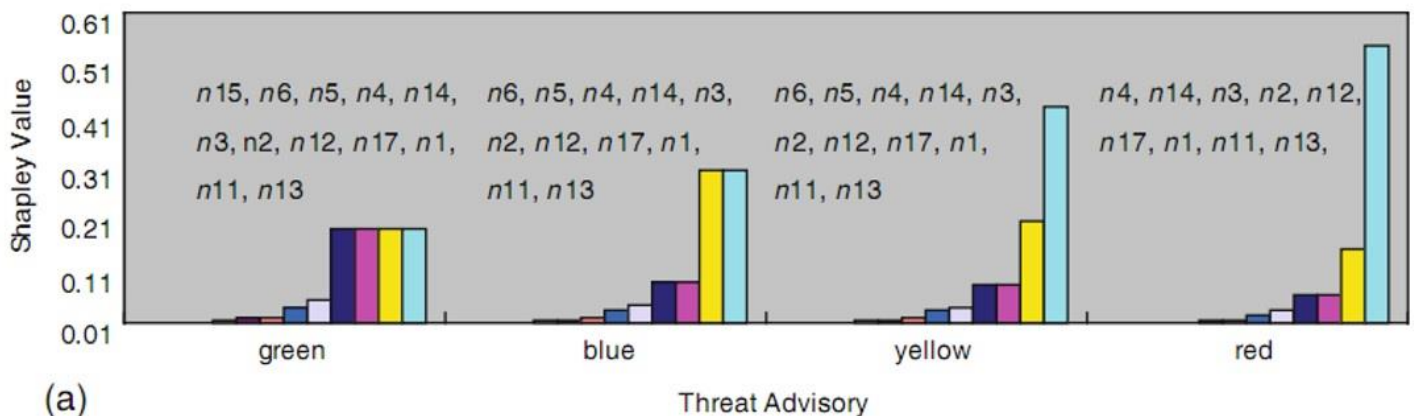
جدول ۴: پارامترهای عامل مهاجم و نقطه تعادل نش هر یک

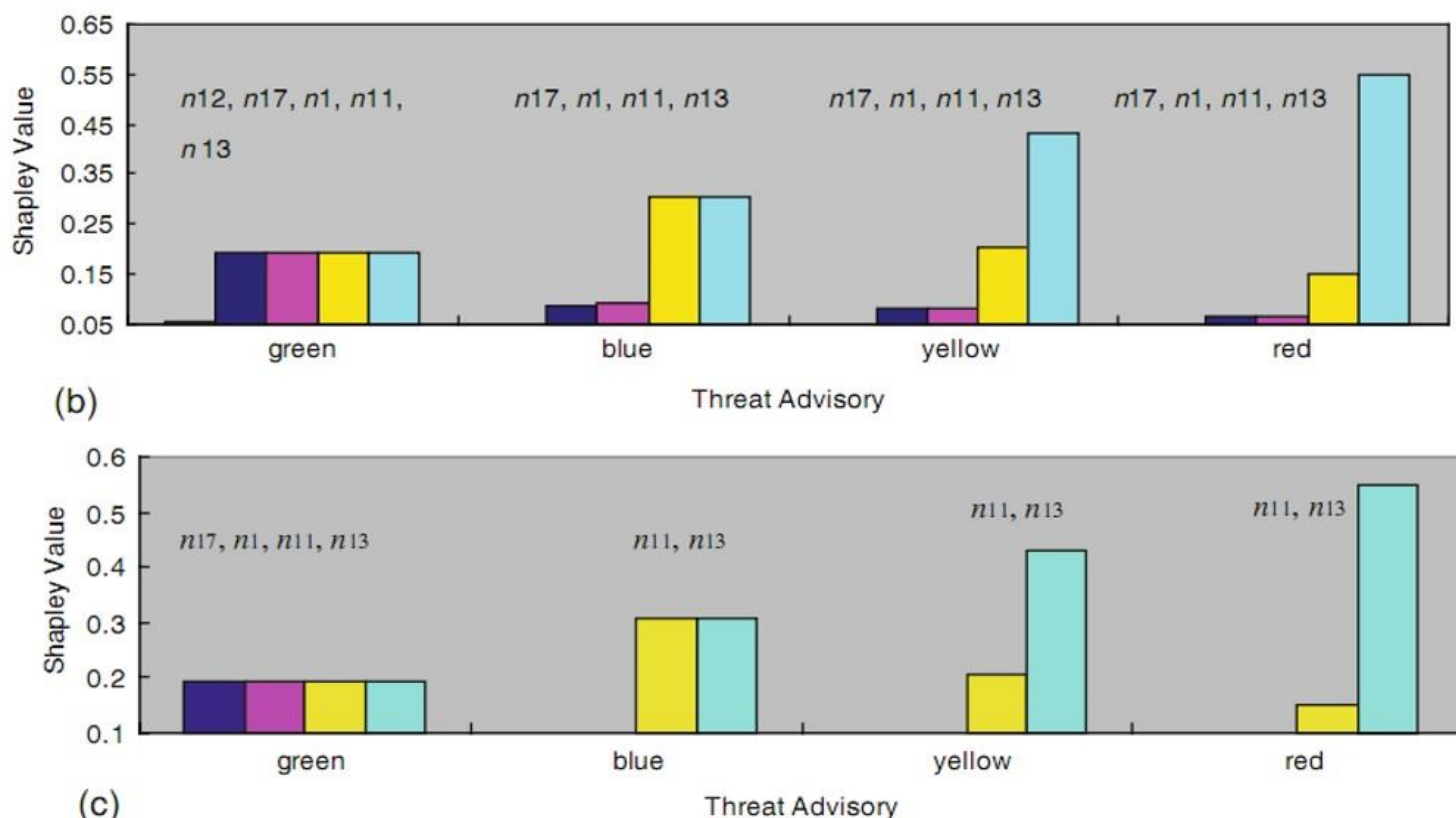
Agent	Attacker and Agent's params										Nash v_i
	b_1	$-b_2$	$f_i\%$	$-c_1$	c_2	c_3	$l_i\%$	$m_i\%$	$p_d\%$	$p_f\%$	
N_1	۱۰	- ۹۰	۰,۴ ۹	- ۲۵	۷۱	۵۵	۰,۸	۰,۳ ۵	۰,۵	۰, ۷	۱۰,۷ ۳۲
N_2	۲۰	- ۱۰۰	۰,۴ ۴	- ۸۱	۲۰	۱۰	۰,۳	۰,۹	۰,۸	۰, ۲	۳,۹۷
N_3	۳۰	- ۵۰	۰,۵ ۵	- ۵۰	۳۱	۳۶	۰,۳	۰,۴ ۵	۰,۸	۰, ۳	۲,۳۱
N_4	۴۰	- ۷۰	۰,۳ ۵	- ۷۰	۴۰	۵۰	۰,۴	۰,۷	۰,۷	۰, ۴	۲,۱۱
N_5	۵۰	- ۵۰	۰,۲ ۵	- ۵۰	۶۰	۴۰	۰,۶	۰,۵	۰,۸	۰, ۶	۱,۵۸ ۹
N_6	۶۰	- ۹۰	۰,۲ ۹	- ۶۰	۵۰	۶۰	۰,۵	۰,۶	۰,۶	۰, ۵	۲,۵۴ ۷
N_7	۷۰	- ۴۰	۰,۶ ۴	- ۴۰	۹۰	۷۰	۰,۸	۰,۵	۰,۷	۰, ۴	۳,۲۰
N_8	۸۰	- ۵۰	۰,۷ ۹	- ۹۰	۶۰	۵۰	۰,۷	۰,۴	۰,۸	۰, ۳	۱,۲۳ ۵
N_9	۹۰	- ۹۰	۰,۳ ۹	- ۸۰	۷۰	۹۰	۰,۱ ۵	۰,۲	۰,۷	۰, ۹	۲,۲۳ ۱
N_{10}	۱۰	- ۱۱	۰,۰ ۵	- ۵	۱۰	۹۰	۱	۰,۱	۰,۴	۰, ۲	۰,۰۰ ۸۷

Agent	Attacker and Agent's params										Nash
	b_1	$-b_2$	$f_i\%$	$-c_1$	c_2	c_3	$l_i\%$	$m_i\%$	$p_d\%$	$p_f\%$	v_i
N13	۲۵	- ۹۱	۰,۴ ۴	- ۸۱	۱۵	۱۴	۰,۰ ۵	۰,۹ ۹	۰,۵	۰, ۱	۵,۷
N14	۳۵	- ۹۸	۰,۳ ۴	- ۱,۵	۹۸	۱۰	۰,۹ ۱	۰,۹ ۹	۰,۵	۰, ۱	۴۹,۰ ۱
N15	۴۵	- ۵۵	۰,۲ ۱	- ۶۵	۳۳	۳۱	۰,۳ ۵	۰,۶ ۵	۰,۷	۰, ۲	۲,۳۸ ۵
N16	۵۵	- ۴۱	۰,۲	- ۴۱	۵۵	۵۰	۰,۵ ۵	۰,۴ ۵	۰,۷ ۹	۰, ۴۵	۱,۴۵
N17	۶۵	- ۱۰۰	۰,۴	- ۱۰	۷۱	۸۵	۰,۸ ۱	۰,۲ ۱	۰,۵	۰, ۳	۱۰,۰ ۲
N18	۷۵	- ۲۵	۰,۷	- ۳۰	۷۴	۷۹	۰,۷ ۵	۰,۲ ۵	۰,۸	۰, ۲۵	۱,۰۹ ۳
N19	۸۵	- ۱۰	۰,۰ ۵	- ۸۰	۱۵	۱۵	۰,۱ ۴	۰,۸ ۵	۰,۵	۰, ۵	۰,۳۸ ۱
N20	۱۰	- ۱۰	۰,۰ ۱	- ۷۰	۲۰	۵۴	۰,۹	۰,۹	۰,۹	۰, ۲	۴,۰۲ ۱
N13	۲۵	-	۰,۴	-	۱۵	۱۴	۰,۰	۰,۹	۰,۵	۰, ۱	۵,۷

چهارم قرار میگیرند. لذا ائتلافی بدین صورت تشکیل می شود و می توان سطح امنیتی این گروهها را معادل یکدیگر در نظر گرفت. با تغییر در مقدار شیپلی می توان این دسته بندی ها را تغییر داد و شبکه فوقانی مختلفی را برای دفاع در مقابل حملات تشکیل داد. در شکل زیر چندین مقدار شیپلی برای حالات مختلف در نظر گرفته شده که تشکیل ائتلاف در سه وضعیت را نشان می دهند.

مقدار حد آستانه برای هر یک از گروه ۴ را محاسبه کرده و مقادیر $\{۹,۰۸, ۳,۰۲, ۲۰,۱, ۴۰,۷۲\}$ برای k_{in} بدست می آید که مقادیر نش محاسبه شده را در چهار گروه دسته بندی می کند. سپس با استفاده از فرمول مقدار شیپلی و مقادیر k_{in} امکان محاسبه دقیق $SP(i)$ میسر می شود. عاملهای $\{n3, n4, n5, n6, n8, n9, n10, n11, n15, n16, n18, n19\}$ در گروه اول و $\{n2, n13, n20\}$ در گروه دوم و $\{n1, n12, n17\}$ در گروه سوم و $\{n14\}$ در گروه





شکل ۵- سه حالت پیکر بندی ائتلاف برای مقادیر شیپلی (a) بزرگتر از ۰,۷ و (b) بزرگتر از ۰,۸ و (c) بزرگتر از ۰,۹

۷. نتیجه گیری

از شبکه را پوشش دهد. علاوه بر این در صورت تشخیص ضعیف یا اشتباه، عاملها با تشکیل ائتلاف و محاسبه SRV تجمیعی (مقدار شیپلی) درصد خطای تشخیص و هشدارهای غیر صحیح (false alarm) را به حداقل می-رساند. در این مدل مقادیر حد آستانه براساس تشخیص-های عامل WGA (v_i) به روز می‌شود (k_{in}) و در هر بار تشکیل ائتلاف این مقادیر تغییر می‌کند و در عمل سیستم در هر بار از دوره‌های محاسبه SRV گروه‌بندی جدیدی را انجام می‌دهد. برای ادامه این تحقیق، محققین قصد دارند با استفاده از روشهای یادگیری از جمله Reinforcement Learning محاسبه k_{in} را به صورت تدریجی محاسبه کرده و در هر بار محاسبه از داده‌های مراحل قبل نیز بهره برده شود.

در این مقاله یک مدل امنیتی توزیع شده برای تشخیص نفوذ به شبکه DIDS مبتنی بر تئوری بازیها ارائه شده است. در سیستم‌های سنتی IDS مدیران سیستم می‌بایستی دائماً به رصد سیستم پرداخته و کلیه اتفاقات و سناریوهای ممکن را بررسی و پایش نمایند. در این تحقیق، ما مدلی را مبتنی بر تئوری بازیها در فرم گسترده به دو صورت همکارانه و غیر همکارانه پیشنهاد کردیم که در نوع غیر همکارانه (رقابتی) مقدار ریسک امنیتی با استفاده از تعادل نش توسط هر کدام از عاملها محاسبه شده و در بازی همکارانه عاملهایی که در یک گروه قرار می‌گیرند با تشکیل ائتلاف و محاسبه مقدار شیپلی درجه امنیتی ائتلاف را محاسبه می‌نماییم. نظر به اینکه در این مدل عاملهای $WGAs$ درون شبکه گسترده شده‌اند، با تشکیل یک شبکه فوقانی تشخیص نفوذ می‌توانند عمده تراکنش‌ها و داده‌های رد و بدل شده در شبکه را رصد کرده و به صورت توزیع شده بخش عمده‌ای

منابع

۹. D. S. P. J. M. R. K. G. Shaw "Inside the minds of the insider", Security Management, جلد ۴۳, p. ۳۴-۴۴, ۱۹۹۹.
۱۰. T. B. T. Alpcan "A game theoretic approach to decision and analysis in network intrusion detection", IEEE Conference on Decision and Control, p. ۲۵۹۵-۲۶۰۰, ۲۰۰۳.
۱۱. C.-K. W. Yi-Ming Chen "A Game Theoretic Framework for Multi-agent Deployment in Intrusion Detection Systems", Security Informatics, Annals of Information Systems, pp. ۱۱۷-۱۳۳, ۲۰۱۰.
۱۲. D. R. B. Mishra "Cost sharing in a job scheduling problem using the Shapley value", ۲۰۰۵.
۱۳. P. Z. Liu "Incentive-based modeling and inference of attacker intent, objectives and strategies", ACM Transactions on Information and System Security, جلد ۸, p. ۷۸-۱۱۸, ۲۰۰۵.
۱۴. A. S. S. Dixit "Games of Strategy", ۲۰۰۱.
۱۵. R. D. M. A. M. a. T. T. L. McKelvey, "Gambit: Software Tools for Game Theory", ۲۰۰۷, Available: <http://econweb.tamu.edu/gambit>.
۱۶. GAMBIT "GAMBIT", ۲۰۱۳, Available: <http://www.gambit-project.org/>
۱. C. Symantec "Symantec Corporation", ۲۰۱۳, Available: <http://www.symantec.com/index.jsp>.
۲. P. H. V. L. R. J. Z. H. M. Mel "An Overview of Issues in Testing Intrusion Detection Systems", NIST, Gaithersburg, MD, ۲۰۰۲.
۳. A. D. V. M. R. D. Keromytis, "SOS: An architecture for mitigating DDoS attacks," *IEEE Communications*, vol. ۲۲, p. ۱۷۶-۱۸۸, ۲۰۰۴.
۴. G. V. S. Suryawanshi "Mobile Agent for Distributed Intrusion detection System in Distributed System", *International Journal of Artificial Intelligence and Computational Research (IJAI CR.)*, p. ۲۰۱۰.
۵. M. A. M. S. M. A. K. & M. R. M. I. Kamaruzaman Maskat "Mobile Agents in Intrusion Detection System: Review and Analysis", *Modern Applied Science*, شماره ۵, جلد ۵, pp. ۲۱۸-۲۳۱, Dec ۲۰۱۱.
۶. R. B. a. P. Mell "Intrusion detection systems", ۲۰۱۲, Available: <http://www.snort.org/docs/nistids.pdf>.
۷. S. E. Schechter, "Computer Security Strength and Risk: A Quantitative Approach," *PhD Thesis, Harvard University*, ۲۰۰۴.
۸. S. E. Schechter "Toward econometric models of the security risk from remote attacks", *IEEE Security & Privacy*, شماره ۱, جلد ۳, p. ۴۴-۴۰, ۲۰۰۵.

تلفیق مدل تحلیل پوششی داده‌ها و درخت تصمیم به منظور ارزیابی واحدهای مبتنی بر تکنولوژی اطلاعات

امیر امینی*

علیرضا علی‌نژاد**

سمیه شفقی‌زاده***

*دانشجوی کارشناسی ارشد، مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه صنعتی ارومیه، ارومیه، ایران

**دانشیار دانشگاه آزاد اسلامی، واحد قزوین، دانشکده مهندسی صنایع و مکانیک، قزوین، ایران

***دانش‌آموخته کارشناسی ارشد مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه آزاد اسلامی، واحد علوم تحقیقات قزوین، قزوین؛ ایران

تاریخ دریافت: ۹۳/۰۸/۳۰

تاریخ پذیرش: ۹۵/۰۸/۲۳

چکیده

هر سازمان به منظور آگاهی از میزان عملکرد و مطلوبیت فعالیت واحدهای خود به یک نظام ارزشیابی جهت سنجش این مطلوبیت نیاز دارد و این موضوع برای مؤسسات مالی از جمله شرکت‌های مبتنی بر تکنولوژی اطلاعات اهمیت بیشتری دارد. تحلیل پوششی داده‌ها^۱ یک روش غیرپارامتری برای اندازه‌گیری کارایی و بهره‌وری از واحدهای تصمیم‌گیری (DMUs) است. از طرف دیگر تکنیک داده‌کاوی^۲ به (DMUs) اجازه کاوش و کشف اطلاعات معنی‌داری، که قبلاً در پایگاه داده‌های بزرگ پنهان بوده‌اند را می‌دهد. این مقاله یک چارچوب کلی ترکیب تحلیل پوششی داده‌ها با درخت رگرسیون^۳ را برای ارزیابی کارایی و بهره‌وری از DMUها پیشنهاد می‌کند. نتیجه مدل ترکیبی مجموعه‌ای از قوانین است که می‌تواند توسط سیاست‌گذاران برای کشف دلایل DMUهای کارآمد و ناکارآمد مورد استفاده قرار گیرد. به عنوان مطالعه موردی با استفاده از روش پیشنهادی برای بررسی عوامل مرتبط با بهره‌وری نمونه‌ای شامل ۱۸ شعبه از بیمه ایران در شهر تهران انتخاب گردید و پس از مدل‌سازی بر مبنای مدل پیشرفته LVM^۴ ورودی محور با دسترسی پذیری ضعیف در تحلیل پوششی داده‌ها با خروجی نامطلوب محاسبه گردید و با تکنیک درخت تصمیم، به استخراج قوانین برای کشف دلایل افزایش بهره‌وری و پسرقت بهره‌وری می‌پردازد.

واژه‌های کلیدی: تحلیل پوششی داده‌ها، داده‌کاوی، طبقه‌بندی و رگرسیون، درخت تصمیم، خروجی نامطلوب

^۱ Data Envelopment Analysis

^۲ Decision Making Units

^۳ Data Mining

^۴ C&R

^۵ Latent Variable Model

^۶ Undesirable Output

۱- مقدمه

کارا استفاده می‌کنند و نحوه عملکرد شرکت‌های بیمه‌ای در استفاده بهینه از منابع در اختیار آن‌ها از دیگر عوامل مؤثر بر توسعه بازارهای بیمه و افزایش کارایی آنها محسوب می‌گردد.

امروزه شرکت‌ها سعی می‌کنند عملکردشان را اندازه‌گیری کرده و آن را با رقبایشان مقایسه نمایند تا بتوانند جهت حصول به سطح عملکردی که بتواند آنها را در بازار حفظ نماید، اقدام مقتضی را انجام دهند [۲۹].

چارنز و همکاران (۱۹۷۸) به معرفی یک روش غیرپارامتری اندازه‌گیری کارایی تکنیکی با یک مجموعه از واحدهای تصمیم‌گیرنده قابل مقیاس با استفاده از روش تحلیل پوششی داده‌ها پرداختند.

اندازه‌گیری کارایی واحدهای تصمیم‌گیرنده (DMUs)ها مانند بیمارستان‌ها، بانک‌ها، شرکت‌های بیمه و... که با چندین ورودی، چندین خروجی تولید می‌کنند کار بسیار پیچیده‌ای است. تحلیل پوششی داده‌ها از مفهوم ارزیابی کارایی یک گزینه تصمیم‌گیری بر مبنای عملکردش در ایجاد خروجی‌ها بواسطه مصرف ورودی بسط یافته است [۳۵]. کارایی هر واحد تصمیم‌گیری در مقایسه با واحد دیگر به عنوان نسبت مجموع وزنی خروجی‌های آن عضو به مجموع وزنی ورودی‌ها تعریف می‌شود [۲۷].

