فصلنامه علمی- پژوهشی

فناوری اطلاعات و ارتباطات ایران

سال یازدهم، شماره‌های ۳۹و۴۰، بهار و تابستان ۱۳۹۸

صاحب امتیاز: انجمن فناوری اطلاعات و ارتباطات ایران

مدیر مسؤول: دکتر مسعود شفیعی

سردبیر: دکتر محمدشهرام معین

شمارگان: ۱۰۰۰

حروفچینی و صفحه آرایی: سمیه آهاری

هیئت تحریریه:

فرخ حجت کاشانی (استاد)، دانشگاه علم وصنعت ایران

سید احمدرضا شرافت (استاد)، دانشگاه تربیت مدرّس

مسعود شفیعی (استاد)، دانشگاه صنعتی امیرکبیر

محمدرضا عارف (استاد)، دانشگاه صنعتی شریف

کریم فائز (استاد)، دانشگاه صنعتی امیرکبیر

رضا فرجی دانا (استاد)، دانشگاه تهران

کمال محامدپور (استاد)، دانشگاه خواجه نصیرالدین طوسی

فرخ مروستی (استاد)، دانشگاه صنعتی شریف

سیداحمد معتمدی (استاد)، دانشگاه صنعتی امیرکبیر

روزبه معینی (استاد)، دانشگاه صنعتی امیرکبیر

کاظم اکبری (دانشیار)، دانشگاه صنعتی امیرکبیر

کامبیز بدیع (دانشیار)، مرکز تحقیقات مخابرات ایران

علی اکبر جلالی (دانشیار)، دانشگاه علم و صنعت ایران

حمیدرضا ربیعی (دانشیار)، دانشگاه صنعتی شریف

امیر البدوی (دانشیار)، دانشگاه تربیت مدرس

محمد تشنه لب (دانشیار)، دانشگاه خواجه نصیر طوسی

اعضای هیئت مدیره:

- **اعضای اصلی:**

دکتر مسعود شفیعی**(رئیس انجمن)**
دانشگاه صنعتی امیرکبیر

دکتر عباس محمدی
دانشگاه صنعتی امیرکبیر

مهندس ناصر یوسف‌پور
شرکت مخابرات همراه اول

دکتر سید علیرضا هاشمی گلپایگانی
دانشگاه صنعتی امیرکبیر

مهندس فرهاد رضایی
شرکت کوشا تجارت نوپدید

دکتر علیرضا اهری لاحق
شرکت داده پردازی ایران

دکتر محمد جوادیان
دانشگاه صنعتی کرمانشاه

مهندس سیدعلی صموتی
دانشگاه فنی حرفه ای شهید شمسی پور

دکتر علیرضا یاری
پژوهشگاه فناوری اطلاعات و ارتباطات

دکتر اکبر کاری دولت آبادی
دانشگاه شهید ستاری

دکتر جمشید شنبه زاده
دانشگاه خوارزمی

- **اعضای علی‌البدل:**

دکتر مهدی شجری (دانشگاه صنعتی امیرکبیر)

مهندس امیرارسلان داداشی جوردهی (شرکت ارتباطات سیار ایران)

بازرسان:

- **بازرس اصلی:**

دکتر محمود کمره (دانشگاه تهران)

- **بازرس علی‌البدل:**

دکتر امیر شهاب شاهمیری (دانشگاه جامع علمی کاربردی)

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

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

تلفن: ۶۶۴۸۵۸۵۶ ، تلفکس: ۶۶۴۹۵۴۳۳

سال یازدهم، شماره‌های ۳۹و۴۰، بهار و تابستان ۱۳۹۸
صص: ۱-۱۸

فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

گروه‌بندی همسان یادگیرندگان در محیط یادگیری الکترونیکی به کمک روش خوشه‌بندی شورایی

[*] ملیحه کمرئی [**] غلامعلی منتظر

* کارشناسی ارشد مدیریت فنّاوری اطلاعات، گروه مدیریت، دانشگاه تربیت مدرّس، تهران، ایران

**استاد فنّاوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه تربیت مدرّس، تهران، ایران

تاریخ دریافت: ۱۳۹۸/۰۴/۰۱ تاریخ پذیرش: ۱۳۹۹/۰۳/۱۹

چکیده

یادگیری الکترونیکی به عنوان روشی نوین در امر آموزش و یادگیری در طی سالیان اخیر مورد استقبال فراوانی قرار گرفته است. اکثر سامانه‌های یادگیری الکترونیکی، صرف نظر از تفاوت‌های فردی یادگیرندگان محتوای آموزشی مشابهی را برای همهٔ یادگیرندگان ارائه می‌دهند در حالی‌که در آموزش‌های مبتنی بر وب، علاوه برافزایش فرصت یادگیری باید به ارتقای بازدهی یادگیری نیز توجه شود. سامانه‌های یادگیری تطبیقی برای ارائهٔ محتوای تطبیق‌پذیر با هر یادگیرنده، نیازمند گروه‌بندی یادگیرندگان با علایق مشابه است و برای تحقق این هدف، می‌توان از سبک‌های یادگیری یادگیرندگان بهره گرفت. گروه‌بندی خودکار یادگیرندگان در این محیط به کمک روش‌های خوشه‌بندی امکان پذیر است. به دلیل متفاوت بودن نتایج روش‌های خوشه بندی در تکرارهای مختلف، در این پژوهش از روش خوشه‌بندی شورایی برای ترکیب نتایج خوشه‌بندی پنج روش FCM، K-means ، KNN , SVM و K-medoids برای گروه‌بندی یادگیرندگان در محیط یادگیری الکترونیکی استفاده شده است. نتایج ارزیابی تجربی روش خوشه‌بندی پیشنهادی بر اساس سه شاخص «دیویس – بولدین»، « خلوص و تجمع» و « واریانس » نشان می‌دهد که این روش، کاهش هزینهٔ محاسباتی و دقت و سرعت بیشتری نسبت به سایر روش‌های متداول در شناسایی گروه‌ها داشته است.

واژه‌های کلیدی: یادگیری الکترونیکی، گروه‌بندی یادگیرندگان، خوشه‌بندی، سبک یادگیری، رویکرد شورایی.

۱-مقدمه

در سال‌های اخیر استفاده از یادگیری الکترونیکی در نهادهای آموزشی و محیط‌های تجاری با سرعت زیادی رو به افزایش است. این محیط، فارغ از محدودیت زمان و مکان، دسترسی یادگیرندگان به منابع یادگیری را تسهیل کرده است.

مفهوم یادگیری الکترونیکی به استفاده از فناوری‌های اطلاعات و ارتباطات به منظور ارائهٔ طیف وسیعی از آموزش‌ها اشاره دارد و می‌تواند شامل تمامی فعالیت‌های آموزشی افراد یا گروه‌های برخط و برون‌خط، همزمان یا غیرهمزمان باشد[۲].

اکثر سیستم های یادگیری الکترونیکی، صرف نظر از تفاوت‌های فردی یادگیرندگان مانند توانایی‌ها، اهداف، میزان دانش، سبک یادگیری، پیش زمینه و... محتوای آموزشی یکسانی را برای تمامی یادگیرندگان ارائه

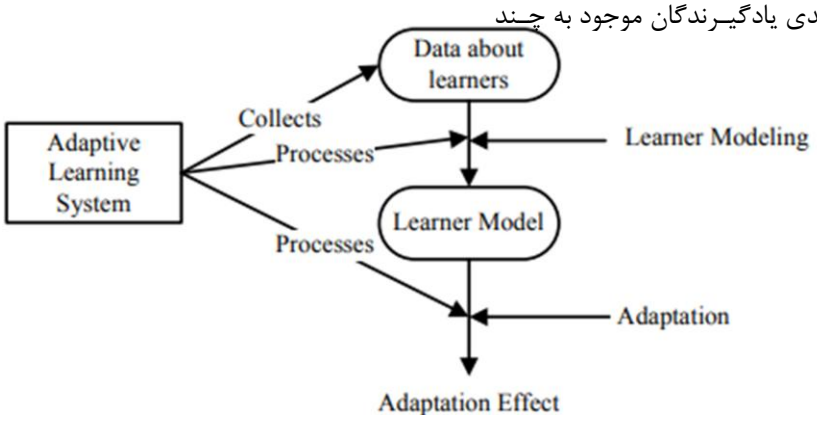
نویسنده عهده‌دار مکاتبات : غلامعلی منتظر montazer@modares.ac.ir