در طول دهه گذشته DEA به عنوان یک ابزار مدیریتی قدرتمند برای ارزیابی عملکرد به طور گسترده، توجه قابل ملاحظه‌ای را به دست آورده است و از طرف دیگر درخت تصمیم یک تکنیک داده کاوی برای طبقه‌بندی نمونه‌ها یا عملکرد ثابت تکه‌ای در رگرسیون غیرخطی است، این الگوریتم توسط براینمن^۷ و همکاران (۱۹۸۴) ابداع شده است. برای ارزیابی کارایی شرکت‌های بیمه از مدل تحلیل پوششی داده‌ها با خروجی نامطلوب استفاده شده به این دلیل که فرآیند تولید صنعت بیمه، در حالت کلی شامل دو فعالیت است؛ یکی فعالیت بازاریابی که مربوط می‌شود به بازاریابی، بیمه‌کردن و پرداختن ادعای خسارت‌ها که به عنوان خروجی

اندازه‌گیری راندمان یا کارایی، همیشه یکی از مباحث مهم مدیریت بوده است. آنچه پیکره دفاعی یک واحد اقتصادی را استحکام می‌دهد و بر دیگر رقبا ارتقاء می‌بخشد، نگاه تیزبینانه‌ای است که برنامه‌ریزان آن واحد با ایجاد بستری مناسب سعی می‌کنند که ضمن شناسایی نقاط ضعف و قوت موجود، درصدد رفع نقایص و ارتقاء قوت در کل مجموعه برآیند و این میسر نمی‌شود، مگر با استفاده از روش‌هایی که در نهایت صحت و به دور از هرگونه اغواگری، موجب شناسایی واحدهای کارا از ناکارا شود.

امروزه صنایع با یک محیط به شدت رقابتی مواجه‌اند. صنعت بیمه نیز از این محیط به شدت رقابتی بی بهره نبوده و در سال‌های اخیر شاهد رقابت فزاینده‌ای بین شرکت‌های بیمه بوده‌ایم، در مواجهه با یک محیط رقابتی، تدوین استراتژی‌های رقابتی، تقویت فعالیت‌های شرکت و ارتقای کیفیت خدمات برای بقا ضروری است، در جهان امروز، بیمه از مهمترین بخش‌های اقتصادی در جامعه است که علاوه بر تسهیل روابط بین الملل در پیشگیری از تحقق ریسک و نیز تأمین آرامش خاطر برای فعالیت‌های مختلف نقش شگرفی را ایفا می‌نماید. شعب مختلف یک شرکت بیمه به‌رغم استفاده از منابع مادی و انسانی مشابه، عملکردهای متفاوتی دارند. یکی از دلایل تفاوت در عملکرد شعب و نمایندگی‌های بیمه در نحوه به‌کارگیری منابع انسانی و مدیریت می‌توان دانست [۴].

این تحقیق به دنبال ارزیابی کارایی شعب بیمه ایران است تا واحدهای تصمیم‌گیرنده بتوانند به کمک نتایج تحقیق یا بسط آن اهداف زیر را پیگیری نمایند:

- ایجاد فضای رقابت سالم بین شعب با توجه به نتایج تحقیق
- استفاده از تجربیات واحدهای کارا و انتقال تجربه و دانش آن‌ها به سایر شعب
- اهداف کاربردی مستتر در این ارزیابی، استفاده از نتایج جهت سیاست‌گذاری شرکت و استفاده از آن به عنوان محک مدیران شعب است.

اما این سؤال مطرح می‌شود که شرکت‌های بیمه‌ای موجود در کشور تا چه اندازه از منابع در دسترس خود به صورت

^۷ Breiman

نگاهی ساده به بهره‌وری، نسبت بازده به منابع مصرف شده است. لذا اگر با هزینه یکسان محصول بیشتری تولید شود یا مقدار معینی محصول با هزینه کمتری تولید شود، منافعی برای کل جامعه حاصل می‌شود که می‌تواند برای تأمین زندگی بهتر برای افراد جامعه مؤثر باشد [۱].

اندازه‌گیری کارایی و روش‌های محاسباتی آن از اواخر قرن بیستم توجه بیشتر اقتصاددانان را برانگیخته است [۳۷].

روش‌های اندازه‌گیری اغلب برآورد مرز کارایی تولید و به دست آوردن منحنی مرزی یا تابع تولید مرزی و کلاً به دو روش پارامتریک و ناپارامتریک بوده است. روش تحلیل پوششی داده‌ها^۹ که در سال ۱۹۷۸ چارنر و همکارانش^{۱۰} ارائه کردند، نخست برای ارزیابی اقتصادی و فنی واحدهای تولیدی معرفی گردید. امروزه این روش برای ارزیابی عملکرد واحدهای خدماتی، اعم از دولتی و غیردولتی و واحدهای تولیدی [۲۵] و [۳۴] و [۱۹] و [۶] کاربرد گسترده‌ای دارد. کائو و هوآنگ [۲۴] برای ارزیابی کارایی شرکت‌های بیمه غیرعمر تایوان از مدل دو مرحله‌ای و در واقع DEA دو مرحله‌ای استفاده کرده‌اند. مدل ارتباطی DEA ارتباطی متداول است تا بتواند رابطه فیزیکی بین فرآیند کل و زیرتوسعه یافته مدل DEA دو مرحله‌ای، محدودیت‌های DEA و فرآیندهای جزء را توصیف کند. در مدل ارتباطی هر دو زیر فرآیند، به محدودیت فرآیند کل اضافه می‌شوند.

در صنعت بیمه نیز تحقیقات متعددی از این روش جهت ارزیابی عملکرد نمایندگی‌های بیمه بهره گرفته‌اند [۵] و [۲۸] و [۳۹].

در همین سال تحقیقات متعددی به طرق گوناگون در زمینه ارزیابی کارایی شرکت‌های بیمه با استفاده از روش تحلیل پوششی داده‌ها در کشورهای آلمان [۲۸]، چین [۱۴]، تایوان [۱۸]، ایالات متحده [۳۰] و بسیاری از کشورهای دیگر انجام شده است.

که تولید کمتر آنها به صرفه‌تر است و آن را عوامل نامطلوب می‌نامیم و دیگری فعالیت سرمایه‌گذاری که عبارت است از منافع حاصل از سرمایه‌گذاری منابع مالی مؤسسه بیمه.

این تحقیق بر آن است تا با استفاده از مدل LVM^۸، تغییرات کارایی و بهره‌وری شعب بیمه ایران در طی سال (۱۳۸۹-۱۳۸۷) در شهر تهران را باهم مقایسه نموده، شاخصی برای تعیین کارایی و ناکارایی یک واحد با استفاده از تغییرات مرز حاصل از ورودی واحدها و تغییرات کارایی آنها صورت دهد و در نهایت با ارائه نتایج، کارایی و بهره‌وری شعب را مورد بحث و بررسی قرار داده و جایگاه هر یک از شعب را در این مدل تعیین کند.

نهاده‌هایی که در این تحقیق مورد استفاده قرار گرفته‌اند، تعداد نمایندگی‌ها، هزینه‌های اداری و بیمه‌گری هستند و ستانده‌های مورد استفاده، درآمد حاصل از حق بیمه و ارزش وام‌های پرداختی به عنوان خروجی مطلوب و ارزش خسارت پرداختی به عنوان خروجی نامطلوب در نظر گرفته شده است، که شناسایی شاخص‌ها و معیارهای عملکردی از طریق مصاحبه و نظرسنجی با کارشناسان و مدیران ارشد بیمه ایران و مطالعه اسنادی صورت پذیرفت. این داده‌ها با مراجعه مستقیم امور مالی و شعب بیمه ایران در شهر تهران جمع‌آوری شده است.

تحقیق حاضر از نظر هدف، از نوع تحقیقات کاربردی محسوب می‌شود. همچنین از نظر نحوه گردآوری داده‌ها این تحقیق توصیفی و از نوع علی-مقایسه‌ای می‌باشد. جامعه آماری این تحقیق کلیه شعب بیمه ایران، در شهر تهران می‌باشند.

۲. بیان مسأله

مفهوم کارایی در اقتصاد، تخصیص مطلوب منابع است. اما از نظر اهداف کاربردی، تعاریف گوناگونی بیان شده است. به طور کلی کارایی، معرف نسبت ستانده‌ها به نهاده‌ها در مقایسه با یک استاندارد مشخص است [۷].

در عصر ما کارایی و بهره‌وری از بالاترین اهداف مدیران و ارزشمندترین مقصد همه سازمان‌ها می‌باشد.

تلاش برای افزایش بهره‌وری جدی‌ترین مبارزه‌ای است که مدیریت در آستانه قرن بیست و یکم با آن روبروست. در

^۹ Data Envelopment Analysis

^{۱۰} Charnes et al, ۱۹۷۸

^۸ Latent Variable Model

پوششی داده‌ها و درخت تصمیم به عنوان یک روش الگوبرداری در فرایند خدمت استفاده کرده‌اند [۳۲]. بابازاده و همکاران (۲۰۱۵) رویکرد یکپارچه دوگانه‌ای از تحلیل پوششی داده‌ها و تکنیک‌های برنامه‌ریزی ریاضی را به منظور طراحی راهبردی شبکه زنجیره تأمین زیست - دیزل در ایران ارائه داده‌اند [۸].

۳. مواد و روش

در این مقاله، دو روش تلفیقی تحلیل پوششی داده‌ها و داده کاوی که برای تخمین بهره‌وری واحدهای تصمیم گیرنده (DMUs) مورد استفاده قرار می‌گیرد، در نظر گرفته شده است. در روش تحلیل پوششی داده‌ها، تخمین کارایی و صرفه‌های مقیاس، با استفاده از سیستم برنامه‌ریزی ریاضی صورت می‌گیرد. در این روش محدودیتی در انتخاب تعداد نهاده‌ها و ستانده‌ها نداریم، البته به شرط آنکه تعداد واحدهای تصمیم‌گیرنده به اندازه کافی زیاد باشند، هیچ گونه نیازی به انتخاب نوع تابع تولید نیست و تکنیک داده کاوی اجازه کشف و بررسی معنادار اطلاعاتی که قبلاً از یک پایگاه داده بزرگ پنهان بوده‌اند را می‌دهد، C&R همان درخت تصمیم در داده‌کاوی است که توسط برایمن و همکاران در سال ۱۹۸۴ توسعه یافته است و بعدها توسط رایپلی^{۱۶} بهبود یافته است.

۳-۱- روش تحلیل پوششی داده‌ها (DEA)

به طور کلی شالوده روش ناپارامتری برای اندازه‌گیری کارایی با انتشار مقاله‌ای از فارل^{۱۷} (۱۹۵۷) بنیان نهاده شد. وی با یک رهیافت صرفاً ریاضی، روش جدیدی را برای اندازه‌گیری کارایی مدل در مقابل روش‌های پارامتری معرفی نمود [۲۳]. تحلیل پوششی داده‌ها روشی ناپارامتریک بر مبنای برنامه‌ریزی خطی است و نخستین بار توسط چارلز و همکاران ارائه شد [۱۳].

چن و همکارانش^{۱۱}، کارایی شرکت‌های بیمه تایوان را با استفاده از روش دومرحله‌ای تحلیل پوششی داده‌ها بررسی کردند [۱۴].

گونزالس و همکاران^{۱۲} در سال ۲۰۰۹ به تحلیل DEA از خطر، هزینه و درآمد در بیمه پرداختند، با استفاده از تحلیل پوششی داده‌ها به مطالعه در زمینه تجزیه و تحلیل نتایج حاصل از DEA در زمینه چند متغیره، تحلیل آماری با استفاده از تحلیل عاملی تفسیر شده و تأثیر عوامل خطر در بهره‌وری مشخص شده است.

زای^{۱۳}، کارایی شرکت‌های بیمه دولتی و خصوصی ایالات متحده را با استفاده از روش تحلیل پوششی داده‌ها مقایسه کرد و دریافت به‌رغم کارایی بالای شرکت‌های بیمه خصوصی، علت تمایل ایشان به دولتی شدن به دلیل دسترسی به منابع و سرمایه بیشتر است [۳۸].

باروز و همکارانش^{۱۴} نیز در همین سال با استفاده از روش دو مرحله‌ای تحلیل پوششی داده‌ها کارایی شرکت‌های بیمه یونان را در دوره ۹ ساله بررسی کردند [۹].

حمید اسفاندرانی [۴] در تحقیق خود که جامعه آماری آن شامل ۲۴ شعبه و ۱۰۰ نمایندگی واقع در تهران شرکت بیمه ایران می‌باشد. در این تحقیق انتخاب نهاده‌ها و ستانده‌ها با توجه به تحقیقات مشابه بوده و مصاحبه و نظرسنجی از کارشناسان و خبرگان بیمه انجام گرفته است. در نهایت با توجه به تجزیه و تحلیل‌های انجام گرفته از مدل‌های متعدد DEA، مدل BCC نهاده‌گرا به منظور اندازه‌گیری کارایی شبکه فروش بیمه عمر مدلی مناسب تشخیص داده شده است.

شاهرودی و همکاران^{۱۵} (۲۰۱۱) به تجزیه و تحلیل کارایی بیمه‌های خصوصی در ایران با روش تحلیل پوششی داده‌ها، به محاسبه کارایی بیمه‌های خصوصی بین سال‌های ۲۰۰۶-۲۰۰۷ با روش DEA دومرحله‌ای پرداختند. سؤال و همکاران (۲۰۰۷) از تلفیق تحلیل

^{۱۱} Chen et al, ۲۰۰۹

^{۱۲} MM Segovia-Gonzalez, Contreras and C Mar-Molinero

^{۱۳} Xie, ۲۰۱۰

^{۱۴} Barros et al, ۲۰۱۰

^{۱۵} Shahroudi, ۲۰۱۱

^{۱۶} Ripley

^{۱۷} Farrell ۱۹۵۷

به معنای بازده به مقیاس ثابت و $u_0 < 0$ نشان دهنده بازده به مقیاس کاهش است.

مدل (۱) یک مدل وزنی است و پوششی این مدل در مدل (۲) نشان داده شده است.

کارایی تکنیکی: یک مدل برنامه ریزی خطی برای بازده به مقیاس متغیر: مدل پوششی، ورودی محور

$$\begin{aligned} \min h_0 &= \varepsilon (s^+_r + s^-_i) \\ \text{St:} & \\ \sum_{j=1}^n \lambda_j x_{ij} + s^-_i &= h_0 \cdot x_{ij}, \quad \forall i = 1, \dots, m \\ \sum_{j=1}^n \lambda_j y_{rj} - s^+_r &= y_{rj}, \quad \forall r = 1, \dots, s \\ \sum_{j=1}^n \lambda_j &= 1 \\ \lambda_j &\geq 0, \quad \forall j = 1, \dots, n \end{aligned} \quad (2)$$

در این مدل، محدودیت اول نشان‌دهنده این است که همه ترکیب‌های ورودی مشاهده شده برای ژامین واحد تصمیم‌گیرنده بر روی مرز یا زیر مرز تولید قرار دارد و دومین محدودیت نشان می‌دهد که سطح‌های خروجی از مشاهدات ناکارایی با سطح‌های واحد مرجع قابل مقایسه است، که از ترکیب محدب خروجی مشاهده تشکیل شده است.

محدودیت سوم معرف بازده به مقدار متغیر است و محدودیت آخر تضمین می‌کند که همه مقادیر تولید اوزان تحدب بزرگتر مساوی یک است، بنابراین واحد فرضی مرجع درون مجموعه امکان است، DMU_{j_0} کاراست اگر و تنها اگر $h_0 = 1$ و تمام متغیرهای کمکی برابر صفر باشند $(\forall s^+_r, s^-_i = 0)$ [۲۵].

معرفی مدل

۳-۳ مدل پیشنهادی روش ناپارامتری خروجیهای نامطلوب با اصل دسترس پذیری ضعیف

• دسترس پذیری ضعیف ورودی‌ها و متغیر پنهان برتولت و پان برای زمانی که تکنولوژی شامل خروجی‌های نامطلوب باشد، روشی را معرفی کردند که براساس اصول زیر ساخته می‌شود [۱۷].

فرض کنید در یک فعالیت تولیدی، P ورودی در تولید، Q خروجی مطلوب و R خروجی نامطلوب به کار برده شود. $U \in R^R_+$ ، $Y \in R^Q_+$ ، $X \in R^P_+$ بردار ورودی مصرف شده و به ترتیب نشان‌دهنده بردارهای خروجی‌های مطلوب و خروجی‌های نامطلوب باشند، همچنین $u \in R^P_+$ ، $v \in R^Q_+$

امروزه روش تحلیل پوششی داده‌ها به عنوان یک روش مناسب برای ارزیابی عملکرد واحدها شناخته شده است که در آن بر پایه اطلاعات موجود، مرز کارا به صورت تجربی برآورد می‌شود. از آنجایی که در دستیابی به تابع مرزی، همه داده‌ها پوشش داده می‌شوند، آن را تحلیل پوششی داده‌ها نام نهاده‌اند.

کارایی تکنیکی در روند تولید صنعت بیمه، ورودی‌ها را به خروجی تبدیل می‌کند (هزینه‌های اداری و بیمه‌گری و تعداد نمایندگی‌ها به عنوان ورودی و ارزش وام‌های پرداختی و درآمد حاصل از حق بیمه به عنوان خروجی مطلوب و ارزش خسارت پرداختی به عنوان خروجی نامطلوب). ارتباط بین ورودی‌ها و خروجی‌ها می‌تواند به صورت یک تابع تولید که بیشترین خروجی ممکن برای یک سطح از ورودی‌های ارائه شده را نشان می‌دهد بیان شود، کارایی تکنیکی یک اندازه‌گیری است که نشان می‌دهد، چگونه میزان حداکثر خروجی از ورودی‌های دسترس به دست می‌آید.

۳-۲ کارایی تکنیکی با چندین ورودی و چندین خروجی

فرض کنید، n شعبه بیمه وجود دارد $(j=1, 2, \dots, n)$ با m ورودی $(x_{ij}, i=1, 2, \dots, m)$ و s خروجی تولید می‌کند $(y_{rj}, r=1, 2, \dots, s)$. کارایی تکنیکی شعبه j_0 را در مقایسه با گروه همتا از n شعبه را به صورت زیر اندازه‌گیری می‌کند:

• کارایی تکنیکی: یک مدل برنامه ریزی خطی برای بازده به مقیاس متغیر (VRS)

$$\text{Max } h_0 = \sum_{r=1}^s u_r y_{rj_0} + u_0 \quad (1)$$

st:

$$\sum_{i=1}^m v_i x_{ij_0} = 1$$

$$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} + u_0 \leq 0, \quad j=1, 2, \dots, n$$

$$u_r, v_i \geq \varepsilon \quad i=1, \dots, m, r=1, \dots, s$$

u_0 free

در این مدل، علامت u_0 بازده به مقیاس را تعریف می‌کند: $u_0 > 0$ به معنای بازده به مقیاس افزایشی است؛ $u_0 = 0$

می‌شود [۲۸].

مالمکوئیست به صورت بیشترین فاکتور به وسیله ورودی‌هایی که روی یک دوره که می‌تواند کاهش یابد، تعریف می‌شود و تولید کردن همان خروجی در یک دوره دوم را می‌تواند محاسبه کند.

فرض کنید تکنولوژی تولید در دوره K ، به طوری که نمایش اصلی ضریب کاهش به این صورت باشد.

$$\lambda_j^k (X_j^L, Y_j^L, U_j^L) \quad (4)$$

به طوری که مقادیر هدف در دوره L هستند، به بیان مختصرتر با ساده‌سازی نماد اصلی را به $\lambda_j^k (U_j^L)$ نشان می‌-

دهیم، که این نشان می‌دهد اندازه‌گیری عملکرد برای خروجی نامطلوب در دوره L در حالی که با تکنولوژی مرجع دوره K محاسبه شده است.

به طور مشابه، $\alpha_j^k (X_j^L)$ اندازه‌گیری عملکرد برای ورودی‌ها در دوره L چنانچه تکنولوژی مرجع در دوره K محاسبه شده است.

در حالت کلی شاخص بهره‌وری مالم کوئیست (MPI) به دو مؤلفه تجزیه می‌شود:

تغییرات کارایی تکنیکی^{۱۸}

و تغییر مرز تولید کارایی^{۱۹}

که در واقع نماد اصلی را با مؤلفه‌های MPI در معادله (۳) ترکیب می‌کنیم. هم هدف کاهش α و هم متغیر پنهان λ ضرایب کاهشی برای این روش تجزیه هستند. وقتی از یک LVM متراکم ضرایب کاهش همبستگی نسبی ناشی می‌شود، که به نوبه خود تا به حداقل رساندن واریانس در این مدل کمک می‌کند.

$w \in R_+^R$ به ترتیب نشان‌دهنده قیمت ورودی‌ها، قیمت خروجی‌های مطلوب و قیمت خروجی‌های نامطلوب باشند و مجموعه مشاهدات با S نشان داده شود.

تمام اصول موضوعه، همراه یک فاکتور انقباضی برای خروجی‌های نامطلوب برای تشکیل مدل LV ضروری است. هیلو و ویمن ادعا کردند که اگر قید ورودی‌ها یا خروجی‌های نامطلوب را با فرم $X = \sum_{j=1}^J Z_j X_j$ تساوی جایگزین شود مدل DEA حاصل، دسترسی‌پذیری ضعیف را نشان می‌دهد.

VRS LV Min α :

$$\{\forall DMU \parallel j=1, 2, \dots, J: K, L\}$$

$$St. \sum_{j=1}^J Z_j X_{pj}^t = \alpha X_p^t, p=1, 2, \dots, P \quad (3)$$

$$\sum_{j=1}^J Z_j Y_{qj}^t \geq Y_q^t, q=1, 2, \dots, Q$$

$$\sum_{j=1}^J Z_j U_{rj}^t \leq \lambda U_r^t, r=1, 2, \dots, R$$

$$\sum_{j=1}^J Z_j = 1$$

$$Z$$

$$Z_j \geq 0$$

$$Latent Variable \cdot \leq \lambda = \frac{\sum_{j=1}^J Z_j U_{rj}^t}{U_r^t} \leq 1$$

$$\{\forall DMU \parallel r=1, 2, \dots, P\}$$

در این مقاله یک مدل جدید برای ارزیابی کارایی، خروجی‌ها به عنوان ورودی پیشنهاد شده است، با در نظر گرفتن یک فاکتور انقباضی برای خروجی‌های نامطلوب، در اینجا با استفاده از شاخص مالم کوئیست به بررسی پراکندگی بین دوره‌ها نیز می‌پردازیم.

پس از به دست آورد مقادیر کارایی واحدهای تصمیم‌گیرنده، بهره‌وری را برای بازه زمانی سال‌های ۱۳۸۸-۱۳۸۷ و ۱۳۸۹-۱۳۸۸ با توجه به فرمول‌های زیر به دست می‌آوریم.