می‌دهند، در حالی‌که از مهم‌ترین اولویت‌های برنامه‌های یادگیری الکترونیکی، ارائهٔ محتوای شخصی شده برای تک‌تک مخاطبان است[۳]. با افزایش کاربرد هوش مصنوعی در آموزش، سامانه‌های یادگیری تطبیقی، آموزش‌های سفارشی را متناسب با تفاوت‌های فردی یادگیرندگان ارائه می‌دهند. شکل ۱، فرایند انطباق در سامانه‌های یادگیری تطبیقی را نشان می‌دهد. در مرحلهٔ اول، سامانه داده‌هایی را از یادگیرندگان جمع‌آوری و پس از پردازش داده‌ها، مدل‌سازی می‌کند. مدل‌سازی می‌تواند بر اساس اطلاعات مختلفی مانند سبک یادگیری[1]، سبک شناختی[2] و ... افراد صورت پذیرد[۴].

ارائهٔ محتوای آموزشی باید متناسب با توانایی‌ها و استعدادهای یادگیرندگان باشد و از آنجا که یادگیرندگان دارای طیف متنوعی هستند و ارائهٔ محتوای شخصی شده و متناسب با هر فرد، امری دشوار و پرهزینه است؛ لذا یکی از مؤثرترین روش‌های ارائهٔ محتوای شخصی شده، گروه بندی یادگیرندگان با علایق مشابه است[۳]. با این روش می‌توان برای هر گروه از یادگیرندگان، آموزش مشابهی را با هزینه ای کمتر ارائه داد. برای گروه‌بندی یادگیرندگان می‌توان از روش‌های متداول خوشه بندی استفاده کرد[۵].

خوشه‌بندی، تقسیم‌بندی یادگیـرندگان موجود به چـند گروه، به گونه‌ای است که یادگیرندگان در یک گروه بیشترین شباهت را و در گروه‌های مختلف، حداکثر تفاوت ممکن را از هم داشته باشند. از مهم‌ترین و پرکاربردترین الگوریتم های خوشه بندی می‌توان به روش‌های K-means، FCM و K-medoids اشاره کرد[۶].

یکی از مشکلات روش‌های خوشه‌بندی، وابستگی پاسخ آنها به نقطهٔ شروع اولیه است که باعث می‌شود در تکرارهای مختلفِ الگوریتم، پاسخ های متفاوتی حاصل شود[۷]. بعضی الگوریتم‌های خوشه بندی به داده‌های دور از مرکز و نویز، حساس هستند[۶]. در این مقاله برای حل این مشکلات، روشی جدید برای گروه‌بندی یادگیرندگان بر مبنای سبک یادگیری آن‌ها در گروه‌های همسان ارائه شده است. در این روش با ترکیب نتایج روش‌های خوشه‌بندی مختلف شامل: FCM، K-means، KNN، SVM و K-medoids احتمال انتخاب یک روش ضعیف را کاهش می‌دهیم و با این کار، ضمن اینکه از پیچیده‌تر کردن روش‌ها جلوگیری شده است، با اعمال الگوریتم های پرکاربرد و ساده‌تر خوشه‌بندی، نتیجهٔ بهتری حاصل می‌شود زیرا روش‌های مختلف می‌توانند معایب همدیگر را پوشش داده و نتیجهٔ بهتری را حاصل کنند.