شاخص بهره‌وری مالم کوئیست برای ارزیابی تغییرات تکنولوژی و اثر تغییرات روی ورودی‌ها و خروجی‌ها استفاده

^{۱۸}Technical Efficiency Change

^{۱۹}Efficient Production Frontier shift

نشان‌دهنده مسأله تصمیم داده شده است به طوری که هر یک از گره‌های غیر برگ با یکی از متغیرهای تصمیم در ارتباط است، هر شاخه از گره غیر برگ با زیرمجموعه‌ای از مقادیر متغیر تصمیم‌گیری مربوطه در ارتباط است، و هر گره برگ با یک مقدار از متغیر هدف (متغیر وابسته) وابسته شده است.

برای هر برگ درخت مرتبط با مقدار میانگین متغیر هدف است، بنابراین یک درخت، یک روش جایگزین برای مدل‌های خطی پیوسته برای مشکلات رگرسیون و مدل‌های خطی لجستیک برای مسائل طبقه بندی است.

به طور کلی، درختان C&R برخی مزایا بیش از مدل‌های رگرسیون دارند. اولاً، مدل ایجاد شده توسط یک درخت قابل فهم‌تر و نسبتاً ساده برای تفسیر غیرآماري است [۲۲].

ثانیاً، ذاتاً غیر پارامتری است یعنی بدون نیاز به هیچ فرضی با توجه به توزیع اساسی از مقادیر متغیرهای مستقل ساخته شده است. بنابراین، درختان C&R می‌توانند داده‌های عددی که چولگی زیاد و یا چند وجهی هستند و همچنین پیش‌بینی‌های رده‌ای با هر دو ساختار ترتیبی و غیرترتیبی را اداره کند.

ثالثاً، درختان C&R روش‌های پیچیده‌ای برای مقابله با متغیرهای از دست رفته در مقایسه با رگرسیون دارد.

در رگرسیون، داده‌هایی که دارای هر گونه مقدار از دست رفته باشند به طور خودکار حذف می‌کنیم.

بنابراین، درختان C&R می‌توانند حتی هنگامی که متغیرهای مستقل مهم برای برخی از DMUها ناشناخته است ایجاد شوند.

در نهایت درختان C&R، نسبتاً یک روش یادگیری ماشین خودکار است. درختان C&R، کارایی محاسباتی ارائه می‌کنند تا جایی که آنها زمان کمتری را در محاسبات و ذخیره‌سازی کمتری نیاز داشته باشند.

در تولید یک درخت C&R، مجموعه داده‌ها حداقل به دو بخش تقسیم می‌شوند: مجموعه داده‌های آموزشی و مجموعه داده‌های آزمایشی (عموماً به عنوان مجموعه داده‌های تست ارجاع داده می‌شود) [۲۲]. سپس تحت دو فاز فرایند اصلی: مرحله رشد و مرحله هرس قرار می‌گیرد [۲۶].

به طور کلی معرفی تکنولوژی متغیر پنهان تحلیل پوششی داده‌ها یک گام نخست به سوی نه تنها تحلیل خروجی‌های نامطلوب بلکه برای گنجاندن پیامدهای خارجی بر روی شرکت‌ها و جامعه است.

این مدل با استفاده از مدل کاهش متغیر پنهان متراکم، تئوری تولید کاهش همزمان خروجی‌های نامطلوب و ورودی‌ها را از طریق رابطه علی با ورودی‌ها نشان می‌دهد. نتایج بهره‌وری DMUها از رابطه (۵) به دست می‌آید، چنانچه $MPI > 1$ (progress) نشان‌دهنده روند پیشرفت در طی دوره و اگر $MPI = 1$ (No change) نشان‌دهنده عدم تغییر و اگر $MPI < 1$ (Regress) نشان‌دهنده پسرفت در طی دوره است.

۳-۴- ترکیب DEA با تکنیک داده‌کاوی

محققان DEA در گذشته عمدتاً بر روی ارزیابی و کنترل عملکرد متمرکز شده‌اند و تنها تلاش اندکی از ترکیب DEA با تکنیک داده‌کاوی گزارش شده است.

برای مثال، سون و مون (۲۰۰۴) امکان استفاده از درخت تصمیم‌گیری را با DEA برای پروژه‌های R&D مورد بررسی قرار دادند، وقتی که یک شرکت سعی بر انتقال یا گسترش تکنولوژی جدید دارد.

مطالعات بسیار اندکی برای DEA برای اهداف پیشگويانه به کار رفته است از جمله، پیش‌بینی شکست بانک (بار و همکاران، ۱۹۹۴) و پیش‌بینی شکست اتحادیه اعتباری (پایل و پاردی، ۱۹۹۷). اما هیچ مطالعه‌ای پیرامون DEA و درخت C&R برای کارایی و بهره‌وری شعب بیمه گزارش نشده است.

۳-۵- طبقه بندی و درخت رگرسیون (C&R)

تکنیک داده‌کاوی به DMUها اجازه کشف و کاوش معنادار اطلاعاتی که قبلاً در پایگاه داده‌های بزرگ پنهان بوده‌اند را می‌دهد.

C&R درخت تصمیم‌گیری است که به طور معمول در داده‌کاوی مورد استفاده قرار می‌گیرد، به وسیله براینم و همکاران در سال ۱۹۸۴ توسعه داده شده است و سپس توسط رایپلی در سال ۱۹۹۶ بهبود یافت. ساختار درختی

^{۲۰} Sohn & Moon

سومین مؤلفه از یک درخت C&R مجموعه داده آموزش است. این یک پایگاه داده شامل مقادیر متغیرهایی برای هم متغیر خروجی و هم متغیرهای پیش‌بینی‌کننده (مستقل)، از یک گروه از DMUهایی که می‌خواهیم نتایج خروجی را پیش‌بینی کنیم.

مؤلفه چهارم از درخت C&R، آزمون یا مجموعه داده‌های بیشتر که متشکل از DMUهایی است که ما خواهان پیشگویی دقیق آنها هستیم. این مجموعه داده‌های آزمون ممکن است در عمل وجود نداشته باشد در حالی که به طور معمول برای باوریم که یک مجموعه داده تست مورد نیاز است تا یک طبقه یا قوانین تصمیم را بدان تنفیذ کنیم، یک پایگاه داده آزمایش جداگانه همیشه نیاز به تعیین عملکرد یک قوانین تصمیم ندارد.

شکل ۱ مراحل ارزیابی کارایی و بهره‌وری شعب بیمه با استفاده از DEA/C&R را نشان می‌دهد. همانطور که در شکل مشاهده می‌شود، ابتدا از DEA برای اندازه‌گیری کارایی و بهره‌وری هر شعبه با ۳ ورودی (هزینه‌های اداری، هزینه‌های بیمه‌گری، تعداد نمایندگی‌ها) و ۳ خروجی (درآمد حاصل از حق بیمه، ارزش وام‌های پرداختی، ارزش خسارت پرداختی) استفاده شده است، به عنوان نتایج این مرحله را به سه گروه شعب کارآمد، ناکارآمد و بدون تغییر تقسیم می‌کنیم.

بنابراین در مرحله دوم تحلیل درخت C&R از عوامل محیطی وابسته به سیستم بیمه همچون قدمت شعب، درجه شعب، تعداد بیمه‌نامه‌های صادره و.... به عنوان عوامل ورودی به تحلیل درخت C&R و نمرات بهره‌وری بدست آمده در مرحله نخست به عنوان خروجی درخت C&R (برای لیست کامل متغیرها به جدول ۱ مراجعه کنید).

واضح است که، این یک چارچوب کلی می‌تواند برای هرگونه تجزیه و تحلیل در بخش بیمه و بانک و یا بخش‌های دیگر کاربرد داشته باشد.

در مرحله اول این روش، نمرات بهره‌وری DMUها به عنوان متغیر هدف (مستقل) وارد درخت رگرسیون می‌شوند. سپس عوامل محیطی در مرحله دوم به عنوان متغیرهای پیش‌بینی (وابسته) از یک طرف و نتایج بهره‌وری در مرحله اول از طرف دیگر وارد درخت رگرسیون می‌شوند و در نهایت

در مرحله رشد C&R، یک درخت از مجموعه داده‌های آموزشی ساخته می‌شود. در این مرحله، هر گره برگ با یک کلاس در ارتباط است با تقسیم‌بندی بیشتری از برگ مفروض در تعدادی از موارد به یک یا دو گره بعدی که در زیر بعضی آستانه مشخص شده، منجر می‌شود.

در این مرحله، درخت C&R برخلاف صحت یا مجموعه داده‌های آموزشی برطبق تولید یک زیر درخت با کمترین نرخ خطای مجموعه داده‌ها آزمایشی ارزیابی شده است. برای الگوریتم دقیق درختان C&R، هند و همکاران (۲۰۰۱) را می‌بینیم.

در این مقاله، از تحلیل درخت C&R برای کشف و بررسی عوامل درونی و بیرونی همچون تعداد خسارت‌های پرداختی، تعداد و تحصیلات کارکنان، درجه شعب، قدمت شعب، تعداد وام‌ها، تعداد بیمه‌نامه‌ها و... که بر روی بهره‌وری شعب بیمه مؤثرند، استفاده می‌شود.

مقدار هدف برای درخت رتبه‌کارایی و بهره‌وری به دست آمده از DEA است. بنابراین، DMUها به دو دسته گروه‌های پیشرفت بهره‌وری و پسرفت بهره‌وری تقسیم می‌شود و بنابراین هدف درخت ما یک متغیر گسسته (طبقه‌ای) است.

۳-۶- ترکیب DEA با درخت C&R

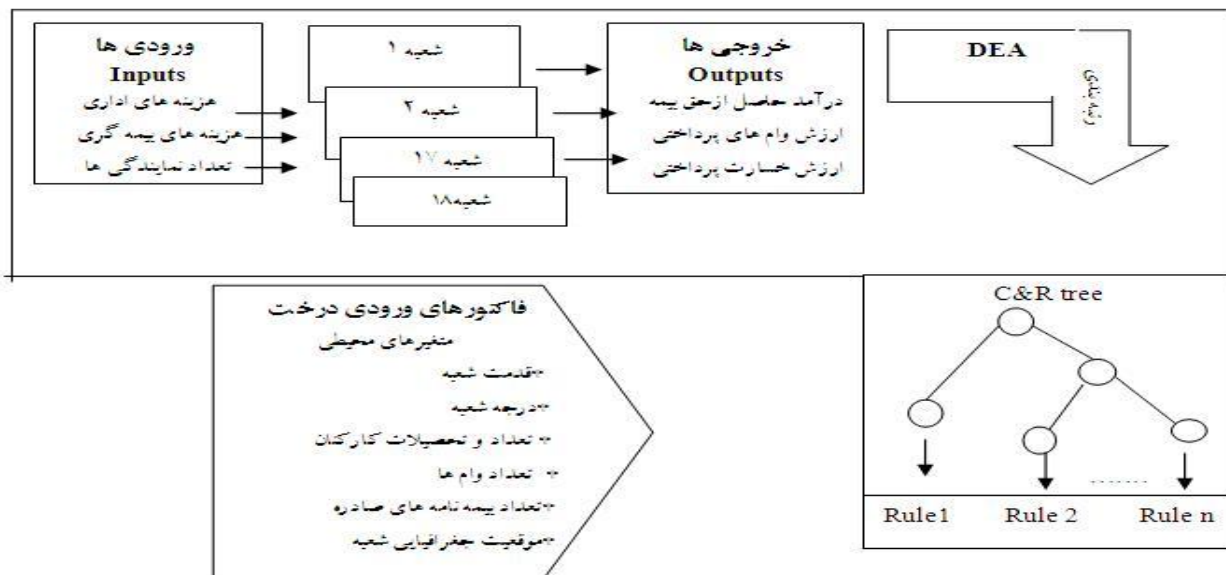
درخت C&R که در این مقاله پیشنهاد شده است، شامل چهار مؤلفه اصلی است.

مؤلفه اول، متغیر خروجی یا متغیر وابسته است. به طور کلی این متغیر مشخصه‌ای است که امیدواریم بتوانیم بر اساس متغیرهای مستقل یا پیش‌بینی‌کننده‌ها، پیش‌بینی کنیم. در این مطالعه متغیر خروجی همان نمرات به دست آمده از بهره‌وری است که به سه دسته پیشرفت بهره‌وری ($1 > \text{هدف}$)، پسرفت ($1 < \text{هدف}$)، بدون تغییر ($1 = \text{هدف}$) دسته‌بندی می‌شود.

مؤلفه دوم از درخت C&R، متغیرهای مستقل یا پیش‌بینی‌کننده است، تعدادی متغیرهای مستقل ممکن بسته به هدف از تجزیه و تحلیل وجود دارد.

در این مطالعه موردی، متغیرهای مستقل عوامل درونی و بیرونی هستند که در جدول ۱ ذکر شده است.

مجموعه قوانینی از درخت استخراج می‌شود.



شکل ۱: DEA/C&R متدلوژی برای ارزیابی شعب بیمه ایران

جدول ۱: عوامل ورودی در درخت تصمیم (۱۳۸۷-۱۳۸۹)

متغیر	نوع متغیر	مینیمم	ماکزیمم	میانگین	انحراف استاندارد (Std)
قدمت شعبه	عددی	۱۳۸۶	۱۳۲۴	۱۳۷۲	۱۲,۹۳۶
درجه شعبه	قیاسی	۱	۳		
موقعیت جغرافیایی	قیاسی	۱	۵		
تعداد پرسنل	عددی	۱۰	۱۵۹	۴۸,۶۶	۳۹,۰۶
تخصیلات کارکنان	قیاسی	۱	۶		
جنسیت کارکنان	قیاسی	۱	۲		
تعداد وام‌ها	عددی	۲	۱۰۰۷	۳۵۴,۹۶۳	۳۲۵,۱۹۱
تعداد بیمه نامه‌های صادره	عددی	۱۲	۶۸۹۴	۱۲۵۱,۵۶	۱۷۸۰,۵۱۱
تعداد خسارت پرداختی	عددی	۶	۱۷۹	۲۶۳,۴۶۳	۲۴۷,۷۲۴۲

بهره‌وری (Regress) ($\text{Target} < 1 \rightarrow \text{MPI} < 1$) و بدون تغییر ($\text{Target} = 1$) ($\text{MPI} = 1$) (No change) تقسیم می‌شوند.

یک درخت C&R دقیق، به یک پایگاه داده بزرگ نیاز دارد؛ بنابراین چون نمونه ما تنها شامل ۱۸ شعبه بیمه است، پس با افزایش پایگاه داده اصلی با ۱۰۰ بار استفاده از تکنیک نمونه‌برداری مجدد بوت استرپینگ که توسط افرون و تیبشیرانی توصیف شد است، استفاده می‌شود [۱۶].

بنابراین در مرحله دوم، به طور تصادفی ۱۸ واحد (با جایگزینی) انتخاب می‌کنیم و این نمونه‌ها را ۱۰۰ بار تا به دست آوردن ۱۸۰۰ واحد تکرار می‌کنیم. بعد از نمونه‌برداری مجدد^۱ ۱۰۰ بار پایگاه داده اصلی، پایگاه داده به دو گروه آموزشی و آزمایشی به نسبت ۷ به ۳ تقسیم می‌شود.

در مرحله سوم با استفاده از نمرات بهره‌وری طبقه‌بندی شده (> 1 ، $= 1$ ، < 1) به عنوان متغیرهای هدف درخت C&R و دیگر متغیرهای غیرقابل کنترل مانند درجه شعب، موقعیت جغرافیایی، تعداد کارکنان و... به عنوان ورودی درخت تصمیم قوانین منطقی استخراج می‌شود.

۱. درجه شعبه: ۱، مجتمع؛ ۲، ممتاز؛ ۳، درجه ۱
 ۲. موقعیت جغرافیایی شعبه: ۱، شمال؛ ۲، جنوب؛ ۳، شرق؛ ۴، غرب؛ ۵، مرکز
 ۳. تحصیلات کارکنان: ۱، زیردیپلم؛ ۲، دیپلم؛ ۳، فوق دیپلم؛ ۴، لیسانس؛ ۵، فوق لیسانس؛ ۶، دکترا
 ۴. جنسیت کارکنان: ۱، زن؛ ۲، مرد
 ۴. یافته‌های پژوهش

۴-۱- روش بوت استرپینگ DEA/C&R برای ارزیابی شعب بیمه

یکی از مشکلات استفاده از DEA/C&R این است که در بسیاری از مطالعات DEA، داده‌های کافی برای تولید درخت تصمیم در دسترس وجود ندارد. بنابراین، روش Bootstrapping زیر برای افزایش تعداد DMUها قبل از ایجاد درخت تصمیم C&R پیشنهاد شده است [۱۵].

این روش از سه مرحله تشکیل شده است، مرحله اول: مقادیر کارایی و بهره‌وری را برای هر شعبه محاسبه نموده و بر طبق این نمرات شعب را به سه گروه پیشرفت بهره‌وری (Progress) ($\text{Target} > 1 \rightarrow \text{MPI} > 1$)، پسرفت

جدول ۲: متغیرهای ورودی/خروجی در مدل DEA

متغیر	مینیمم	ماکزیمم	میانگین	انحراف استاندارد
ورودی‌ها				
هزینه های بیمه گری	۳۵۸۳۳.۳۳	۵۴۵۳۴۲۵۵۴۴	۵۳۷۲۷۳۹۱۵	۹۷۶۳۹۳۲۳۷
هزینه های اداری	۷۱۶۶۶.۶۷	۲۶۵۸۳۷۸۵۱۶۷	۲۱۴۶۹۷۰۵۵۲	۴۴۴۰۷۹۶۸۰۷
تعداد نمایندگی ها	۲	۲۷۱	۹۲.۱۲۹۶۳	۵۴.۲۸۹۲۵
خروجی‌ها				
درآمد حاصل از حق بیمه	۳۲۵۶۰۰۰	۳۷۰۷۰۶۴۵۸۱۱	۴۶۱۸۹۰۲۷۶۹	۶۴۰۸۱۹۶۴۹۵
ارزش وام های پرداختی	۱۹۵۵۷۱۰۰	۱۸۵۰۵۴۷۹۱۰۰	۵۸۲۴۷۷۸۲۲۹	۵۳۲۵۶۸۱۱۰۶
ارزش خسارت های پرداختی	۸۲۷۰۴۰۰	۲۰۴۰۰۲۱۰۳۰۰	۳۴۹۱۵۱۸۴۶۴	۳۶۴۸۱۶۴۱۷۶

^۱ Re-sampling

انتخاب تصادفی ۱۸ واحد با جایگذاری و تکرار این نمونه‌ها ۱۰۰ بار یک پایگاه داده با ۱۸۰۰ واحد به دست آمد. که این اطمینان را برای رسیدن به دقت بالاتری در پیش‌بینی درخت C&R به ما خواهد داد.

• مرحله سوم: تحلیل C&R
بر طبق ارزیابی DEA شعب بیمه به سه گروه، پیشرفت بهره‌وری ($MI > 1$)، پسرفت بهره‌وری ($MI < 1$) و بدون تغییر بهره‌وری ($MI = 1$) تقسیم می‌شود. این گروه‌ها، به عنوان متغیر هدف در درخت C&R مورد استفاده قرار می‌گیرند نمرات کارایی و بهره‌وری شعب با شاخص مالمکوویست در جدول ۳ نشان داده شده است.

در طی بررسی روند بهره‌وری در سال‌های ۱۳۸۸-۱۳۸۷، ۵ واحد روند پیشرفت و ۱۳ واحد روند پسرفت داشته‌اند و در طی سال‌های ۱۳۸۹-۱۳۸۸، ۸ واحد بهره‌ور و ۱۰ واحد عدم بهره‌وری را نشان می‌دهند.

به طور میانگین، ۳۶٪ واحدها دارای پیشرفت بهره‌وری و ۶۶٪ واحدها پسرفت بهره‌وری را نشان می‌دهند. البته به دلیل پراکندگی بالای مقادیر داده‌های ورودی درخت تمام مقادیر قبل از ورود به درخت نرمال شده و سپس وارد درخت شده است تا سبب کاهش دقت پیش‌بینی نشود.

• مرحله دوم: بوت استرپینگ (Bootstrapping)
همان‌طور که در بخش قبل به طور کامل توضیح داده شد، با

جدول ۳: نمرات بهره وری به دست آمده از مدل LVM با شاخص مالکوییست

واحد	از سال	تا سال	TC	EC	MI
DMU ^۱	۱۳۸۷	۱۳۸۸	۰,۰۱۲۹	۱۱,۷۲۳۶	۰,۱۵۱۲
DMU ^۱	۱۳۸۸	۱۳۸۹	۲,۴۰۷۷	۰,۱۱۱۱	۰,۲۶۷۵
DMU ^۲	۱۳۸۷	۱۳۸۸	۰,۰۰۶۳	۴۰,۴۵۳۱	۰,۲۵۴۹
DMU ^۲	۱۳۸۸	۱۳۸۹	۰,۹۰۶۷	۰,۴۰۹۷	۰,۳۷۱۵
DMU ^۳	۱۳۸۷	۱۳۸۸	۰,۰۱۲۲	۱۳۲,۲۴۶۹	۱,۶۱۳۴
DMU ^۳	۱۳۸۸	۱۳۸۹	۰,۹۰۷۵	۰,۲۹۹۷	۰,۲۷۲
DMU ^۴	۱۳۸۷	۱۳۸۸	۰,۰۱۲۳	۳,۰۹۴۶	۰,۰۳۸۱
DMU ^۴	۱۳۸۸	۱۳۸۹	۱,۰۵۱۱	۰,۴۴۸۳	۰,۴۷۱۲
DMU ^۵	۱۳۸۷	۱۳۸۸	۰,۰۱۰۳	۲۱,۲۲۴۳	۰,۲۱۸۶
DMU ^۵	۱۳۸۸	۱۳۸۹	۰,۹۴۶۱	۰,۲۵۶۵	۰,۲۴۲۷
DMU ^۶	۱۳۸۷	۱۳۸۸	۰,۰۰۶۳	۱۱,۹۶۰۶	۰,۰۷۵۴
DMU ^۶	۱۳۸۸	۱۳۸۹	۲,۵۴۶	۰,۶۳۷	۱,۶۲۲۷
DMU ^۷	۱۳۸۷	۱۳۸۸	۰,۰۰۸۵	۴,۵۷۹۲	۰,۰۳۸۹
DMU ^۷	۱۳۸۸	۱۳۸۹	۱,۰۸۳۹	۰,۲۳۲۸	۰,۲۵۲۳
DMU ^۸	۱۳۸۷	۱۳۸۸	۰,۰۰۰۵	۱۰,۰۳۵۴	۰,۰۰۰۲
DMU ^۸	۱۳۸۸	۱۳۸۹	۲,۲۵۵	۰,۷۹۶۴	۱,۷۹۵۳
DMU ^۹	۱۳۸۷	۱۳۸۸	۰,۰۰۰۵	۳,۹۵۲۵	۰,۰۱۹۸
DMU ^۹	۱۳۸۸	۱۳۸۹	۰,۹۵۳۳	۰,۵۳۴۷	۰,۵۰۹۷
DMU ^{۱۰}	۱۳۸۷	۱۳۸۸	۰,۰۰۰۸	۱۱,۷۱۷۷	۰,۰۹۳۷
DMU ^{۱۰}	۱۳۸۸	۱۳۸۹	۱,۶۰۹۲	۰,۲۶۴۶	۰,۴۲۵۸
DMU ^{۱۱}	۱۳۸۷	۱۳۸۸	۰,۰۰۷۴	۰,۴۹	۰,۰۰۳۶
DMU ^{۱۱}	۱۳۸۸	۱۳۸۹	۱,۰۳۶۲	۰,۲۵۷۳	۰,۲۶۶۶
DMU ^{۱۲}	۱۳۸۷	۱۳۸۸	۰,۰۱۱۸	۱,۱۷۷۴	۰,۰۱۳۹
DMU ^{۱۲}	۱۳۸۸	۱۳۸۹	۱,۲۴۲۱	۱,۳۸۶۷	۱,۷۲۲۵
DMU ^{۱۳}	۱۳۸۷	۱۳۸۸	۰,۰۰۵۵	۰,۲۵۹۹	۰,۰۰۱۴
DMU ^{۱۳}	۱۳۸۸	۱۳۸۹	۱,۱۹۵۵	۱,۲۰۵۹	۱,۴۴۱۷
DMU ^{۱۴}	۱۳۸۷	۱۳۸۸	۰,۰۱۱۴	۱۴,۶۹۷۷	۰,۱۶۷۶
DMU ^{۱۴}	۱۳۸۸	۱۳۸۹	۱,۲۴۴۵	۱,۲۸۸	۱,۶۰۲۳
DMU ^{۱۵}	۱۳۸۷	۱۳۸۸	۰,۰۱۰۷	۱۱,۰۴۸	۱,۱۸۲
DMU ^{۱۵}	۱۳۸۸	۱۳۸۹	۱,۳۷۹	۰,۲۳۱۷	۰,۴۵۷۴
DMU ^{۱۶}	۱۳۸۷	۱۳۸۸	۰,۰۱۰۳	۹۷۵,۷۳۹۶۸	۱۰,۵۰۱۲
DMU ^{۱۶}	۱۳۸۸	۱۳۸۹	۳,۱۹۵۲	۲,۰۶۳۵	۸,۰۳۷۷
DMU ^{۱۷}	۱۳۸۷	۱۳۸۸	۰,۰۰۸۲	۳۲۴,۸۱۸۱۸	۲,۶۶۳۵۱
DMU ^{۱۷}	۱۳۸۸	۱۳۸۹	۲,۰۸۳۵	۲,۰۶۳۵	۶,۳۶۲۸
DMU ^{۱۸}	۱۳۸۷	۱۳۸۸	۰,۰۱	۲۰۳,۱۴۷	۲,۰۳۱۴۷
DMU ^{۱۸}	۱۳۸۸	۱۳۸۹	۳,۰۲۹۹	۰,۳۹۳۹	۱,۱۹۳۵

۵. بحث و نتیجه‌گیری

با توجه به انتخاب درخت C&R، با شاخصی به نام (Impurity) بر حسب کمترین مقدار ناخالصی در هر گره، برای تمام خصیصه‌ها^{۲۲} میزان ناخالصی را قبل و بعد از شاخه‌زدن محاسبه نموده و خصیصه‌ای که ناخالصی را بیشتر کاهش دهد را انتخاب می‌کند. به این ترتیب چندین درخت تصمیم‌گیری رگرسیون، برای هر دوره رسم می‌شود.

۵-۱- تحلیل نتایج درخت رگرسیون طی دوره (۱۳۸۷-۱۳۸۸)

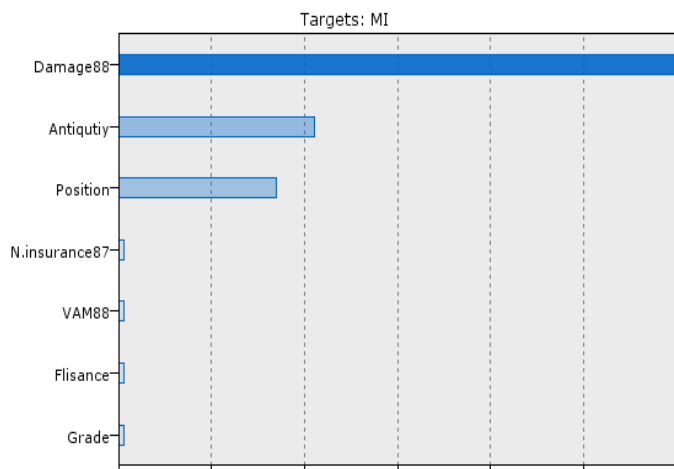
ابتدا درخت پیش‌بینی را برای سال (۱۳۸۷-۱۳۸۸) با متغیرهای، درجه شعبه، قدمت، موقعیت مکانی، تعداد کارمندان فوق لیسانس و دکترا، تعداد وام‌های این دوره، تعداد بیمه‌نامه‌های صادر شده در این دوره، تعداد خسارت‌های پرداختی در این دوره، به عنوان ورودی و طبقه‌بندی بهره‌وری به عنوان خروجی درخت در نظر گرفته می‌شود. توجه داشته باشید که یکی از ویژگی‌های ذاتی درخت، حذف برخی از ویژگی‌ها برحسب اهمیت یا حداقل همبستگی است، بنابراین درخت تنها با متغیرهای تعداد بیمه‌نامه‌های صادره، قدمت و موقعیت مکانی به عنوان متغیرهای محیطی وارد می‌کنیم.

شکل ۲، اهمیت متغیرهای محیطی را نشان می‌دهد. در این شکل مشاهده می‌کنیم، که تعداد خسارت‌های پرداختی دارای بیشترین اهمیت در تعیین طبقه‌بندی را دارد (۵۹٪) و قدمت با (۲۱٪) و موقعیت جغرافیایی (۱۷٪) در جایگاه دوم و سوم و بقیه متغیرها با (۵٪) در مقام‌های بعدی اهمیت قرار دارند، اما چون متغیرهای بعدی در جایگاه یکسان اهمیت قرار دارند تا رسیدن به نتیجه مطلوب و دقت بیشتر درخت را با متغیرهای هم‌رده دوباره می‌سازیم. جدول ۳-۶ دقت پیش‌بینی درخت تولید شده را نشان می‌دهد. به واسطه ۱۸۰۰ مورد:

۱۲۹۰ مورد در طی سال ۱۳۸۸-۱۳۸۷، شاخص بهره‌وری زیر یک و ۵۱۰ مورد بالای یک را پیش‌بینی می‌کند و از ۱۲۴۶ مورد داده برای آزمایش (Training) داده‌ها صورت گرفته است، ۱۱۵۶ مورد به صورت صحیح پیش‌بینی شده است و با

دقت ۹۲٫۷۸٪ این پیش‌بینی همراه است و ۹۰ مورد به اشتباه پیش‌بینی شده است. سطح دقت کلی درخت پیش‌بینی C&R، ۹۸٫۰۲٪ است، که سطح اطمینان بالایی را نشان می‌دهد. در شکل ۳ درخت تولید شده C&R، با ۸ گره نمایش داده شده است.

Variable Importance



شکل ۲: اهمیت متغیرهای محیطی درخت (۱۳۸۷-۱۳۸۸)

جدول ۴: دقت پیش‌بینی درخت تولید شده (شکل ۶-۲)

Results for output field MI

Comparing \$R-MI with MI

Partition'	1_Training	2_Testing
Correct	1,226 98.39%	541 97.65%
Wrong	20 1.61%	13 2.35%
Total	1,246	554

^{۲۲} Attributes

باشد، آنگاه شعبه دارای پیشرفت بهره‌وری است (Progress)؛ (۷۲ مورد).

۵-۲- استخراج قوانین برای شعب با پسرفت بهره‌وری

شعب با روند پسرفت بهره‌وری (در مجموع ۹۱۴ مورد از ۱۲۴۶)، اگر قوانین به صورت زیر دنبال شود:

- قانون ۴: اگر تعداد خسارت پرداختی بزرگتر مساوی ۰,۰۲۰ و قدمت شعبه کمتر مساوی سال ۱۳۸۰ و موقعیت مکانی شعبه (مرکز، شرق، شمال و جنوب) باشد، آنگاه شعبه دارای روند پسرفت بهره‌وری است (Regress)؛ (۹۱۲ مورد).

۵-۳- تحلیل نتایج درخت رگرسیون طی دوره (۱۳۸۸-۱۳۸۹)

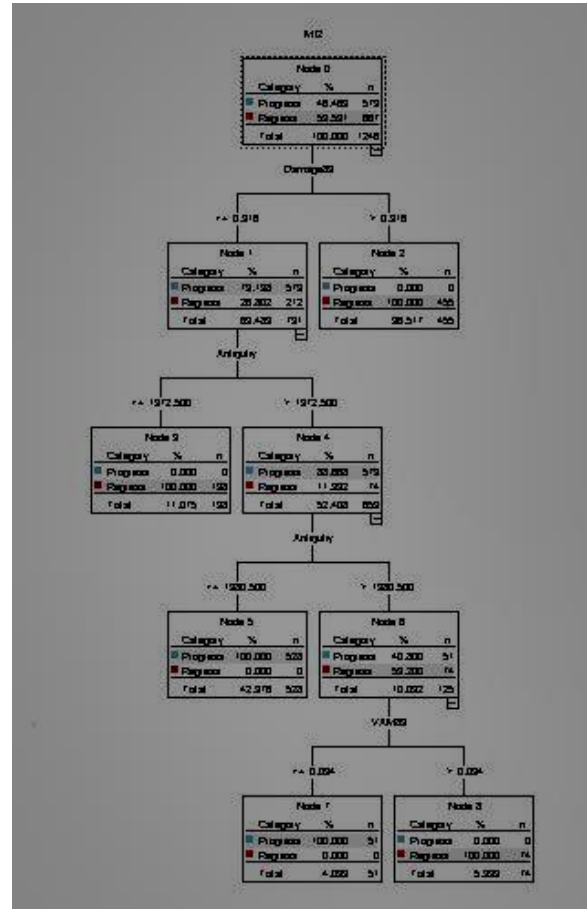
جدول ۵ دقت پیش‌بینی درخت تولید شده را نشان می‌دهد. به واسطه ۱۸۰۰ مورد:

۹۹۷ مورد در طی سال ۱۳۸۸-۱۳۸۹، شاخص بهره‌وری زیر یک و ۸۰۳ مورد بالای یک را پیش‌بینی می‌کند و از ۱۲۴۶ مورد داده برای آزمایش (Training) داده‌ها صورت گرفته است. کل ۱۲۴۶ مورد به صورت صحیح پیش‌بینی شده است و با دقت ۱۰۰٪ این پیش‌بینی همراه است.

سطح دقت کلی درخت پیش‌بینی C&R، ۱۰۰٪ است، که سطح اطمینان بالایی را نشان می‌دهد.

در شکل ۵ درخت تولید شده C&R، با ۵۷۹ مورد پیشرفت بهره‌وری و ۶۶۷ مورد کاهش بهره‌وری، با ۸ گره نمایش داده شده است.

با توجه به نمودار تشکیل شده در شکل ۴ که اهمیت متغیرهای ورودی را در درخت رگرسیون نشان می‌دهد، تعداد خسارات پرداختی در سال ۱۳۸۹ با (۶۷٪) در جایگاه اول و قدمت شعبه با (۳۰٪) اهمیت در جایگاه دوم و بقیه متغیرها (تعداد کارمندان با تحصیلات دکترا و فوق لیسانس، درجه شعبه و تعداد بیمه‌نامه‌های صادره در سال ۸۹) با (۶٪) در جایگاه بعدی با اهمیت یکسان قرار دارند.



شکل ۳: درخت تولید شده با متغیرهای محیطی (۱۳۸۸-۱۳۸۷)

بر طبق این درخت قوانین زیر استخراج می‌شود:

استخراج قوانین برای شعب با پیشرفت بهره‌وری

شعب با روند پیشرفت بهره‌وری (در مجموع ۳۲۵ مورد از ۱۲۴۶)، اگر قوانین به صورت زیر دنبال شود:

- قانون ۱: اگر تعداد خسارت پرداختی کوچکتر مساوی ۰,۰۲۰ باشد، آنگاه شعبه دارای پیشرفت بهره‌وری است (Progress)؛ (۱۹۲ مورد).

• قانون ۲: اگر تعداد خسارت پرداختی هر شعبه بزرگتر از ۰,۰۲۰ و قدمت تأسیس مربوط به کمتر مساوی از سال ۱۳۸۰ باشد و موقعیت مکانی در غرب شهر تهران باشد، آنگاه شعبه دارای پیشرفت بهره‌وری است. (Progress)؛ (۵۰ مورد).

• قانون ۳: اگر تعداد خسارت پرداختی شعب بیشتر از ۰,۰۲۰ و قدمت تأسیس مربوط به بیشتر از سال ۱۳۸۰

استفاده از درخت C&R می‌تواند درک بهتری از نتایج به دست آمده از تحلیل پوششی داده‌ها را نشان دهد و با بررسی عوامل مربوط به کارایی و بهره‌وری و عدم بهره‌وری که یک مسئله مهم برای سیاست‌گذاران است، در نظر گرفته شود.

همچنین ما روش پیشنهادی را در صنعت بیمه مورد مطالعه قرار داده، اما روش پیشنهادی در این مطالعه به طور بالقوه دارای کاربردهای بسیار گسترده‌تری است.

روش پیشنهادی DEA/C&R یک چارچوب برای هر مطالعه دیگری است که می‌تواند برای ارزیابی کارایی و بهره‌وری واحدهای تصمیم‌گیرنده (DMUs) مورد استفاده قرار گیرد.

نتایج مدل ترکیبی، یک مجموعه قوانینی است که می‌تواند توسط سیاست‌گذاران برای کشف دلایل کارایی و ناکارایی واحدهای تصمیم‌گیرنده مورد استفاده قرار گیرد. اگرچه، برخی از محدودیت‌ها، زمانی که DEA با C&R ترکیب می‌شود، وجود دارد.

ایجاد یک درخت خوب و قابل اطمینان C&R، معمولاً به یک پایگاه داده بزرگ با مشاهدات زیادی نیازمند است. در اغلب گزارشات DEA در ادبیات موضوع، تعداد واحدهای تصمیم‌گیرنده (DMUs) به اندازه کافی بزرگ نیست که بتواند یک درخت C&R را تغذیه کند.

در این مطالعه، روش بوت استرپینگ (Bootstrapping) را برای رفع این مشکل پیشنهاد کردیم؛ با این حال به نظر می‌رسد که این موضوع نیاز به تحقیقات بیشتری برای برنامه‌های کاربردی دارد.

۷-۵- نتیجه‌گیری

تحلیل پوششی داده‌ها (DEA)، یک ابزار مدیریتی برای ارزیابی کارایی و بهره‌وری است. این مقاله به معرفی یک چارچوب ترکیب DEA با طبقه‌بندی و تحلیل رگرسیون می‌پردازد.

در حالی که، DEA نتایج قابل قبول و ارزشمندی ارائه می‌دهد، از طرفی تحلیل C&R، مبتنی بر تحلیل یافته‌های حقایقی اضافی که در مطالعات قبلی نامشخص بودند را برای ما آشکار می‌کند. برخلاف مطالعات قبل در حوزه تحلیل پوششی داده‌ها و صنعت بیمه که به شناسایی تأثیر یکسانی

۵-۴- استخراج قوانین برای شعب با پیشرفت بهره‌وری (۵۷۹ مورد)

• قانون ۱: اگر تعداد خسارات در سال ۱۳۸۹ کمتر مساوی ۰,۳۱۶ باشد و قدمت تأسیس مربوط به سال‌های بعد ۱۳۷۲ و قبل ۱۳۸۰ باشد، آنگاه شعبه دارای روند پیشرفت بهره‌وری است (Progress)؛ (۵۲۸ مورد).

• قانون ۲: اگر تعداد خسارت در سال ۱۳۸۹ کمتر مساوی ۰,۳۱۶ باشد و قدمت تأسیس مربوط به سال‌های بعد ۱۳۸۰ باشد و تعداد وام‌های پرداختی کمتر مساوی ۰,۰۳۴ باشد، شعبه دارای پیشرفت بهره‌وری است (Progress) (۵۱ مورد).

۵-۵- استخراج قوانین برای شعب با پسرفت بهره‌وری (۶۶۷ مورد)

• قانون ۳: اگر تعداد خسارت پرداختی سال ۱۳۸۹ کمتر مساوی ۰,۳۱۶ باشد و قدمت تأسیس شعبه مربوط به سال‌های قبل از ۱۳۷۲ باشد، شعبه دارای روند کاهشی یا پسرفت بهره‌وری است (Regress)؛ (۱۳۸ مورد).

• قانون ۴: اگر تعداد خسارت پرداختی سال ۱۳۸۹ کمتر مساوی ۰,۳۱۶ و قدمت تأسیس شعبه مربوط به سال‌های بعد از ۱۳۷۲ باشد و تعداد وام‌های پرداختی بیشتر از ۰,۰۳۴ باشد، شعبه دارای روند کاهشی یا پسرفت بهره‌وری است (Regress)؛ (۷۴ مورد).

• قانون ۵: اگر تعداد خسارت پرداختی بیش از ۰,۳۱۶ پس شعبه دارای روند پسرفت بهره‌وری است (Regress)؛ (۴۵۵ مورد).

۵-۶- ارزیابی نهایی

به طور خلاصه این مقاله، به معرفی ترکیب DEA و رویکرد درخت C&R می‌پردازد. مطالعه موردی ما در این مطالعه شعب بیمه ایران در شهر تهران می‌باشد.

به طور کلی، تحلیل پوششی داده‌ها نمرات کارایی و با استفاده از شاخص مالکویست، نمرات بهره‌وری را می‌توان محاسبه نمود، اما این روش نمی‌تواند عوامل مربوط به ناکارایی و عدم بهره‌وری را توضیح دهد، خصوصاً اگر این عوامل در شکل متغیرهای غیرعددی مانند درجه شعب باشند.

قوانین برای سیاست‌گذاران بیمه بسیار مفید و روند تصمیم‌گیری آنان را تحت‌الشعاع قرار می‌دهد.

۵-۸- پیشنهادات آتی

تعدادی موضوعات فراتر از محدوده تحقیق ما وجود دارد که از لحاظ اهمیت کاربردی برای اشخاصی که با درخت C&R به تحلیل و ارزیابی موضوعات می‌پردازند، بسیار حائز اهمیت است. این عوامل عبارتند از فاکتورهای مستقل برای بخش‌های بیمه و استفاده از قوانین مختلف و اندازه‌گیری دقیق همچنین بهبود روش بوت استرپینگ، این موارد می‌تواند زمینه را برای توسعه مطالعات آینده گسترده‌تر کند. در تحقیقات آتی می‌توان، پایگاه داده‌ای با رکوردهای بالا انتخاب نمود که مجبور به استفاده از بوت‌استرپینگ برای تولید پایگاه داده معتبر نبوده و به راحتی با دقت بالا داشته باشیم، البته استفاده از شبیه‌سازی داده‌ها در این تحقیق جزء محدودیت‌های تحقیق در زمینه دسترسی به داده‌های مورد نظر بوده است.

همچنین می‌توان از درخت تصمیم فازی به جای درخت تصمیم در حالت دقیق^{۲۷} استفاده نمود که برای داده‌های کیفی صنعت بیمه، نتایج مفیدی را ارائه می‌دهد و یا خروجی‌های درخت را نتایج کارایی و ناکارایی مدل DEA قرار داد و یا بسته به نوع داده‌ها و اهمیت متغیرهای ورودی از درختان دیگری چون C4.5، C5، CART و... استفاده نمود.

منابع

۱. آذر عادل؛ مؤمنی منصور. ۱۳۸۳. اندازه‌گیری بهره‌وری در شرکت‌های تولیدی به وسیله مدل‌های تحلیل پوششی داده‌ها، دو ماهنامه علمی پژوهشی دانشور رفتار، دانشگاه شاهد، سال یازدهم، شماره ۸.
۲. امامی میبیدی علی. ۱۳۸۴. اصول و اندازه‌گیری کارایی و بهره‌وری (علمی- کاربردی)، مؤسسه مطالعات پژوهش‌های بازرگانی، چاپ دوم، تهران.
۳. اسفاندرانی حمید. ۱۳۹۰. طراحی شبکه فروش بیمه‌های عمر (مورد مطالعاتی شرکت سهامی بیمه ایران به روش

از عوامل بر روی کارایی می‌پردازد، درخت C&R، مبتنی بر تحلیل پیشنهادهایی است که سطح^{۲۳} تأثیر عوامل وابسته به کارایی و بهره‌وری را بر روی شعب نشان می‌دهد.