شکل ۱- فرایند انطباق در سامانه‌های یادگیری تطبیقی[۴]

[1].learning style

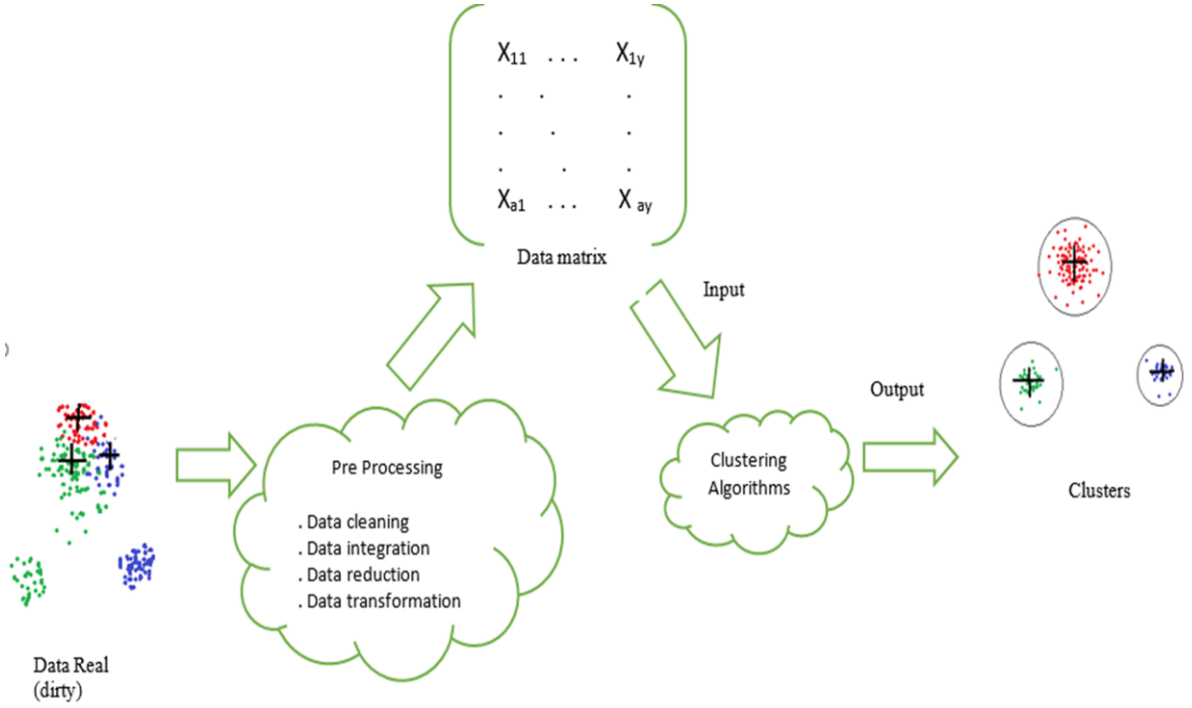
[2]. cognitive style

در این روش ابتدا شاخص‌های رفتاری متناظر با ابعاد مختلف سبک یادگیری مدل فلدر- سیلورمن[3] شناسایی شده و سپس یادگیرندگان برمبنای این رفتارها با استفاده از پنج روش معرفی شده و با استفاده از روش شورایی[4] گروه‌بندی می‌شوند.

ادامهٔ مقاله به صورت زیر تدوین شده است: در بخش ۱، موضوع گروه‌بندی یادگیرندگان تبیین شده است، سبک یادگیری مدل فلدر- سیلورمن و شاخص‌های رفتاری متناظر با آن در محیط یادگیری الکترونیکی بیان شده است و روش شورایی معرفی شده است، نتایج استفاده از این روش در یک دورهٔ آموزشی برخط در بخش ۳ بیان شده و در نهایت بخش۴ به مقایسهٔ نتایج این شیوه با روش‌های دیگر، بحث و نتیجه‌گیری اختصاص یافته است.