با بررسی اهمیت متغیرها با نرم‌افزار کلمنتاین^{۲۴} و تأثیر قدرت آنها در وابستگی متغیرها با کمترین میزان ناخالصی تا رسیدن به گره هدف، این روند می‌تواند به تفکیک دقیق عوامل محیطی با نتایج کارایی و بهره‌وری شعب که از مرحله قبل و با روش DEA به دست آمده، تحلیلی قابل تعمق و با کمترین میزان خطا را در پی داشته باشد.

با بررسی مطالعات گذشته بر روی صنعت بیمه، تاکنون تنها به ارزیابی شعب بر اساس شاخص‌های مؤثر و تنها به تحلیل کارایی یا ناکارایی و یا بهره‌وری و عدم بهره‌وری در طی دوره پرداخته شده است و به کشف عوامل محیطی مرتبط با پیشرفت بهره‌وری (Progress) و یا پسرفت بهره‌وری (Regress) پرداخته نشده است.

به عنوان مثال، با این که تعداد خسارت و تعداد وام‌های پرداختی و قدمت تأسیس شعبه به عنوان شاخص مهم بیمه در روند کارایی و ناکارایی محسوب نمی‌شود، اما همین طور که از قوانین استخراج شده استنباط می‌کنیم، به عنوان عوامل وابسته در این روند با سطح اهمیت متفاوت دخیل هستند.

از طرف دیگر، توانستیم به استخراج قوانین برای هر واحد تصمیم‌گیرنده (DMU) با استفاده از متغیرهای عددی^{۲۵} و رده‌ای^{۲۶} که دارای درصدی از اهمیت بودند، پردازیم و برای شناسایی بهره‌وری و عدم بهره‌وری در انتخاب شعب استفاده کنیم.

برخلاف مطالعات قبلی کاربردهای DEA، که تمرکز فقط در زمینه عددی برای محاسبه کارایی و بهره‌وری تصمیم‌گیری می‌شد، در این مطالعه با استفاده از درخت C&R، به توسعه عوامل مرتبط با بهره‌وری و عدم بهره‌وری می‌پردازد و هرقانونی که می‌تواند برای یک واحد بهره‌ور به دست آید با استفاده از متغیرهای عددی و - متغیرهای قیاسی را استخراج کند. بدیهی است که این

^{۲۳}Level

^{۲۴}clementine

^{۲۵}Numerical

^{۲۶}Categorical

^{۲۷} Crisp

۱۳. Charnes A., W. W. Cooper, Rhodes E. ۱۹۷۸. Measuring the Efficiency, European Journal of Operations Research, No ۲.
۱۴. Chen, Y., Cook, W.D., Li, N., Zhu, J. ۲۰۰۹. Additive efficiency decomposition in two stage DEA. European Journal of Operational Research.
۱۵. Emrouznejad A., Anouze A. ۲۰۱۰. Data envelopment analysis with classification and regression tree-a case of banking efficiency, Expert Systems.
۱۶. Efron, B. and Tibishirani R. ۱۹۹۳. An Introduction to the Bootstrap, New York: Chapman and Hall.
۱۷. Fakhari A., Eftekhari Moghadam A.M. ۲۰۱۳. Combination of classification and regression in decision tree for multi-labeling image annotation and retrieval, Applied Soft Computing.
۱۸. Fan Ch.K, & Cheng, Sh.W. ۲۰۰۹. Using Analytic Hierarchy Process Method and Technique for Order Preference by Similarity to Ideal Solution to Evaluate Curriculum in Department of Risk Management and Insurance, J. Soc. Sci., ۱۹(۱).
۱۹. Farzipoor R. ۲۰۰۷. Suppliers selection in the presence of both cardinal and ordinal data, European Journal of Operational Research.
۲۰. Emrouznejad A., DEA Home page <http://www.deazone.com/tutorial>.
۲۱. Han J., Kamber M. 2006. Data Mining Concept and Techniques. 2nd Edition. San Francisco, Elsevier.
۲۲. Hand D.J., Manilla H., Smyth P. ۲۰۰۱. Principles of Data, Cambridge, MA: MIT Press.
۲۳. Hosseini Bamakan, S. M. Gholami, P. ۲۰۱۴. A Novel Feature Selection Method Based on an Integrated Data Envelopment Analysis and Entropy Model, ۲nd International Conference on Information Technology and Quantitative Management, ITQM ۲۰۱۴, Procedia Computer Science ۳۱, ۶۳۲ - ۶۳۸.
۲۴. Hwang S., Kao T.L. ۲۰۰۷. Measuring Managerial Efficiency in Non-Life Insurance Companies: An Application of Two-Stage Data Envelopment Analysis, International Journal of a Management, vol. 1, No. 3.
۲۵. Jahanshahloo GR., Alirezaee MR. ۱۹۹۲. Measuring the efficiency of academic units at the Teacher Training University, Proceedings of 26th Annual Iranian math conference.
- تحلیل پوششی داده‌ها (DEA)، پایان‌نامه کارشناسی‌ارشد، دانشگاه تربیت مدرس.
۴. دعایی حبیب ا.؛ نیکخواه فرخانی زهرا. ۱۳۸۸. ارزیابی عملکرد عملیاتی و منابع انسانی نمایندگی‌های بیمه کارافرین در استان خوزستان با نگرش چندگانه به روش تحلیل پوششی داده‌ها، فصلنامه صنعت بیمه، سال بیست و چهارم، شماره ۴۳، پاییز و زمستان، شماره مسلسل ۹۶-۹۵.
۵. سلطان‌پناه هیرش و همکاران. ۱۳۸۶. ارزیابی کارایی شعب بیمه البرز با استفاده از تحلیل پوششی داده‌ها، فصلنامه صنعت بیمه، سال بیست و دوم، شماره ۴، زمستان، شماره مسلسل ۸۸، ۱۵۱-۱۷۷.
۶. مشیری سعید؛ رضوان مهدی. ۱۳۸۵. اثر به‌کارگیری فناوری ارتباطات و اطلاعات در کارایی صنعت خدمات هوایی ایران، فصلنامه پژوهش‌های اقتصادی ایران، سال هشتم، ش ۲۶.
۷. مهرگان محمدرضا. ۱۳۸۳. مدل‌های کمی در ارزیابی عملکرد سازمان‌ها، انتشارات دانشگاه تهران.
۸. Babazadeh, R. Razmi, J. Rabbani, M. Pishvae, M. S. ۲۰۱۵. An integrated data envelopment analysis-mathematical programming approach to strategic biodiesel supply chain network design problem, Journal of Cleaner Production (Article in press).
۹. Barros, C. P., Nektarios, M. & Assaf, A. ۲۰۱۰. Efficiency in the Greek insurance industry, European Journal of Operational Research DEA technology. Omega 41, PP. ۳۱۵-۳۲۵.
۱۰. Bretholt A, Pan J. 2013. Evolving the latent variable model as a an environmental ۹.
- Charnes, A, Cooper, WW, Rhodes, E. ۱۹۷۸. Measuring the efficiency of decision making units, European journal of operational research, ۲, ۴۲۹-۴۴۴.
۱۱. Barr, R., L.M. Siford and T.F. Simes. ۱۹۹۴. Forecasting bank failure: a non-parametric approach, Recherches Economiques de Louvain.
۱۲. Breiman, L., J. Friedman, R. Olshenand C. Stone. ۱۹۸۴. Classification and Regression Trees, Pacific Grove, CA: Wadsworth-Monterey.

۳۳. Shahroudi K., Taleghani M., Mohammadi G. ۲۰۰۱. Efficiency Decomposition in Data Envelopment Analysis: An application to Insurance companies in Iran.
۳۴. Sueyoshi T., Goto P. ۲۰۰۹. DEA-discriminate analysis Methodological comparison among eight discriminant analysis approaches, *European Journal of Operational Research*.
۳۵. Tavana, M., Keramatpour, M., Santos-Arteaga, F.J., Ghorbaniane, E. ۲۰۱۵. A Fuzzy Hybrid Project Portfolio Selection Method Using Data Envelopment Analysis, TOPSIS and Integer Programming, *Expert Systems with Applications* (Article in press).
۳۶. Torgo L. ۱۹۹۷. Functional models for regression tree leaves, *Proceedings of the 14th International Conference on Machine Learning*.
۳۷. Wang, C. H. Chuang, J. J. ۲۰۱۵. Integrating decision tree With back propagation network to conduct business diagnosis and performance simulation For solar companies, *Decision Support Systems* (Article in press).
۳۸. Xie X. ۲۰۱۰. Are publicly held firms less efficient? Evidence from the US property-liability insurance industry, *Journal of Banking & Finance*.
۳۹. Yao, Sh., Han, zh., & Feng, G. ۲۰۰۷. On technical efficiency of China's insurance industry after WTO accession, *China Economic Review*.
۲۶. Kim, H. and G.J. Koehler. ۱۹۹۵. Theory and practice of decision tree induction, Omega.
۲۷. Lee, S. ۲۰۱۰. Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C controls, *Decision Support Systems*, ۴۹, ۴۸۶-۴۹۷.
۲۸. Luhnen, M. ۲۰۰۹. Efficiency and Competition in Insurance Markets, *Dissertation no. ۳۶۷۵*.
۲۸. Mahlberg, B & Url, Th. ۲۰۱۰. Single Market effects on productivity in the German insurance industry, *Journal of Banking & Finance*.
۲۸. Malmquist, S. ۱۹۵۳. Index numbers and indifference surfaces. *Trabajos de Estadística*.
۲۹. Nagurur, N. N. Rajbhanari, B. ۲۰۰۱. Data envelopment analysis for the performance evaluation of air conditioning and refrigeration companies in Thailand, *Business Performance Management*.
۳۰. Park, J. Lee, D.S. Christakis, N, and Barabasi, A.L.. ۲۰۰۹. The impact of cellular networks on disease comorbidity, *Molecular Systems Biology*.
۳۱. Pille, P. and Paradi J. ۱۹۹۷. Facets at the frontier and efficiency measurement in DEA, Paper presented at the Fifth European Workshop on Efficiency and Productivity Analysis.
۳۲. Seol, H. Choi, J. Park, G. Park, Y. ۲۰۰۷. A framework for benchmarking service process using data Envelopment analysis and decision tree, *Expert Systems with Applications*, ۳۲, ۴۳۲-۴۴۰.

درک اعتماد اولیه کاربران به شبکه‌های اجتماعی

* محسن اکبری

** کامران زاهدفر

*** زهرا ایاغ

* استادیار و عضو هیات علمی دانشگاه گیلان

** دانشجوی دکتری مدیریت بازرگانی، سیاست‌گذاری بازرگانی، دانشگاه تهران

*** کارشناس ارشد مدیریت بازرگانی، دانشگاه گیلان

تاریخ دریافت: ۹۴/۰۲/۲۲ تاریخ پذیرش: ۹۴/۰۴/۰۹

چکیده

شبکه‌های اجتماعی یکی از انواع رسانه‌های اجتماعی محسوب می‌شوند که در سال‌های اخیر در زندگی کاربران اینترنتی آنچنان تأثیرگذار بوده اند که شکل‌های ارتباطات اجتماعی نیز از این رسانه‌های جدید تأثیر پذیرفته است و در آینده نقش به مراتب مهم‌تری را بازی خواهند کرد. از سوی دیگر اعتماد یکی از عوامل اصلی ایجاد رضایت و تکرار استفاده‌ی کاربران از خدمات برخط است. لذا این پژوهش مدل بسط احتمالی را برای شناخت عوامل مؤثر بر اعتماد اولیه کاربران به شبکه‌های اجتماعی (فیس‌بوک) به کار گرفته است و بیان می‌کند که کاربران از طریق دو مسیر مرکزی و جانبی نگرش خود را تغییر می‌دهند. نشانه‌های مرکزی شامل کیفیت اطلاعات و کیفیت خدمات و نشانه‌های جانبی شامل کیفیت سیستم، امنیت ساختاری و شهرت هستند. خودکارآمدی اثرات نشانه‌های مرکزی و نشانه‌های جانبی را بر اعتماد اولیه تعدیل می‌کند. جامعه آماری این پژوهش، کاربران ایرانی فیس‌بوک می‌باشند. داده‌های مورد نیاز پژوهش از طریق ارتباط با صفحات شخصی افراد و صفحات عمومی (مربوط به گروه‌ها، شهرها، فعالیت‌ها، دانشگاه‌ها و ...) فعال در این شبکه‌ی اجتماعی جمع‌آوری گردید. بدین منظور لینک پرسشنامه‌ی آنلاین طراحی شده بر روی docs.google.com در اختیار آنان قرار گرفت. در این پرسشنامه از مقیاس ۵ گزینه‌ای لیکرت از کاملاً مخالفم (۱) تا کاملاً موافقم (۵) استفاده شد و در پایان تعداد ۴۰۶ پرسشنامه جمع‌آوری گردید که با توجه به جدول جرسی و مورگان، برای انجام پژوهش در جامعه‌ی بی‌نهایت، مطلوب می‌باشد. روایی پرسشنامه از طریق نظرخواهی از خبرگان و اساتید دانشگاه و پایایی آن از طریق محاسبه‌ی آلفای کرونباخ تأمین گردید و نشان دهنده‌ی حد مطلوب پایایی است. داده‌ها با نرم افزار آماری SmartPLS^۲ تحلیل شد. طبق نتایج به دست آمده از میان متغیرهای مستقل این پژوهش خودکارآمدی، شهرت و امنیت ساختاری به ترتیب با ضریب مسیر ۰/۴۷، ۰/۲۳ و ۰/۲۰ بیشترین اثر را بر اعتماد اولیه دارند و کیفیت سیستم کمترین اثر را بر اعتماد اولیه با مقدار ۰/۰۷ دارد. همچنین نتایج بیانگر این مطلب بوده که خودکارآمدی تنها اثر کیفیت سیستم را بر اعتماد اولیه تعدیل می‌کند. بنابراین از آن جایی که خودکارآمدی جدای از نقش تعدیل‌گریش به طور مستقیم نیز بیشترین تأثیر را بر اعتماد اولیه دارد؛ طراحان شبکه‌های اجتماعی باید به این عامل توجه ویژه‌ای داشته باشند؛ به گونه‌ای که یادگیری آن آسان و فرآیند استفاده از آن برای کاربران واضح و قابل فهم باشد.

واژه‌های کلیدی: اعتماد اولیه، خودکارآمدی، فیس‌بوک، مدل بسط احتمالی

۱- مقدمه

اینترنت به سرعت تبدیل به بخشی از زندگی روزمره شده است [۱]. رشد ارتباطات الکترونیکی اثرات قابل توجهی بر فعالیت‌های روزمره‌ی انسان دارد. کارشناسان این حوزه سعی می‌کنند این تکنولوژی را برای تسهیل امور روزانه به کار بگیرند؛ به طوری که صاحبان صنایع، سازمان‌های خدماتی و سایر مراکز قادر به برقراری ارتباط با مشتریان خود در اولین زمان و با کمترین هزینه باشند [۲]. شبکه‌های اجتماعی مانند فیس‌بوک در سال‌های گذشته به طور گسترده‌ای مورد پذیرش واقع شده اند؛ چرا که نقش مهمی در پر کردن مرزهای بین فردی و تشویق ارتباطات بین المللی ایفا می‌کنند و به افراد اجازه می‌دهند تا ایده‌ها و فعالیت‌هایشان را در صفحات شخصی خود به اشتراک بگذارند [۳].

اما عواملی مانند، آزار و اذیت و سرقت اطلاعات شخصی نگرانی‌های عمیقی در میان کاربران فیس‌بوک ایجاد کرده و سبب شده تمایل زیادی به افشای اطلاعات شخصی در شبکه‌های اجتماعی وجود نداشته باشد [۴]. با وجود افزایش تعداد دوستداران فیس‌بوک و نقش آن در ترویج فرهنگ کشورها (البته به شرطی که از آن درست استفاده شود)، نقش اعتماد اولیه در فیس‌بوک تا حد زیادی نادیده گرفته شده است. هدف از این مطالعه بررسی دلایل اعتماد کاربران به شبکه‌های اجتماعی خارج از کشور است. در صورتی که تعداد زیادی از شبکه‌های داخلی نتوانسته است که اعتماد کاربران را جلب نمایند. تحقیقات موجود نیز غالباً بر اعتماد مصرف‌کنندگان در حوزه تجارت الکترونیک متمرکز شده اند [۵، ۶، ۷] و یا اعتماد کاربران به بانکداری موبایل را مطالعه کرده‌اند [۸]. برای درک بهتر اعتماد اولیه کاربران فیس‌بوک، این پژوهش با توجه به مدل به کار گرفته شده بیان می‌کند که افراد از طریق دو مسیر مرکزی و جانبی نگرش خود را تغییر می‌دهند. نشانه‌های مرکزی عبارتند از کیفیت اطلاعات و کیفیت خدمات و نشانه‌های جانبی کیفیت سیستم، امنیت ساختاری و شهرت هستند. اینکه کاربران نشانه‌های مرکزی یا نشانه‌های جانبی را انتخاب کنند توسط مدل بسط احتمالی تعیین می‌شود که به انگیزش و توانایی آنها بستگی دارد.

۲- ادبیات نظری و پیشینه‌ی پژوهش

۲-۱- اعتماد اولیه^۱

اعتماد اولیه به عنوان یک کانال عاطفی/شناختی تأثیر بالایی بر روابط بین کاربران و وب سایت‌ها دارد؛ چرا که مدل ذهنی کاربران بر اساس این اعتماد اولیه شکل می‌گیرد. شکل گیری اعتماد اولیه به دلیل فراگیر بودن دارای اهمیت زیادی می‌باشد [۹]. سه عامل توانایی، خیرخواهی و صداقت در اعتماد به یک گروه یا یک سازمان مؤثر هستند. توانایی به معنی مهارت‌ها و شایستگی‌های ارائه دهنده‌گان خدمات است. صداقت بیان می‌کند که ارائه‌دهندگان خدمات باید از اصول اخلاقی و معنوی مورد پذیرش کاربران پیروی کنند. خیرخواهی بیانگر حسن نیت ارائه دهنده‌گان خدمات نسبت به کاربران می‌باشد [۱۰].

مطالعات مختلف عوامل تأثیرگذار بر اعتماد را شناسایی نموده‌اند؛ از جمله آن‌ها می‌توان به شهرت، سودمندی وبسایت، سهولت استفاده، کنترل‌های امنیتی [۷، ۴۳، ۴۷]، خودکارآمدی [۱۲]، کیفیت خدمات [۱۳، ۱۴، ۱۵]، کیفیت سیستم [۱۶، ۴۳] و کیفیت اطلاعات [۱۱، ۴۷] اشاره نمود. مک‌نایت و همکارانش نیز سه عامل برای ایجاد اعتماد پیشنهاد کردند: امنیت ساختاری، شهرت وب و کیفیت وبسایت. به اعتقاد آن‌ها هر سه عامل به طور قابل توجهی اعتماد را تحت تأثیر قرار می‌دهد. این عوامل، به ویژه کیفیت وبسایت و شهرت، اهرم قدرتمندی هستند که ارائه‌دهندگان خدمات برای ایجاد اعتماد می‌توانند از آن‌ها استفاده نمایند تا بر احساسات منفی در مورد ایمنی محیط وب غلبه کنند [۵].

۲-۲- مدل بسط احتمالی^۲

مدل بسط احتمالی یک نظریه در مورد فرآیندهای فکری است که در سال ۱۹۸۱ توسط پتی و کاسیوپو برای توضیح

^۱ Initial trust^۲ Elaboration Likelihood Model (ELM)

فیس‌بوک به کار می‌بریم و انتظار داریم که فرآیندهای توسعه اعتماد را مشخص نماید.

۳-۲ - نشانه‌های مرکزی^۷

کیفیت اطلاعات^۸ با مسائلی مانند بهنگام بودن، کامل بودن، دقت، صحت و ارتباط داده‌های تولید شده توسط یک سیستم اطلاعاتی مرتبط است [۲۰، ۲۱، ۲۲]. فقدان اطلاعات مربوط ممکن است اعتماد کاربر را به توانایی‌های ارائه‌دهنده کاهش دهد [۲۳] و بر عکس ارائه‌ی اطلاعات شفاف و کامل به افزایش اعتماد به وب سایت کمک می‌کند [۲۴]. از آنجایی که کاربران نیاز به صرف زمان و تلاش زیاد در مورد درک کیفیت اطلاعات دارند؛ این متغیر ممکن است از طریق مسیر مرکزی اعتماد اولیه را تحت تأثیر قرار دهد. لذا پیش‌بینی می‌کنیم:

فرضیه ۱: کیفیت اطلاعات تأثیر مثبتی بر اعتماد اولیه دارد. کیفیت خدمات^۹ تابعی از تفاوت بین انتظارات و ادراکات مشتریان از خدمات واقعی است [۲۵]. پاراسورامان و همکاران^{۱۰} در سال ۱۹۸۸، ۵ بعد را برای کیفیت خدمات شامل ملموس بودن^{۱۱}، قابلیت اطمینان^{۱۲}، پاسخگویی^{۱۳}، امنیت^{۱۴} و همدلی^{۱۵} پیشنهاد نمودند. کیفیت خدمات یک ساختار پایدار است که شامل عملکرد کیفیت در تمامی فعالیت‌های انجام شده است [۲۶]. چون کاربران نیاز به صرف زمان و هزینه برای ارزیابی کیفیت خدمات دارند؛ کیفیت خدمات ممکن است به عنوان یک نشانه مرکزی

نظریه‌های متضاد، با پیشنهاد راه‌هایی برای تغییر نگرش در منبع و پیام مطرح شد. این نظریه، یک نظریه روان‌شناسی اجتماعی است که به طور گسترده‌ای برای تغییر رفتار افراد و همچنین تحول در تبلیغات به کار می‌رود. این نویسندگان دو خط‌مشی برای تغییر نگرش توصیف نمودند: مسیر مرکزی^۳ و مسیر جانبی^۴. مسیر مرکزی نیاز به اطلاعات و استدلال زیاد دارد؛ در حالی که مسیر جانبی از استنباط و ابتکارات برای تغییر نگرش استفاده می‌کند [۱۷]. با این حال معمولاً تغییرات ادراک و نگرش از طریق مسیر مرکزی با ثبات‌تر و پایدارترند و پیش‌بینی بهتری از رفتارهای بلندمدت نسبت به مسیر جانبی دارند. بر اساس این نظریه دریافت‌کنندگان اطلاعات می‌توانند به طور گسترده‌ای توانایی و انگیزه خود را شرح و بسط دهند که به نوبه خود فرآیندهای شکل‌گیری و یا تغییر نگرش را محدود می‌کند [۱۸]. یی و همکاران^۵ این نظریه را به منظور بررسی شکل‌گیری اعتماد در اطلاعات بهداشتی مبتنی بر وب به کار گرفتند و پیشنهاد کردند که دو عامل کیفیت استدلال و تخصص منبع به واسطه‌ی کیفیت اطلاعات و ریسک بر اعتماد تأثیر می‌گذارند [۱۹]. باتاچرج و سانفورد^۶ برای بررسی پذیرش فناوری اطلاعات از این نظریه استفاده نمودند و بیان کردند کاربرانی که ارتباط شغلی و تخصص نسبتاً بالا در استفاده از فناوری اطلاعات دارند تمایل دارند که از مسیر مرکزی استفاده کنند؛ در حالی که کسانی که ارتباط شغلی و تخصص کمی در فناوری اطلاعات دارند بر نشانه‌های جانبی برای تصمیم‌گیری در پذیرش فناوری اطلاعات تکیه می‌کنند [۱۸]. بنابراین در این پژوهش نظریه بسط احتمالی را به منظور آزمون اعتماد کاربران در

^۷. Central cues

^۸. Information quality

^۹. Service quality

^{۱۰}. Parasuraman and et al

^{۱۱}. Tangibles

^{۱۲}. Reliability

^{۱۳}. Responsiveness

^{۱۴}. Assurance

^{۱۵}. Empathy

^۳. Central route

^۴. Peripheral route

^۵. Yi and et al

^۶. Bhattacharjee and Sanford

عمل کند. کیم و همکارانش^{۱۶} بیان کردند که کیفیت خدمات بر اعتماد تأثیر می‌گذارد [۲۷]. لذا پیش‌بینی می‌کنیم:

فرضیه ۲: کیفیت خدمات تأثیر مثبتی بر اعتماد اولیه دارد.

۴-۲- نشانه‌های جانبی^{۱۷}

کیفیت سیستم^{۱۸} با اشکالات موجود در سیستم، نرخ پاسخگویی سیستم و سهولت استفاده مرتبط است [۲۰]. در طراحی یک وب‌سایت باید مقدار اطلاعات نمایش داده شده، رنگ و گرافیک مد نظر قرار بگیرد [۲۸، ۴۸]. کیفیت سیستم ممکن است بر اعتماد اولیه از طریق مسیر جانبی تأثیر بگذارد. چون کاربران به آسانی نشانه‌های اطلاعاتی مربوط به کیفیت سیستم را به دست می‌آورند. کیفیت سیستم به طور قابل توجهی اعتماد کاربران را تحت تأثیر قرار می‌دهد [۲۹، ۳۰]. لذا پیش‌بینی می‌کنیم:

فرضیه ۳: کیفیت سیستم تأثیر مثبتی بر اعتماد اولیه دارد.

شهرت^{۱۹} یک دارایی با ارزش است که به سرمایه‌گذاری بلند مدت و تلاش فراوان نیاز دارد [۳۱] و به این معنی است که افراد تا چه اندازه ارائه دهنده‌ی خدمات را صالح، درستکار و خیرخواه می‌دانند [۳۲]. اطلاعات دست دوم مانند بازخورد دوستان و تبلیغات دهان به دهان^{۲۰} می‌تواند بر اعتماد کاربران تأثیر بگذارد. تحقیقات نشان داده که شهرت و تبلیغات دهان به دهان در شبکه‌های اجتماعی فرآیند اعتماد را تحت تأثیر قرار می‌دهد. مشتریانی که تجربه قبلی نداشته باشند؛ بر شهرت ارائه‌دهنده‌ی تکیه می‌کنند [۲۴]. شهرت اثر خود را به عنوان نشانه‌ی جانبی اعمال می‌کند. اثرات مثبت و قابل توجهی از شهرت درک شده بر اعتماد مشاهده شده است [۲۴، ۳۱، ۴۹]. لذا پیش‌بینی می‌کنیم:

فرضیه ۴: شهرت تأثیر مثبتی بر اعتماد اولیه دارد. امنیت ساختاری^{۲۱} به معنی برداشت‌های افراد از ایمنی محیط وب می‌باشد و به طور قابل توجهی اعتماد افراد را تحت تأثیر قرار می‌دهد [۵]. در واقع امنیت ساختاری مکانیزم اعتماد مبتنی بر نهاد است و از سنت جامعه‌شناختی به وجود می‌آید [۳۴]. جامعه‌شناسان دریافته‌اند که اعتماد در مردم توسط ساختارهای نهادی حمایت می‌شود تا محیطی امن و مطمئن را برای کاربران ایجاد کند [۵، ۳۳]. همچنین مهر و موم شخص ثالث^{۲۲} نیز بر اعتماد افراد تأثیر می‌گذارد [۳۵]. امنیت ساختاری ممکن است از طریق مسیر جانبی بر اعتماد اولیه تأثیر بگذارد [۸]. لذا پیش‌بینی می‌کنیم:

فرضیه ۵: امنیت ساختاری تأثیر مثبتی بر اعتماد اولیه دارد.

خودکارآمدی^{۲۳} یک جزء کلیدی در نظریه شناختی اجتماعی و به عنوان اعتقاد فرد به توانایی‌هایش برای سازماندهی و اجرای رفتارهای خاص به منظور دستیابی به اهداف مشخص تعریف شده است [۳۶] و نقش مهمی در کاهش عدم اطمینان و افزایش اعتماد کاربر ایفا می‌کند [۱۲]. در این پژوهش خودکارآمدی به معنی توانایی و مهارت‌های فرد برای کار با فیس‌بوک می‌باشد. لذا پیش‌بینی می‌کنیم که:

فرضیه ۶: خودکارآمدی تأثیر مثبتی بر اعتماد اولیه دارد. همان‌طور که گفته شد؛ مدل بسط احتمالی دو مسیر را برای تغییر نگرش توصیف می‌کند: مسیر مرکزی و مسیر جانبی. هنگامی که کاربران فیس‌بوک درک کنند که اطلاعات مربوط به رفتار آن‌ها است و یا اینکه توانایی بالایی دارند، نگرش خود را از طریق مسیر مرکزی و در غیر این صورت، از طریق مسیر جانبی تغییر می‌دهند. خودکارآمدی به عنوان توانایی درک شده کاربران ممکن

^{۱۶}. Kim and et al

^{۱۷}. Peripheral cues

^{۱۸}. System quality

^{۱۹}. Reputation

^{۲۰}. Word of mouth advertising

^{۲۱}. Structural assurance

^{۲۲}. Third-Party Seals

^{۲۳}. Self-efficacy

پرسشنامه‌ای است. در پژوهش حاضر، بنا به اهداف و روش اجرایی آن، جامعه آماری، کاربران ایرانی فیس‌بوک می‌باشند. این پژوهش با استفاده از روش نمونه‌گیری در دسترس صورت گرفت. داده‌های مورد نیاز پژوهش از طریق ارتباط با صفحات شخصی افراد و صفحات عمومی (مربوط به گروه‌ها، شهرها، فعالیت‌ها، دانشگاه‌ها و ...) فعال در این شبکه‌ی اجتماعی جمع‌آوری گردید. بدین منظور لینک پرسشنامه‌ی آنلاین طراحی شده بر روی تارنمای docs.google.com در اختیار آنان قرار گرفت. در این پرسشنامه از مقیاس ۵ گزینه‌ای لیکرت از کاملاً مخالفم (۱) تا کاملاً موافقم (۵) استفاده شد و در پایان تعداد ۴۰۶ پرسشنامه جمع‌آوری گردید که با توجه به جدول جرسی و مورگان، برای انجام پژوهش در جامعه‌ی بی‌نهایت، مطلوب می‌باشد. روایی^{۲۴} اولیه‌ی پرسشنامه از طریق نظرخواهی از خبرگان و اساتید دانشگاه مرتبط با این حوزه تأیید شد و پایایی^{۲۵} اولیه‌ی آن نیز از طریق محاسبه آلفای کرونباخ تأمین گردید که نتایج آن در جدول ۲ ذکر شده است و نشان دهنده‌ی حد مطلوب پایایی است.

ابعاد	فراوانی	درصد	ابعاد	فراوانی	درصد
جنسیت					
مرد	۱۸۳	۴۵/۱	زیر دبلم	۳	۰/۷
زن	۲۲۳	۵۴/۹	دبلم	۱۸	۴/۴
			تحصیلات		
			کارشناسی	۱۵۳	۳۷/۷
			کارشناسی ارشد	۲۱۶	۵۳/۲
			دکتری	۱۶	۳/۹
			س		
			≥ ۲۰	۲۱	۵/۲
			۲۴-۲۰	۲۰۸	۵۱/۲
			مدت		
			۲۹-۲۵	۱۴۱	۳۴/۷
			استفاده		
			۳۹-۳۰	۲۷	۶/۷
			بیش از ۳ سال	۳۱	۷/۶
			≥ ۴۰	۹	۲/۲

جدول ۱. مشخصات جمعیت شناختی نمونه پژوهش

است ایجاد اعتماد اولیه را تعدیل نماید. هنگامی که خودکارآمدی کاربران بالا باشد؛ ممکن است اعتماد را از طریق مسیر مرکزی و در غیر این صورت، از طریق مسیر جانبی شکل دهند.

فرضیه ۷: خودکارآمدی تأثیر کیفیت اطلاعات بر اعتماد اولیه را تعدیل می‌کند.

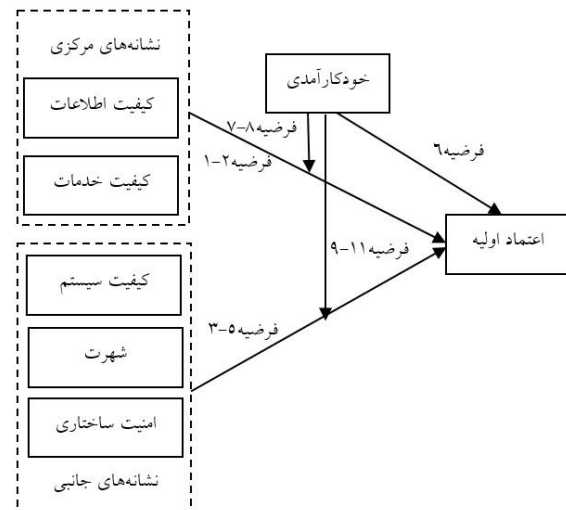
فرضیه ۸: خودکارآمدی تأثیر کیفیت خدمات بر اعتماد اولیه را تعدیل می‌کند.

فرضیه ۹: خودکارآمدی تأثیر کیفیت سیستم بر اعتماد اولیه را تعدیل می‌کند.

فرضیه ۱۰: خودکارآمدی تأثیر شهرت بر اعتماد اولیه را تعدیل می‌کند.

فرضیه ۱۱: خودکارآمدی تأثیر امنیت ساختاری بر اعتماد اولیه را تعدیل می‌کند.

شکل ۱ مدل مفهومی این پژوهش را نشان می‌دهد. نشانه‌های مرکزی کیفیت اطلاعات و کیفیت خدمات و نشانه‌های جانبی شامل کیفیت سیستم، شهرت و امنیت ساختاری هستند. خودکارآمدی تعدیل کننده‌ی اثرات نشانه‌های مرکزی و جانبی بر اعتماد اولیه می‌باشد.



شکل ۱. مدل مفهومی پژوهش

۳- روش‌شناسی پژوهش

این پژوهش از لحاظ هدف کاربردی، از لحاظ ماهیت و روش توصیفی و از لحاظ شیوه گردآوری داده‌ها کتابخانه‌ای و

^{۲۴}. Validity

^{۲۵}. Reliability

جدول ۱ مشخصات جمعیت شناختی نمونه پژوهش را نشان می‌دهد.

۴- تجزیه و تحلیل یافته‌ها

جهت سنجش روابط علی در مدل‌های دارای متغیر تعدیل‌گر، رویکرد حداقل مربعات جزئی^{۲۶} روشی مناسب است [۳۸]. از این رو در این پژوهش به منظور تحلیل داده‌ها از نرم‌افزار اسمارت پی. ال. اس^{۲۷} استفاده گردیده است. به دلیل اینکه مدل‌سازی مسیری PLS شاخصی برای نیکویی برازش^{۲۸} ارائه نمی‌دهد، قبل از بررسی فرضیه‌های پژوهش، مطلوب است بخش‌های بیرونی (مدل اندازه‌گیری^{۲۹}) و درونی (مدل ساختاری^{۳۰}) مدل پژوهش مورد ارزیابی قرار گیرد [۳۹].

جهت ارزیابی مدل‌های اندازه‌گیری، پایایی و روایی آنان مورد بررسی قرار گرفت. به منظور سنجش میزان پایایی از دو معیار تک بعدی بودن و پایایی معرف^{۳۱} استفاده شد. از این‌رو مقادیر دو ابزار آلفای کرونباخ و پایایی مرکب^{۳۲} برای بررسی تک بعدی بودن محاسبه شدند. مقدار مناسب برای آلفای کرونباخ بزرگتر از ۰/۷ [۴۰] و حداقل میزان قابل قبول آن ۰/۶ [۴۱] می‌باشد. همچنین برای پایایی مرکب مقادیر بالای ۰/۷ مناسب ذکر شده است [۴۲]. بنابراین با توجه به مقادیر آلفای کرونباخ و پایایی مرکب گزارش شده در جدول ۲، تک بعدی بودن و پایایی مناسب متغیرهای مکنون پژوهش تأیید شد.

در ادامه جهت تعیین پایایی معرف، بارهای عاملی کلیه معرف‌ها استخراج شد. مطابق جدول ۳ همه‌ی بارهای عاملی مدل حاضر (اعداد پر رنگ) مقادیر مناسب بالاتر از ۰/۴ را

میانگین	انحراف معیار	آلفای کرونباخ	پایایی ترکیبی	وارianس استخراج شده	میانگین
۲/۹۲	۰/۶۹	۰/۸۸	۰/۹۱	۰/۶۸	کیفیت اطلاعات
۴/۰۴	۰/۷۲	۰/۹۱	۰/۹۴	۰/۸۵	خودکارآمدی
۳/۳۴	۰/۷۱	۰/۶۱	۰/۷۹	۰/۵۶	شهرت
۳/۲۶	۰/۶۴	۰/۷۳	۰/۸۲	۰/۵۰	امنیت ساختاری
۳/۵۰	۰/۴۹	۰/۷۵	۰/۸۱	۰/۴۱	کیفیت خدمات
۳/۳۷	۰/۵۶	۰/۶۵	۰/۷۹	۰/۴۹	کیفیت سیستم
۳/۱۳	۰/۶۴	۰/۷۶	۰/۸۴	۰/۵۷	اعتماد اولیه

جدول ۲. مقادیر معیارهای پایایی و روایی

پس از گردآوری داده‌های مورد نیاز پژوهش نیز، روایی و پایایی پرسشنامه با چندین روش مورد بررسی و تأیید قرار گرفت که مشروح نتایج آن در قسمت تجزیه و تحلیل یافته‌ها گزارش شده است.

مدل پژوهش شامل هفت عامل است که هر عامل با چند سنجه و با طیف لیکرت پنج‌گزینه‌ای سنجیده می‌شود. سنجه‌های کیفیت اطلاعات بهنگام بودن، دقت، صحت، کامل بودن و قابلیت اطمینان از تبادل اطلاعات است [۲۱]. پنج بعد کیفیت خدمات شامل ملموس بودن، قابلیت اطمینان، پاسخگویی، امنیت و همدلی [۲۵] و اقلام کیفیت سیستم شامل سرعت دسترسی و سهولت استفاده [۲۰] می‌باشند. آیتم محبوبیت [۲۴] در سنجش شهرت در نظر گرفته شد. امنیت ساختاری شامل امنیت وب (حریم خصوصی)، [۳۷] و همچنین مهر و موم شخص ثالث [۳۵] است. اقلام خودکارآمدی برای اندازه‌گیری توانایی‌ها و مهارت کاربران [۳۶] می‌باشد و اعتماد اولیه با سنجه‌های امنیت ساختاری و شهرت وب سایت سنجیده می‌شود.

^{۲۶}. Partial Least Squares (PLS)

^{۲۷}. Smart PLS

^{۲۸}. Goodness of Fit (GOF)

^{۲۹}. Measurement Model

^{۳۰}. Structural Model

^{۳۱}. Indicator

^{۳۲}. Composite Reliability

درک اعتماد اولیه کاربران به شبکه‌های اجتماعی

مدل‌های اندازه‌گیری حاضر در مدل کلی پژوهش را پذیرفت.

اختیار نموده‌اند [۵۰] که نشانگر پایایی تمامی معرفها است. در مجموع با توجه به نتایج حاصل از بررسی تک بعدی بودن و پایایی، معرفها می‌توان ارزیابی، مطلوب

اعتماد	کیفیت سیستم	کیفیت خدمات	امنیت ساختاری	شهرت	خودکارآمدی	کیفیت اطلاعات	
۰/۳۸	۰/۳۹	۰/۲۶	۰/۶۹	۰/۱۶	۰/۲۷	۰/۲۱	امنیت ساختاری ۱
۰/۵۴	۰/۵۴	۰/۳۷	۰/۸۹	۰/۴۰	۰/۲۶	۰/۳۹	امنیت ساختاری ۲
۰/۵۲	۰/۵۰	۰/۳۴	۰/۸۹	۰/۳۴	۰/۲۶	۰/۳۴	امنیت ساختاری ۳
۰/۱۸	۰/۱۹	۰/۰۵	۰/۴۰	۰/۰۹	۰/۰۶	۰/۰۸	امنیت ساختاری ۴
۰/۳۸	۰/۲۳	۰/۳۰	۰/۵۴	۰/۲۳	۰/۲۱	۰/۲۰	امنیت ساختاری ۵
۰/۶۳	۰/۴۰	۰/۳۶	۰/۲۴	۰/۱۹	۰/۸۹	۰/۲۶	خودکارآمدی ۱
۰/۶۸	۰/۴۱	۰/۳۵	۰/۳۰	۰/۱۹	۰/۹۳	۰/۲۸	خودکارآمدی ۲
۰/۶۹	۰/۴۲	۰/۳۴	۰/۳۳	۰/۲۴	۰/۹۳	۰/۳۰	خودکارآمدی ۳
۰/۳۷	۰/۳۳	۰/۵۶	۰/۲۲	۰/۲۴	۰/۲۴	۰/۳۵	کیفیت خدمات ۱
۰/۵۲	۰/۵۲	۰/۷۱	۰/۲۷	۰/۳۳	۰/۳۸	۰/۴۶	کیفیت خدمات ۲
۰/۳۰	۰/۲۰	۰/۴۴	۰/۱۶	۰/۲۸	۰/۱۳	۰/۲۴	کیفیت خدمات ۳
۰/۳۴	۰/۲۲	۰/۵۶	۰/۲۱	۰/۱۷	۰/۱۹	۰/۲۹	کیفیت خدمات ۴
۰/۳۴	۰/۲۷	۰/۵۸	۰/۲۷	۰/۲۷	۰/۱۶	۰/۳۳	کیفیت خدمات ۵
۰/۲۹	۰/۲۲	۰/۴۷	۰/۲۴	۰/۱۴	۰/۲۱	۰/۱۸	کیفیت خدمات ۶
۰/۳۲	۰/۲۸	۰/۵۷	۰/۰۹	۰/۲۶	۰/۲۱	۰/۱۹	کیفیت خدمات ۷
۰/۳۶	۰/۳۴	۰/۶۱	۰/۱۶	۰/۲۶	۰/۱۹	۰/۲۶	کیفیت خدمات ۸
۰/۳۴	۰/۲۵	۰/۵۱	۰/۲۴	۰/۲۴	۰/۱۸	۰/۲۶	کیفیت خدمات ۹
۰/۴۴	۰/۳۴	۰/۵۸	۰/۲۸	۰/۵۳	۰/۱۷	۰/۳۹	کیفیت خدمات ۱۰
۰/۸۹	۰/۵۷	۰/۵۶	۰/۴۱	۰/۶۲	۰/۷۴	۰/۶۲	اعتماد ۱
۰/۸۵	۰/۵۴	۰/۷۴	۰/۴۰	۰/۴۰	۰/۷۸	۰/۴۷	اعتماد ۲
۰/۶۹	۰/۵۴	۰/۳۶	۰/۶۵	۰/۴۴	۰/۲۵	۰/۳۸	اعتماد ۳
۰/۵۶	۰/۳۹	۰/۲۰	۰/۴۹	۰/۳۱	۰/۱۹	۰/۲۵	اعتماد ۴
۰/۴۷	۰/۴۶	۰/۳۷	۰/۳۸	۰/۳۱	۰/۲۵	۰/۷۲	کیفیت اطلاعات ۱
۰/۵۰	۰/۴۵	۰/۴۹	۰/۳۲	۰/۳۶	۰/۲۴	۰/۸۴	کیفیت اطلاعات ۲
۰/۵۲	۰/۴۴	۰/۴۸	۰/۳۰	۰/۳۸	۰/۲۷	۰/۸۵	کیفیت اطلاعات ۳
۰/۴۸	۰/۳۸	۰/۴۹	۰/۲۵	۰/۳۳	۰/۲۴	۰/۸۶	کیفیت اطلاعات ۴
۰/۴۹	۰/۴۲	۰/۴۴	۰/۳۱	۰/۳۷	۰/۲۳	۰/۸۵	کیفیت اطلاعات ۵
۰/۴۷	۰/۳۶	۰/۵۰	۰/۲۸	۰/۷۹	۰/۲۰	۰/۳۹	شهرت ۱
۰/۴۹	۰/۳۸	۰/۳۴	۰/۴۰	۰/۷۸	۰/۱۲	۰/۳۲	شهرت ۲
۰/۳۶	۰/۲۹	۰/۲۷	۰/۱۴	۰/۶۷	۰/۱۹	۰/۲۳	شهرت ۳
۰/۵۸	۰/۷۶	۰/۴۶	۰/۵۹	۰/۴۸	۰/۲۵	۰/۴۴	کیفیت سیستم ۱
۰/۴۶	۰/۷۷	۰/۴۷	۰/۳۶	۰/۳۱	۰/۳۱	۰/۴۰	کیفیت سیستم ۲
۰/۳۷	۰/۶۵	۰/۳۷	۰/۲۷	۰/۳۰	۰/۱۹	۰/۳۵	کیفیت سیستم ۳
۰/۴۳	۰/۶۰	۰/۲۳	۰/۲۶	۰/۱۸	۰/۵۰	۰/۲۴	کیفیت سیستم ۴