۱-۱-گروه‌بندی یادگیرندگان

گروه‌بندی یادگیرندگان به معنای تفکیک یادگیرندگان به گروه‌های مختلف برای تسهیل فرایند یادگیری و ارائهٔ محتوای یکسان برای افراد یک گروه است. با استفاده از این روش می‌توان در هزینه و زمان صرف شده برای ارائهٔ محتوای شخصی و به صورت جداگانه صرفه جویی کرد و محتوای آموزشی را متناسب با گروه‌های همسان با علایق و نیازهای مشابه آماده کرد. خوشه‌بندی نیز فرایندی است که در آن مجموعه‌ای از اشیا به گروه‌هایی مجزا تقسیم می‌شوند به طوری که اشیای یک‌کلاس تا حدامکان به یکدیگر شبیه و با اشیای دیگر کلاس‌ها، متفاوت است. لذا در این پژوهش از خوشه‌بندی که یکی از روش‌های گروه‌بندی یادگیرندگان، درون مجموعه‌ای از دسته‌هاست، استفاده شده است. شکل ۲، فرایند کلی خوشه‌بندی را نشان می‌دهد.



شکل ۲- فرایند خوشه‌بندی [۶]

[3] Felder-Silverman

[4] Ensemble method

یکی از مرسوم‌ترین روش‌های خوشه‌بندی، روش K-means است. هدف اصلی این روش آن است که که مجموع عدم تشابه بین تمام اشیای یک خوشه از مراکز خوشه های متناظرشان کمترین باشد. این الگوریتم به دلیل سادگی پیاده سازی و سرعت عملکرد، محبوبیت زیادی دارد اما مهم ترین مشکل آن، این است که نتایج خوشه بندی، به انتخاب مراکز خوشه های اولیه حساس است و ممکن است به نقطهٔ بهینهٔ محلی همگرا شود[۶,۵].

الگوریتم خوشه‌بندی Fuzzy C-Means (FCM) یکی از پرکاربردترین الگوریتم‌های خوشه‌بندی فازی است که به شدت مشابه الگوریتم k-means است. FCM یکی از روش‌های خوشه‌بندی تفکیکی است که از معیار شباهت فاصلهٔ اقلیدسی برای اندازه‌گیری شباهت داده‌ها استفاده می کند. در این الگوریتم برای همهٔ داده‌ها، درجه‌ای از عضویت تعریف می‌شود و بر اساس آن داده‌ها با درجهٔ عضویت مشخصی به همهٔ خوشه‌ها تعلق می‌یابند. بعد از هر تکرار، مراکز خوشه‌ها و درجهٔ عضویت‌ها به روزرسانی می‌شوند. مهم‌ترین مشکل این الگوریتم، وابستگی نتیجهٔ خوشه‌بندی به نقطهٔ انتخابی اولیه است[۸].

الگوریتم k -medoids از دیگر الگوریتم‌های کاربردی در خوشه‌بندی است که برخلاف‌الگوریتم k-means، مرکزی‌ترین دادهٔ هر خوشه به عنوان مرکز خوشه در نظر گرفته می‌شود. یکی از مشکلات الگوریتم k -medoids این است که نیازمند اجرای تکرارهای بسیار است و این موضوع، سرعت محاسباتی روش را کاهش می دهد[۶].

ماشین بردار پشتیبان (SVM) یکی از روش‌های خوشه‌بندی است که در زمینه‌های مختلفی مورد استفاده قرار می‌گیرد. SVM داده‌ها را به صورت خطی تقسیم‌بندی می‌کند و برای این امر خطی انتخاب می‌شود که حاشیه اطمینان بیشتری داشته باشد. در این الگوریتم آموزش نسبتاً ساده است و بر خلاف شبکه‌های عصبی در نقاط بیشینهٔ محلی گیر نمی‌افتد[۹].

الگوریتم نزدیک‌ترین K همسایه[5] یکی از مسائل بهینه‌سازی برای پیدا کردن نزدیک‌ترین نقطه‌ها در فضاهای متریک است. پس از تعیین مراکز خوشه‌ها با کمک این الگوریتم می‌توان نزدیک‌ترین نقاط به مرکز هر خوشه را یافت[۸]. در این پژوهش، خوشه‌بندی یادگیرندگان بر مبنای سبک یادگیری و با استفاده از رفتارهای شبکه‌ای آن‌ها انجام می‌گیرد؛ لذا در بخش ۳ به معرفی سبک یادگیری می‌پردازیم.

۱-۲-سبک یادگیری

از مؤثرترین رویکردها در مطالعهٔ یادگیری، سبک‌های یادگیری هستند. سبک یادگیری چگونگی تمایل فرد به یادگیری و نحوهٔ انطباق با محیط را نشان می‌دهد[۹]. از جمله سبک‌های یادگیری می‌توان به مدل کلب[6][۱۰]، مدل هانی و مامفورد[7][۱۱] ، مدل فلدر- سیلورمن[۱۲] و... اشاره کرد [۱۳]. در این پژوهش از سبک یادگیری فلدر – سیلورمن استفاده شده است؛ زیرا از میان سبک های یادگیری ارائه شده، مدل فلدرسیلورمن به جهت استفاده مکرر در تحقیقات علمی و ارزیابی نتایج آنها توصیف مناسبی از سبک یادگیری ارائه می‌دهد ؛ این مدل ارزیابی سبک یادگیری فلدرسیلورمن موجب شده که امکان تعیین عددی آن به صورت خودکار میسر شود. [۱۴].

[5] K-Nearest Neighbor (KNN)

[6] Kolb

[7] Honey & Mumford

عددی از سبک یادگیری یادگیرندگان ارائه می‌دهد که این مهم عامل مفیدی برای استفاده از این مدل در سیستم‌های مبتنی بر رایانه به شمار می‌آید. از سوی دیگر وجود نگاشت‌های رفتاری یادگیرنده برای اندازه‌گیری این مدل دارای پنج بعد پردازش[8] (واکنشی-تأملی)، ورودی[9] (دیداری-شنیداری)، فهم[10] (متوالی-سراسری)، ادراک[11](حسی-شهودی) و سازمان[12] (قیاسی-استقرایی) است.

یادگیرندگان حسی دوست دارند مسائل خود را به طور واقعی و با استفاده از روش‌های تعریف شده (روش‌های روشن و واضح) حل کنند و پیچیدگی را دوست ندارند. اما یادگیرندگان شهودی اغلب ترجیح می‌دهند امکانات و روابط را کشف کنند، خلاق و نوآور باشند و تکرار را دوست ندارند. یادگیرندگان واکنشی، کارهای گروهی را دوست دارند و اطلاعات را به وسیله انجام دادن بهتر درک می‌کنند. در مقابل یادگیرندگان تأملی ترجیح می‌دهند امکانات و ارتباطات را کشف کنند. خلاقیت و نوآور بودن را دوست دارند و ازتکرار متنفرند. یادگیرندگان دیداری آنچه را می‌بینند (یعنی مواردی مانند تصاویر، نمودارها، جداول و فیلم ها) بهتر به یاد می‌آورند. آنان نقشهٔ مفهومی را به کمک فهرست کردن نکات کلیدی و قرار دادن آنها در جعبه‌ها یا جاهای خاص‌خود، برای یادگیری خطی تمایل دارند. معمولاً در حل مسائلی که با آن مواجه می‌شوند، مراحل را به طور منطقی و گام به گام طی می‌کنند تا بتوانند مسائل خود را حل کنند. این دسته از یادگیرندگان مطالب را به چیزهایی که تاکنون یادگرفته ربط می‌دهند و در مقابل یادگیرندگان سراسری، قبل از شروع مطالعه، ابتدا مروری کلی بر مطالب مورد خود در نظر می‌گیرند. یادگیرندگان شنیداری، اطلاعات را بیشتر از طریق کلمات (چه به صورت کلمات نوشته شده یا توضیحات کلامی) بهتر به یاد می‌آورند. این دسته از یادگیرندگان اطلاعات را خلاصه می‌کنند و کارهای گروهی را دوست دارند[۱].