جدول ۳. بارهای عاملی

۷	۶	۵	۴	۳	۲	۱	
						۰/۸۲	کیفیت اطلاعات
					۰/۹۲	۰/۳۰	خودکارآمدی
				۰/۷۵	۰/۲۲	۰/۴۳	شهرت
			۰/۷۱	۰/۳۸	۰/۳۲	۰/۳۷	امنیت ساختاری
		۰/۵۶	۰/۳۹	۰/۵۰	۰/۳۸	۰/۵۵	کیفیت خدمات
	۰/۷۰	۰/۴۲	۰/۵۵	۰/۴۷	۰/۴۴	۰/۵۲	کیفیت سیستم
۰/۷۶	۰/۶۷	۰/۵۴	۰/۶۰	۰/۵۹	۰/۷۲	۰/۶۰	اعتماد

جدول ۴. همبستگی های میان متغیرهای

مکنون و مقادیر AVE

تأیید روایی و پایایی مدل‌های اندازه‌گیری امکان ارزیابی مدل ساختاری را میسر می‌سازد. در این مرحله از معیار ضریب تعیین برای ارزیابی مدل ساختاری پژوهش استفاده شده است. چین^{۳۸} (۱۹۹۸) مقادیر ۰/۱۹، ۰/۳۳ و ۰/۶۷ را به عنوان معیاری جهت ارزیابی ضعیف، متوسط و قوی بودن ضرایب تعیین معرفی کرده است. مطابق شکل ۲، اعتماد اولیه به عنوان تنها متغیر درونزای مدل از ضریب تعیین قوی ۰/۸۵ برخوردار بوده که نشان‌دهنده‌ی ارزیابی بسیار مناسب بخش ساختاری مدل است. پس از تأیید بخش‌های اندازه‌گیری و ساختاری، از شاخص نیکویی برازش (GOF) که به طور همزمان مدل‌های اندازه‌گیری و ساختاری را مد نظر قرار می‌دهد جهت ارزیابی برازش مدل کلی پژوهش استفاده شد [۴۵]. وتزلس و همکاران^{۳۹} سه مقدار ۰/۰۱، ۰/۲۵ و ۰/۳۶ را به عنوان مقادیر ضعیف، متوسط و قوی برای GOF معرفی نموده اند [۴۶]. از آنجایی برای مدل حاضر GOF برابر ۰/۶۹۲ محاسبه گردید، برازش مناسب مدل کلی پژوهش مورد تأیید قرار گرفت. پس از بررسی و تأیید مدل‌های اندازه‌گیری، ساختاری و کلی پژوهش،

جهت سنجش روایی مدل‌های اندازه‌گیری، از روایی همگرا^{۳۳} و روایی واگرا^{۳۴} استفاده شد. برای ارزیابی روایی همگرا، معیار متوسط واریانس استخراج شده^{۳۵} (AVE) مورد بررسی قرار گرفت [۴۲]. از آنجایی که مقادیر AVE بالای ۰/۴ نشان‌دهنده‌ی روایی همگرای قابل قبول است [۴۵]، بر اساس مقادیر AVE گزارش شده در جدول ۲ تمامی متغیرهای مدل‌های اندازه‌گیری دارای حد مطلوبی از روایی همگرا می‌باشند. در ادامه نیز جهت بررسی روایی واگرا از دو روش آزمون بارهای عرضی^{۳۶} (بررسی روایی واگرا در سطح معرف) و معیار فورنل و لاکر^{۳۷} (بررسی روایی واگرا در سطح سازه) استفاده گردیده است. در آزمون بار عرضی انتظار می‌رود بار هر معرف برای هر متغیر مکنون (اعداد پرنرگ جدول ۳) بیشتر از بارهای عرضی یا بار آن معرف برای سایر متغیرهای مکنون باشد [۳۹].

که این مهم مطابق جدول ۳ برای متغیرهای این پژوهش مورد تأیید قرار می‌گیرد و می‌توان دارا بودن روایی واگرای مدل‌های اندازه‌گیری در سطح معرف را نتیجه گرفت. معیار فورنل-لاکر بیان می‌کند که یک متغیر باید در مقایسه با معرف‌های سایر متغیرهای مکنون، پراکندگی بیشتری را در بین معرف‌های خودش داشته باشد. بنابراین از نظر آماری، AVE هر متغیر مکنون باید بیشتر از بالاترین توان دوم همبستگی آن متغیر با سایر متغیرهای مکنون باشد [۵۰]. برای کاهش محاسبات می‌توان از روش معادل یعنی مقایسه جذر AVE با همبستگی‌ها بهره برد. طبق جدول ۴ بر اساس نتایج محاسبه‌ی همبستگی‌ها و جذر AVE (اعداد قطر جدول) می‌توان روایی واگرای مدل‌های اندازه‌گیری در سطح سازه را نتیجه گرفت.

^{۳۳}. Convergent Validity

^{۳۴}. Discriminant Validity

^{۳۵}. Average Variance Extracted (AVE)

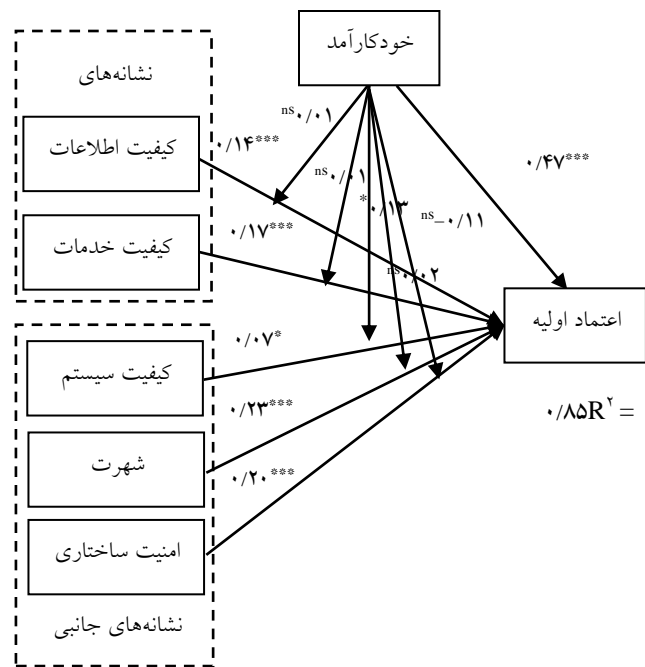
^{۳۶}. Cross-Loadings

^{۳۷}. Fornell & Larcker

^{۳۸}. Chin

^{۳۹}. Wetzels and et al

بررسی و آزمون فرضیه‌های تحقیق در سه مرحله جداگانه انجام شد. در مرحله اول مدل بدون حضور متغیر خودکارآمدی مورد سنجش قرار گرفت. در مرحله دوم این متغیر به عنوان متغیر مستقل در نظر گرفته شد و تنها اثر مستقیم آن بر روی اعتماد اولیه لحاظ گردید. در مرحله سوم این متغیر هم به عنوان متغیر مستقل و هم به عنوان تعدیل‌گر پنج رابطه‌ی دیگر موجود در مدل در نظر گرفته شد. در هر مرحله ضرایب مسیر به طور جداگانه محاسبه گردید. همچنین معناداری ضرایب مسیر هر مرحله از روش بوت استراپ با ۷۵۰ مرتبه تکرار محاسبه شد.



(ns = no significant, * $p < .05$, ** $p < .01$, *** $p < .001$)

شکل ۲. نتایج پایانی مدل پیشنهادی پژوهش

با توجه به نتایج حاصل از اندازه‌گیری مدل در مراحل مختلف، فرضیه‌های پژوهش مورد آزمون قرار گرفت که نتایج آن به تفکیک مراحل در جدول‌های ۵ و ۶ ارائه شده است.

نتیجه	معناداری	ضریب مسیر	فرضیات
***	۳/۸۱	۰/۱۶	H ₁ کیفیت اطلاعات تأثیر مثبتی بر اعتماد اولیه دارد.
***	۴/۹۴	۰/۲۱	H ₂ کیفیت خدمات تأثیر مثبتی بر اعتماد اولیه دارد.
***	۴/۷۶	۰/۲۲	H ₃ کیفیت سیستم تأثیر مثبتی بر اعتماد اولیه دارد.
***	۶/۰۱	۰/۲۱	H ₄ شهرت تأثیر مثبتی بر اعتماد اولیه دارد.
***	۶/۲۴	۰/۲۹	H ₅ امنیت ساختاری تأثیر مثبتی بر اعتماد اولیه دارد.

(ns = no significant, * $p < .05$, ** $p < .01$, *** $p < .001$)

جدول ۵. مرحله اول: آزمون فرضیات اصلی بدون حضور متغیر تعدیل‌گر

فرضیات	خودکارآمدی به عنوان متغیر مستقل		خودکارآمدی به عنوان متغیر مستقل و تعدیل‌گر	
	β	SE	β	SE
H ₁ کیفیت اطلاعات تأثیر مثبتی بر اعتماد اولیه دارد.	0.15	0.05	0.14	0.02
H ₂ کیفیت خدمات تأثیر مثبتی بر اعتماد اولیه دارد.	0.17	0.05	0.17	0.06
H ₃ کیفیت سیستم تأثیر مثبتی بر اعتماد اولیه دارد.	0.07	0.02	0.07	0.02
H ₄ شهرت تأثیر مثبتی بر اعتماد اولیه دارد.	0.23	0.09	0.23	0.09
H ₅ امنیت ساختاری تأثیر مثبتی بر اعتماد اولیه دارد.	0.20	0.05	0.20	0.13
H ₆ خودکارآمدی تأثیر مثبتی بر اعتماد اولیه دارد.	0.47	0.05	0.47	0.05
H ₇ خودکارآمدی تأثیر مثبتی بر اعتماد اولیه را تعدیل می‌کند.			0.01	0.04
H ₈ خودکارآمدی تأثیر مثبتی بر اعتماد اولیه را تعدیل می‌کند.			0.01	0.13
H ₉ خودکارآمدی تأثیر مثبتی بر اعتماد اولیه را تعدیل می‌کند.			0.13	0.28
H ₁₀ خودکارآمدی تأثیر مثبتی بر اعتماد اولیه را تعدیل می‌کند.			0.02	0.07
H ₁₁ خودکارآمدی تأثیر مثبتی بر اعتماد اولیه را تعدیل می‌کند.			-0.11	0.24

(ns = no significant, * $p < .05$, ** $p < .01$, *** $p < .001$)

جدول ۶. بررسی فرضیات بر اساس عملکرد متغیر خود کار آمدی

طراحان شبکه‌های اجتماعی باید به این عامل توجه ویژه‌ای داشته باشند به گونه‌ای که یادگیری آن آسان و فرآیند استفاده از آن برای کاربران واضح و قابل فهم باشد. به منظور تقویت اعتماد اولیه ارائه‌دهندگان خدمات باید به ارتقاء کیفیت سیستم پردازند؛ چرا که اعتماد جزء دارایی‌های بلند مدتی است که می‌توان با انتظار مزایا در آینده، در آن سرمایه‌گذاری کرد و برای توسعه‌ی روابط بلندمدت میان کاربران و ارائه‌دهندگان خدمات اهمیت دارد. به ارائه‌دهندگان خدمات پیشنهاد می‌شود با فراهم نمودن اطلاعات بهنگام، دقیق، صحیح، کامل، ایجاد محیطی امن و مطمئن و سهولت استفاده از وبسایت زمینه افزایش اعتماد کاربران را ایجاد نمایند. باتاچرج و سانفورد بیان کردند کاربرانی که ارتباط شغلی و تخصص نسبتاً بالا در استفاده از فناوری اطلاعات دارند تمایل دارند که از مسیر مرکزی استفاده کنند؛ در حالی که کسانی که ارتباط شغلی و تخصص کمی در فناوری اطلاعات دارند بر نشانه‌های جانبی برای تصمیم‌گیری در پذیرش فناوری اطلاعات تکیه می‌کنند [۱۸]. نتیجه پژوهش ژو^{۴۰} نیز نشان داد که هنگامی که کاربران خودکارآمدی پایینی دارند، بر نشانه‌های اطلاعاتی مانند کیفیت سیستم برای اعتماد اولیه از طریق مسیر محیطی تکیه می‌کنند. همچنین او بیان کرد که اثر شهرت بر اعتماد اولیه توسط خودکارآمدی تعدیل نمی‌شود [۸]. علاوه بر پنج عامل تأثیرگذار بر اعتماد اولیه در مدل پژوهش، عوامل دیگری مانند سهولت استفاده و سودمندی درک شده را نیز می‌توان مورد بررسی قرار داد. در این پژوهش تنها از خودکارآمدی به عنوان متغیر تعدیل‌گر استفاده شد؛ می‌توان متغیرهای دیگری مانند درگیری ذهنی را نیز به عنوان متغیر تعدیل‌کننده مورد توجه قرار داد. به دلیل شرایط و ویژگی‌های شبکه‌های اجتماعی، یکی از محدودیت‌های این پژوهش استفاده از روش نمونه‌گیری در

با توجه به جدول ۶ فرضیه‌های اول، دوم، سوم، چهارم، پنجم، ششم و نهم مورد تأیید قرار می‌گیرد و فرضیه‌های هفتم، هشتم، دهم و یازدهم رد می‌شود.

۵- نتیجه‌گیری

فیس‌بوک یک وبسایت اجتماعی و محصول دنیای ارتباطات و فناوری اطلاعات است. هدف این پژوهش درک اعتماد اولیه کاربران فیس‌بوک و پر کردن شکاف فرآیند توسعه اعتماد می‌باشد. بدین منظور از مدل بسط احتمالی استفاده شد. این نظریه بیان می‌کند که اعتماد اولیه از طریق یک مسیر دوگانه یعنی مسیر مرکزی و مسیر جانبی توسعه می‌یابد. کیفیت اطلاعات و کیفیت خدمات به عنوان نشانه‌های مرکزی و کیفیت سیستم، شهرت و امنیت ساختاری به عنوان نشانه‌های جانبی عمل می‌کنند. خودکارآمدی نیز به عنوان تعدیل‌کننده‌ی احتمالی توسعه‌ی اعتماد به کار گرفته شد.

همان‌گونه که در جدول ۶ نشان داده شد؛ از میان یازده فرضیه این پژوهش فرضیه‌های اول، دوم، سوم، چهارم، پنجم، ششم و نهم مورد تأیید قرار گرفت و فرضیه‌های هفتم، هشتم، دهم و یازدهم رد شد. بدین ترتیب نتایج نشان داد که هم نشانه‌های مرکزی (کیفیت اطلاعات، کیفیت خدمات) و هم نشانه‌های جانبی (کیفیت سیستم، شهرت، امنیت ساختاری) و همچنین خودکارآمدی بر اعتماد اولیه اثرگذار هستند. این نتایج با نتایج سایر پژوهش‌های مربوط به شناسایی عوامل تأثیرگذار بر اعتماد همسو می‌باشد.

از میان متغیرهای مستقل این پژوهش خودکارآمدی، شهرت و امنیت ساختاری به ترتیب با ضریب مسیر ۰/۴۷، ۰/۲۳ و ۰/۲۰ بیشترین اثر را بر اعتماد اولیه دارند و کیفیت سیستم کمترین اثر را بر اعتماد اولیه با مقدار ۰/۰۷ دارد. همچنین نتایج بیانگر این مطلب بوده که خودکارآمدی تنها اثر کیفیت سیستم را بر اعتماد اولیه تعدیل می‌کند. علاوه بر این نتایج نشان داد که خودکارآمدی اثر کیفیت اطلاعات، کیفیت خدمات، امنیت ساختاری و شهرت را بر اعتماد اولیه را تعدیل نمی‌کند و این بدان معنی است که اثر این متغیرها بر اعتماد اولیه تحت تأثیر خودکارآمدی قرار نمی‌گیرد. لذا از آن جایی که خودکارآمدی جدای از نقش تعدیل‌گریش به طور مستقیم نیز بیشترین تأثیر را بر اعتماد اولیه دارد؛

^{۴۰}: Zhou

۸. Zhou, T., Understanding users' initial trust in mobile banking: An elaboration likelihood perspective, *Computers in Human Behavior*, ۲۰۱۲, ۲۸: ۱۵۱۸-۱۵۲۵.
۹. McKnight, D.H. Chervany, N.L., Reflections on an initial trust-building model. *Handbook of trust research*, ۲۰۰۶, ۲۹-۵۱.
۱۰. Mayer, R.C. Davis, J.H. Schoorman, F.D., An integrative model of organizational trust. *Academy of Management Review*, ۱۹۹۵, ۲۰ (۳): ۷۰۹-۷۳۴.
۱۱. Szymczak, H., Küçükbalaban, P., Lemanski, S., Knuth, D., & Schmidt, S., Trusting Facebook in crisis situations: the role of general use and general trust toward Facebook. *Cyberpsychology, Behavior, and Social Networking*, ۲۰۱۶, ۱۹(۱), ۲۳-۲۷.
۱۲. Kim, Y. H. Kim, D. J., A Study of Online Transaction Self-Efficacy, Consumer Trust, and Uncertainty Reduction in Electronic Commerce Transaction. ۳۸th Hawaii International Conference on System Sciences, ۲۰۰۵.
۱۳. Coulter, K.S. Coulter, R.A., The effects of industry knowledge on the development of trust in service relationships. *Int J Res Mark*, ۲۰۰۳, ۲۰: ۳۱-۴۳.
۱۴. Rodolfo, V.C. Leticia, S.A. Ana, M.D.M., Trust as a key factor in successful relationships between consumers and retail service providers. *The Service Industries Journal*, ۲۰۰۵, ۲۵(۱): ۸۳-۱۰۱.
۱۵. Chang, Ch. Sh. Chen, Su. Y. Lan, Yi. T., Service quality, trust, and patient satisfaction in interpersonal-based medical service encounters. *BMC Health Services Research*, ۲۰۱۳, ۱۳(۲۲): ۱۳-۲۲.

دسترس است. همچنین از محدودیت‌های دیگر این پژوهش آن است که داده‌ها از نوع مقطعی می‌باشند. به دلیل وجود محدودیت‌هایی برای دسترسی به شبکه‌ی اجتماعی فیس‌بوک در ایران، پیشنهاد می‌شود که مدل حاضر در یکی از شبکه‌های اجتماعی که چنین محدودیت‌هایی بر آن اعمال نمی‌شود، بار دیگر مورد بررسی قرار گیرد.

منابع

۱. Bargh, J. A. & McKenna, K. Y. A., The Internet and social life. *Annual Review of Psychology*, ۲۰۰۴, ۵۵: ۵۷۳-۵۹۰.
۲. Hanafizadeh, P. Behboudi, M. Abedini Koshksaray, A., Mobile-banking adoption by Iranian bank clients. *Telematics and Informatics*, ۲۰۱۴, ۳۱: ۶۲-۷۸.
۳. Dong, T. P. Cheng, N.C. Jim wu, Y.Ch., A study of the social networking website service in digital content industries: The Facebook case in Taiwan. *Computers in Human Behavior*, ۲۰۱۴, <http://dx.doi.org/۱۰.۱۰۱۶/j.chb.۲۰۱۳.۰۷.۰۳۷>.
۴. Chang, C. Heo, j., Visiting theories that predict college students' self-disclosure on Facebook. *Computers in Human Behavior*, ۲۰۱۴, ۳۰: ۷۹-۸۶.
۵. McKnight, D.H. Choudhury, N.L. Kacmar, C., The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *Journal of Strategic Information Systems*, ۲۰۰۲b, ۱۱: ۲۹۷-۳۲۳.
۶. Ba, S. Whinston, A. Zhang, H., Building trust in online auction markets through an economic incentive mechanism. *Decision Support Systems*, ۲۰۰۳, ۳۵: ۲۷۳- ۲۸۶.
۷. Koufaris, M. Hampton-Sosa, W., The development of initial trust in an online company by new customers. *Information & Management*, ۲۰۰۴, ۴۱: ۳۷۷-۳۹۷.