یادگیرندگان متوالی به درک مطالب در مراحل منظم و اند، نظر خواهند داشت. این دسته از یادگیرندگان کل‌نگرند و در برخورد با مسائل پیچیده، ابتدا می‌کوشند آنها را درک و سپس حل کنند[۱].

فلدر و سولومان در سال ۱۹۹۸ میلادی با حذف بُعد سازمان، پرسش‌نامه‌ای را برای ارزیابی چهار بُعد دیگر طراحی کرده‌اند. این پرسش‌نامه دارای ۴۴ پرسش (۱۱ پرسش دو گزینه‌ای برای هر بُعد با امتیاز ۰ یا ۱) است. مجموع امتیازات ۱۱ پرسش، که عددی بین ۰ تا ۱۱ خواهد بود، نمرهٔ فرد در آن بُعد را نشان می‌دهد[۱۲].

برای گروه‌بندی یادگیرندگان بر مبنای رفتارهای آنان و شناسایی سبک یادگیری با استفاده از الگوهای رفتاری موجود در گروه‌های حاصل، لازم است ابتدا شاخص‌های رفتاری موردنظر برای گروه‌بندی شناسایی شود. این شاخص‌ها، ویژگی‌هایی از یادگیرنده هستند که گروه‌بندی و تفکیک یادگیرندگان در گروه‌های همسان، بر مبنای آنها انجام می‌شود[۱۵]. بدین منظور شاخص‌های رفتاری معرفی شده در جدول ۱ به ابعاد مختلف سبک یادگیری متناظر شده‌اند[۵].

جدول ۱- شاخص رفتاری یادگیری متناظر با ابعاد سبک یادگیری فلدر- سیلورمن [۵]

ابعاد	شاخص‌های متناظر ثبت شده از یادگیرنده در سامانهٔ آموزشی
ادراک	نوع مطالعهٔ مواد (نوع مطالب انتخابی (خلاصه یا جامع و عمیق))
	انواع نمایش اسلاید
	زمان اختصاص یافته برای خواندن مفاهیم و نظریه‌ها (درس-پارهای نظری و غیر عملی)
	زمان اختصاص یافته برای خواندن مثال‌ها و واقعیت‌ها
	زمان اختصاص داده شده برای انجام تمرین‌های اضافی
	تعداد مثال‌های مورد مطالعه اضافی
	زمان اختصاصی به آزمون و بازبینی
پردازش	مشارکت در گفتگوها
	استفاده از رایانامه و اتاق گفتگو
	مشارکت در بحث/کار گروهی

[8] processes
[9] input
[10] understanding
[11] perception
[12] organization

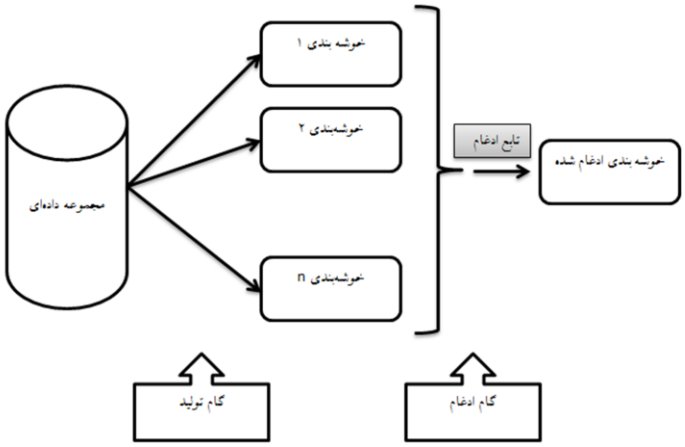
	انتخاب پروژه‌های فردی یا گروهی کلاس
ورودی	گوش دادن به سخنرانی‌ها (پرونده‌های صوتی درس)
	استفاده از محتوای ویدیویی درس
	خواندن محتوای متنی درس
	انواع نمایش اسلایدها (نوع توصیه‌های انتخابی)
	میزان استفاده از نمودار و شکل
فهم	الگوی دسترسی به محتواهای دورهٔ آموزشی
	زمان اختصاص یافته به هدف و کلیّات درس