۲۴. Beldad, A. de Jong, M. Steehouder, M., How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior*, ۲۰۱۰, ۲۶(۵): ۸۵۷-۸۶۹.
۲۵. Parasuraman, A. Zeithaml, V. A. Berry, L. L., SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, ۱۹۸۸, ۶۴(۱): ۱۲-۳۷
۲۶. Prayag, G., Assessing international tourists' perceptions of service quality at Air Mauritius. *International Journal of Quality & Reliability Management*, ۲۰۰۷, ۲۴(۵): ۴۹۲-۵۱۴.
۲۷. Kim, H. W. Xu, Y. Koh, J., A comparison of online trust building factors between potential customers and repeat customers. *Journal of the Association for Information Systems*, ۲۰۰۴, ۵(۱۰): ۳۹۲-۴۲۰.
۲۸. Grabner-Kräuter, S., & Bitter, S., Trust in online social networks: A multifaceted perspective. In *Forum for social economics*, ۲۰۱۵, Vol. ۴۴, No. ۱, pp. ۴۸-۶۸.
۲۹. Bart, Y. Shankar, V. Sultan, F. Urban, G.L., Are the drivers and role of online trust the same for all Web sites and consumers? A large-scale exploratory empirical study. *Journal of Marketing*, ۲۰۰۵, ۶۹: ۱۳۳-۱۵۲.
۳۰. Vance, A. Christophe, E.-D.-C. Straub, D. W., Examining trust in information technology artifacts: The effects of system quality and culture. *Journal of Management Information Systems*, ۲۰۰۸, ۲۴(۴): ۷۳-۱۰۰.
۳۱. Teo, S.H. Liu, J., Consumer trust in e-commerce in the United States, Singapore and China. *The international journal of management science*, ۲۰۰۷, ۲۲
۱۶. Ayyash, M. M. Ahmad, K. Singh, D., Investigating the Effect of Information Systems Factors on Trust in E-Government Initiative Adoption in Palestinian Public Sector. *Research Journal of Applied Sciences, Engineering and Technology*, ۲۰۱۳, ۵(۱۵): ۳۸۶۵-۳۸۷۵.
۱۷. Petty, R. Cacioppo, J., Issue involvement as a moderator of the effects on attitude of advertising content and context. *Advances in Consumer Research*, ۱۹۸۱, ۸: ۲۰-۲۴.
۱۸. Bhattacharjee, A. Sanford, C., Influence processes for information technology acceptance. An elaboration likelihood model. *MIS Quarterly*, ۲۰۰۶, ۳۰(۴): ۸۰۵-۸۲۵.
۱۹. Yi, M.Y. Yoon, J.J. Davis, J.M. Lee, T., Untangling the antecedents of initial trust in Web-based health information: The roles of argument quality, source expertise, and user perceptions of information quality and risk. *Decision Support Systems*, ۲۰۱۳, ۵۵: ۲۸۴-۲۹۵.
۲۰. Seddon, P.B., A respecification and extensions of the Delone and Mclean model of IS success. *Information Systems Research*, ۱۹۹۷, ۸(۳): ۲۴۰-۲۵۳.
۲۱. Nicolaou, A. I. McKnight, D.H., Perceived Information Quality in Data Exchanges: Effects on Risk, Trust, and Intention to Use. *Information Systems Research*, ۲۰۰۶, ۱۷(۴): ۳۳۲-۳۵۱.
۲۲. McKinney, V. Yoon, K. Zahedi, F.M., The measurement of Web customer satisfaction: An expectation and disconfirmation approach. *Information System Research*, ۲۰۰۲, ۱۳(۳): ۲۹۶-۳۱۵.
۲۳. Ha, H.Y., Factors influencing consumer perceptions of brand trust online. *Journal of Product & Brand Management*, ۲۰۰۴, ۱۳(۵): ۳۲۹-۳۴۲.

۴۱. Nunnally, J. C., & Bernstein, I. H., Psychometric theory (۳rd ed). New York: McGraw-Hill, ۱۹۹۴.
۴۲. Nunnally, J. C., Psychometric theory. New York, NY: McGraw-Hill, ۱۹۷۸.
۴۳. Ha, H. Y., John, J., John, J. D., & Chung, Y. K., Temporal effects of information from social networks on online behavior: the role of cognitive and affective trust. *Internet Research*, ۲۰۱۶, ۲۶(۱), ۲۱۳-۲۳۵.
۴۴. Magner, N. Welker, R. B. & Campbell, T. L., Testing a model of cognitive budgetary participation -processes in a latent variable structural equations framework. *Accounting and Business Research*, ۱۹۹۶, ۲۷ (۱): ۴۱-۵۰.
۴۵. داوری، علی؛ رضازاده، آرش؛ مدل‌سازی معادلات ساختاری با نرم افزار PLS، تهران: انتشارات سازمان جهاد دانشگاهی، ۱۳۹۲.
۴۶. Wetzels, M., Odekerken-Schroder, G., & Van Oppen, C., Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, ۲۰۰۹, ۳۳(۱): ۱۷۷.
۴۷. Amiri Aghdaie, F. Fathi, S. Piraman, A., Factors affecting the attitude of trust in Internet purchasing from the perspective of consumers. *Institute of Interdisciplinary Business Research*, ۲۰۱۱, ۳ (۵): ۲۰۸-۲۲۱.
۴۸. Xiaoni, Z. Kellie, B.K. Robert, J.P., Information Quality of Commercial Web Site Home Pages: An Explorative Analysis. Retrieved July ۱, from: <http://aisel.aisnet.org/icis۲۰۰۰/۱۶>.
۴۹. Casalo, L. V. Flavian, C. Guinaliu, M., The influence of satisfaction, perceived reputation and trust on a consumer's commitment to a website. *Journal of*
۳۲. Doney, PM. Cannon, JP., an examination of the nature of trust in buyer-seller relationships. *Journal of Marketing*, ۱۹۹۷, ۶۱(۲): ۳۵-۵۱.
۳۳. Shi, S., & Chow, W. S., Trust development and transfer in social commerce: prior experience as moderator. *Industrial Management & Data Systems*, ۲۰۱۵, ۱۱۵(۷), ۱۱۸۲-۱۲۰۳.
۳۴. Shapiro, S.P., The social control of impersonal trust. *American Journal of Sociology*, ۱۹۸۷, ۹۳ (۳): ۶۲۳-۶۵۸.
۳۵. Head, M. Hassanein, K., Trust in e-Commerce: Evaluating the Impact of Third-Party Seals. *Quarterly Journal of Electronic Commerce*, ۲۰۰۲, ۳(۳): ۳۰۷-۳۲۵.
۳۶. Bandura, A., Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, ۱۹۷۷, ۸۴: ۱۹۱-۲۱۵.
۳۷. McKnight, D. H. Choudhury, V. Kacmar, C., Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, ۲۰۰۲a, ۱۳(۳): ۳۳۴-۳۵۹.
۳۸. Zhou, Zh., Zhang Q., Su Ch., Zhou N., How do brand communities generate brand relationships? Intermediate mechanisms, *Journal of Business Research*, ۲۰۱۲, ۶۵: ۸۹۰-۸۹۵.
۳۹. آذر، عادل؛ غلامزاده، رسول؛ قنواتی، مهدی، مدلسازی مسیری-ساختاری در مدیریت، کاربرد نرم‌افزار Smart PLS تهران، انتشارات نگاه دانش، ۱۳۹۱.
۴۰. Cronbach, L.J., Coefficient Alpha and the Internal Structure of Tests. *Psychometrical*, ۱۹۵۱, ۱۶: ۲۹۷- ۳۳۴.

equation models with unobservable variables and measurement error. *Journal of Marketing Research*, ۱۹۸۱, ۱۸(۱): ۳۹-۵۰.

Marketing Communications, ۲۰۰۷, ۱۳(۱): ۱-۱۷.

۵۰. Fornell, C., & Larcker, D., Structural

راهنمای نگارش

مقالات فصلنامه علمی - پژوهشی فناوری اطلاعات و ارتباطات ایران

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

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

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

و - مراجع: شامل همه کتابها، مقاله‌ها، گزارشها و صفحه‌های وب مورد استناد در مقاله.

یادآوری ۱: 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).

یادآوری ۲: شیوه نگارش مراجع و ارجاع به آن در متن، مطابق استاندارد نشریات IEEE است.

یادآوری ۳: در مقاله‌های فارسی لازم است همه 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، مصوبات فرهنگستان زبان و ادب فارسی است.

یادآوری ۴: شیوه نگارش فارسی فصلنامه براساس «دستور خط فارسی» مصوب 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 همه مقاله‌ها (خواه فارسی، خواه انگلیسی) باید دارای چکیده دو زبانه باشند.

۶- لازم است نگارندگان مقاله خود را در محیط MS-Word به صورت دو ستونه، با فاصله خطوط ۱ (Single) و با قلم بی-لوتوس (B Lotus) ۱۱ (برای مقاله فارسی) و Times New Roman ۱۱ (برای مقاله انگلیسی) و با رعایت حاشیه 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

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

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

۱- مقاله‌های ارسالی به فصلنامه باید دربرگیرنده نتایج نو در هر یک از حوزه‌های تخصصی فصلنامه باشد و نباید قبلاً در هیچ نشریه دیگری به چاپ رسیده یا به طور همزمان به سایر نشریه‌های داخلی و خارجی برای داوری ارسال شده باشد. شایان ذکر است انتشار مقاله کاملی که خلاصه 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 یافته مقاله (طول عنوان نباید از ۱۵ واژه بیشتر باشد).

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

۸- هر مقاله پس از بررسی سردبیر به سه داور ارسال می‌شود. جمع‌بندی نظریات داوران برعهده سردبیر فصلنامه است. فصلنامه حق پذیرش، ویرایش یا رد 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 نویسنده عهده‌دار مکاتبات ارسال کند. ضمناً دسترسی به نسخه الکترونیکی مقاله برای همه نویسندگان مقاله امکان‌پذیر خواهد بود.

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

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

۷- حق نشر مقاله‌ها برای فصلنامه محفوظ است؛ از این رو ضروری است نویسنده عهده‌دار مکاتبات به نمایندگی از همه نویسندگان «مجوز حق انتقال نشر» را امضا کند، در غیر این صورت فصلنامه از چاپ مقاله پذیرفته شده معذور است.

معرفی انجمن فناوری اطلاعات و ارتباطات ایران

اهداف

- انجمن فناوری اطلاعات و ارتباطات ایران در سال ۱۳۸۳ با اهداف زیر تشکیل گردید:
- ایجاد ارتباط با پژوهشگران و افراد خیره در حوزه‌های مرتبط با فناوری اطلاعات (IT)، فناوری ارتباطات (CT) و فراهم نمودن زمینه‌های شناسایی و همکاری مشترک بین آنها.
 - همکاری با نهادهای اجرائی، علمی و پژوهشی مرتبط با فناوری اطلاعات و ارتباطات، اقدام در جهت پیوند و همکاری دستگاههای اجرائی با مجموعه‌های علمی - تحقیقاتی.
 - توسعه فرهنگ استفاده مطلوب از توانمندیهای فناوری اطلاعات و ارتباطات.
 - ارزیابی و تحلیل عوامل رشد کشورهای مشابه دیگر و استفاده مطلوب از تجربیات دیگران در توسعه کشور در حوزه ارتباطات و فناوری اطلاعات.

برنامه‌ها

- انجمن به منظور تحقق اهداف خود، برنامه‌های زیر را در اولویت فعالیت‌هایش قرار داده است.
- ایجاد و تقویت ارتباط علمی، پژوهشی و آموزشی در سطوح ملی، منطقه‌ای و بین‌المللی.
 - برگزاری همایشهای علمی در سطح ملی، منطقه‌ای و بین‌المللی.
 - ایجاد تسهیلات برای نشر نتایج آخرین دستاوردهای علمی و پژوهشی مرتبط با نیازهای کشور در حوزه ارتباطات و فناوری اطلاعات (انتشار کتب و نشریات علمی).
 - برگزاری کارگاههای تخصصی در حوزه ارتباطات و فناوری اطلاعات.

برنامه‌ریزی فعالیتها

- انتشار خبرنامه انجمن
- تشکیل کمیته‌های تخصصی به منظور 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 فناوری اطلاعات و دانش (IKT)
- تشکیل جلسات با مسئولین مملکتی ذیربط جهت شناساندن اهداف انجمن و مشخص شدن 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 حوزه ارتباطات و فناوری اطلاعات دعوت می‌شود با عضویت در انجمن، ما را در تحقق اهداف والای آن یاری نمایند.

نشانی دبیرخانه: تهران - خیابان حافظ - دانشگاه صنعتی امیرکبیر - ساختمان ابوریحان - طبقه ششم - اتاق ۶۱۲

تلفن: ۰۲۱-۶۶۴۸۵۸۵۶، تلفکس: ۰۲۱-۶۶۴۹۵۴۳۳

Understanding the users' initial trust of social networks

* Mohsen Akbari¹, ** Kamran Zahedfar, ***Zahra Ayagh

*Assistant Professor, Gilan University

*Phd Student, Business Management, Tehran University

*** Master of Science, Business Management, Gilan University

Abstract

Social networks are one of the types of social media which in recent years has been influential in the lives of internet users so that forms of social communication is influenced by the new media and in the future will play a more important role. On the other hand the trust is one of the major causes of satisfaction and repeat use of the online service. This research employs elaboration likelihood model to identify the factors affecting the users' initial trust in Facebook and it express that users change their attitude via a dual route including central route and peripheral route. Central cues include information quality and service quality, whereas peripheral cues include system quality, structural assurance and reputation. Self-efficacy moderates the effects of central cues and peripheral cues on initial trust. The populations of this study are Iranian users on Facebook. Data for this study through contacts with personal pages and pages of general (relating to groups, cities, activities, university, etc.) active in the social network were collected. For this purpose, link of online questionnaire to provide them on the website docs.google.com. In this questionnaire was used a five -point Likert scale from strongly disagree (1) to strongly agree (5) and finally 406 questionnaires were collected according to schedule and Morgan Jersey, which are well to conduct research in the infinite. Questionnaire validity by the opinions of experts and university professors and its reliability through Cronbach's alpha were provided and represents optimum reliability. Data were analyzed with Smart PLS statistical software. According to the results of the independent variables in this study, self-efficacy, reputation and structural assurance respectively with path coefficient 0.47, 0.23 and 0.2 have the greatest effect on the initial trust and system quality has minimal impact on initial trust with amount 0.07. The results also indicate that self-efficacy moderates the effect of system quality on initial trust. Therefore, since the self-efficacy apart from the moderating role directly also has the greatest impact on initial trust; social network designers should be a special attention to this factor so that it's learn is easy and it's use process is clear and understandable for users.

Keywords: Initial trust, Self-efficacy, Facebook, Elaboration likelihood model.

1. Corresponding author (Akbarimohsen@gmail.com)

Integrating Data Envelopment Analysis and Decision Tree Models in Order to Evaluate Information Technology-Based Units

* Amir Amini¹, ** Alireza Alinezhad, ***Somayeh Shafaghizadeh

*MSc Student of Industrial Engineering, Department of Industrial Engineering, Urmia University of Technology, Urmia, Iran

**Associate Professor, Faculty of Industrial and Mechanical Engineering, Qazvin Branch, Islamic Azad University, Qazvin

***MSc graduate of industrial engineering, Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Qazvin

Abstract:

In order to evaluate the performance and desirability of the activities of its units each organization needs an evaluation system to assess this desirability and it is more important for financial institutions, including information technology-based companies. Data envelopment analysis (DEA) is a non-parametric method to measure the effectiveness and efficiency of decision-making units (DMUs). On the other hand, data mining technique allows DMUs to explore and discover meaningful information, which had previously been hidden in large databases. . This paper presents a general framework for combining DEA and regression tree for evaluating the effectiveness and efficiency of the DMUs. Resulting hybrid model is a set of rules that can be used by policy makers to discover reasons behind efficient and inefficient DMUs. Using the proposed method for examining factors related to productivity, a sample of 18 branches of Iran insurance in Tehran was elected as a case study. After modeling based on advanced model the input oriented LVM model with weak disposability in data envelopment analysis was calculated using undesirable output, and by use of decision tree technique deals with extracting and discovering the rules for the cause of increased productivity and reduced productivity.

Key words: Data envelopment analysis, Data mining, Classification and regression, Decision tree, Undesirable output

1. Corresponding author (A.Amini@ine.uut.ac.ir)

A Novel Model for detecting intrusion with Mobile Agent and Game theory

* Amin Nezarat ¹, ** Mehdi Raja, *** G.H Dastghaibyard

* Assistant Professor, Department of Computer Engineering, Payam e Noor University, Iran

** Department of Computer Engineering, Shiraz University, Iran

*** Associate Professor, Department of Computer Engineering, Shiraz University, Iran

Abstract:

Network intrusion detection systems are tools used to protect network resources against attacks. Due to the widespread nature of attacks on the Internet and the deformation and type of distributed attacks, the architecture of such systems is also distributed. In this paper, we propose a Mobile Agent based methodology that acts as a sensor for detecting non-valid movements. The effective factors in identifying widespread attacks on the network are moving from one node to another, and each time they have an overlay security network using a Cooperative game and communicate with one another, after reaching Shapley Value, they can identify and report an attack. In this paper, we suggest that the WGA attempts, in a non-cooperative game with an attacker, to build a legendary relationship to calculate its Nash equilibrium and maximize profits to distinguish between actual requests and attacks.

Key words: Mobile Agents, IDS, Game Theory, Multi Agent Systems, Nash Equilibrium, Shapley Value, Network Security

1. Corresponding author (amin_nezarat@hotmail.com)

Extracting the Information Technology architecture of hospitals with approach of Implementing in Iran

* Atefeh.S Haghithoseini, ** Hosein Boubarshad, *** Fatemeh Saghafi, ** Hadi Zare

*Master of Science of Faculty of New Science and Technology of University of Tehran

** Assistant Professor of faculty of new science and technology of university of Tehran

***Assistant Professor of faculty of Mangement of university of Tehran (Corresponding author)

Abstract:

Nowadays, smart and fast services to patients and move on to the next generation hospital are the essential parts of the health field. The production of information technology architecture for hospital organizations is the foundation of accessible smart services and providing services with more speed and higher quality than the traditional systems. In this paper, the target is to present an indigenous IT architecture based on important criteria and metrics in Iranian hospitals. The TOGAF architecture is used and then it is adopted and localized for Shariati hospital based on Iran's indigenous conditions. For this aim, expert's views and 134 questionnaire tools are applied and the results are analyzed through suitable statistical tests. So IT architecture is designed by consisting up a conceptual model of four input and four infrastructure layer. The results show that 111 items are verified to apply to the hospital among 145 possible items. The customization framework is called the Hospital IT Architecture. The components provide the framework approved in 8 layers and 11 components can be used in hospitals to implement enterprise architecture. The proposed architecture in 8 layers and 11 components are designed and it could be applied as an indigenous reference architecture for implementation of enterprise architecture in Iranian Hospital organizations.

Key words: Next generations hospitals; information technology architecture; TOGAF architecture

A proper method for the advertising email classification based on user's profiles

*Rahim Hazratgholizadeh, ** Mohammad Fathian

* Master of Science, Industrial Engineering, Iran University of Science and Technolog

**Full Professor, Iran University of Science and Technolog

Abstract:

By the increasing of video communication in portable and functional devices, encoders design with low complexity and high performance are required. H.264 / AVC standard offers higher compression efficiency than previous standards. But this standard by employing several powerful coding techniques, considerably increased complexity at the encoder. This paper presents a new algorithm to reduce the complexity of the H.264/AVC encoder.

The proposed method uses simple directional masks, neighboring blocks modes and detection of 4×4 and/or 16×16 intra estimation modes with determination of quantization parameters for fast mode selection in Intra-Frame Modes prediction. Experimental results show that the proposed method reduces maximum 29% of the encoding time, while has little effect on visual quality and PSNR.

Key words: Electronic Commerce, Electronic Advertising, spam classification, Data mining, profile

New changes of local binary patterns and classification and segmentation of seabed images

* BabakGoudarzi ¹, ** Reza.Javidan, ***M.Javad Dehghani

* Master of Science, Department of Computer Engineering and Information Technology, Shiraz University of Technology

** Associate Professor, Department of Computer Engineering and Information Technology, Shiraz University of Technology

***Associate Professor, Department of Computer Engineering and Information Technology, Shiraz University of Technology

Abstract:

Texture analysis plays an important role in image processing. According to the super-textured appearance of sonar images, texture analysis techniques are a good choice for analyzing the acoustic images of the seabed. Local binary pattern operator is a very effective and multi-resolution texture descriptor. This descriptor gains the appropriate information from lighting and image modes. While many developments have been provided for local binary patterns, these operators are sensitive to noise. It also sometimes leads to the description of different patterns of structure with a homogeneous binary code that inevitably reduces their discrimination ability. This research provides an overview of the local binary pattern method, which includes several of the newer variables. Then, to overcome the inefficiencies of variety of local binary patterns, a robust framework of binary pattern called completed robust binary pattern has been proposed that the value of each central pixel is replaced with the average values of the gray intensity of three-to-three square in vicinity. The proposed method is a quick tool with high accuracy for classification of seabed images that comparison of simulation results with other well-known methods, indicating the efficiency of proposed algorithm.

Key words: Local Binary Pattern(LBP), Classification, Segmentation, Seabed, Acoustic Texture Images.

1. Corresponding author (BabakGoudarzi2010@gmail.com)

Contents

**Iranian Journal of
Information Technology & Communication**
No. 25-26, Vol.7, September-March 2016

▪ New changes of local binary patterns and classification and segmentation of seabed images	1
Babak Goudarzi, Reza Javidan, M. Javad Dehghani	
▪ A proper method for the advertising email classification based on user's	21
Rahim Hazratgholizadeh, Mohammad Fathian	
▪ Extracting the Information Technology architecture of hospitals with approach of Implementing in Iran	37
Atefeh.S Haghithoseini, Hosein Boubarshad, Fatemeh Saghafi, Hadi Zare	
▪ A Novel Model for detecting intrusion with Mobile Agent and Game theory	47
Amin Nezarat, Mehdi Raja, G.H Dastghaibfard	
▪ Integrating Data Envelopment Analysis and Decision Tree Models in Order to Evaluate Information Technology-Based Units	61
Amir Amini, Alireza Alinezhad, Somayeh Shafaghizadeh	
▪ Understanding the users' initial trust of social networks	81
Mohsen Akbari, Kamran Zahedfar, Zahra Ayagh	
▪ Instruction for Authors	99
▪ Abstracts in English	105