۱-۳-روش شورایی برای گروه‌بندی یادگیرندگان

یکی از روش های مورد استفاده برای افزایش دقت خوشه‌بندی، ترکیب نتایج چندین روش خوشه‌بندی است[۱۸,۱۶]. موفقیت رویکرد شورایی در یادگیری نظارت شده، انگیزه‌ای جدّی برای بهره مندی از این روش در یادگیری بی‌نظارت[۱۳]، ایجاد کرد.در مطالعات سال‌های اخیر، رویکرد شورایی[۱۴] نسبت به دیگر روش‌های خوشه‌بندی از محبوبیت بیشتری برخوردار بوده است. ایدهٔ خوشه‌بندی شورایی از ترکیب نتایج چندین خوشه‌بندی برای حصول راه حلی برتر گرفته شده است.

روش‌های مختلفی برای ترکیب نتایج خوشه‌بندی ها وجود دارد به طور مثال، توابع اجماع[۱۵] مختلفی برای ترکیب نتایج خوشه‌بندی پیشنهاد شده است زیرا پس از تشکیل خوشه‌های حاصل ازالگوریتم های مختلف، نیازمند تابعی هستیم که نتایج خوشه‌بندی ها را ترکیب و نتیجهٔ نهایی را جاصل کند[۱۷]. یکی از ساده ترین روش‌ها، روش رأی اکثریت است [۱۸] که دراین روش، برای هرداده، بهترین کلاس را که توسط الگوریتم های مختلف انتخاب شده در نظر می‌گیرند و برچسب کلاسی را انتخاب می‌کند که بیشترین تکرار را در میان مجموعه خروجی‌های جدید داشته باشد.

روش خوشه‌بندی شورایی شامل دو گام اصلی است[۱۹]:
تولید[۱۶]: در این گام افرازهای مجموعه داده‌ای تشکیل می‌شود؛ تابع ادغامی[۱۷]: در این گام افراز جدید، با یکپارچه‌سازی همه افرازهای به دست آمده در گام تولید محاسبه می‌شود.
شکل ۳ روند کلی روش خوشه‌بندی شورایی را نشان می‌دهد[۱۹].



شکل ۳-روند انجام خوشه‌بندی شورایی[۱۹]

[1] unsupervised learning
[14] Ensemble method
[15] consensus functions
[16] Generation
[17] Consensus Function

در ادامه، گام‌های روش خوشه‌بندی شورایی، که در این مقاله مورد استفاده قرار گرفته است، بیان می‌شود.

رویکرد تولید

«تولید»، اولین گام در روش خوشه‌بندی شورایی است که در آن مجموعهٔ خوشه‌بندی‌های مورد نظر برای ادغام ایجاد می‌شود. با توجه به اینکه نتایج نهایی به خوشه‌بندی تولید در این گام وابسته است، انتخاب فرایند تولید مناسب از اهمیت به‌سزایی برخوردار است. در حالت کلی محدودیتی برای روش‌های استفاده شده در این گام وجود ندارد و می‌توان از نتایج اجراهای مختلف یک الگوریتم خوشه‌بندی و یا چند الگوریتم خوشه‌بندی مختلف استفاده کرد. ضمن اینکه استفاده از زیرمجموعه‌های مختلف از ویژگی‌های متناظر با هر داده، تنظیم متفاوت پارامترهای مختلف روش مورد استفاده و ...، دیگر رویکردهایی هستند که در این گام استفاده می‌شود. البته استفاده از چند روش خوشه‌بندی مختلف که به تولید اطلاعات بیشتر دربارهٔ مجموعهٔ داده‌ای منجر شود، رویکردی است که بیشتر مورد تأکید قرار گرفته است[۱۹]. در این مقاله، از پنج روش خوشه‌بندی K-means، FCM، KNN-K، medoids و SVM در گام تولید روش خوشه‌بندی شورایی استفاده شده است.

تابع ادغام

انتخاب تابع ادغامِ مناسب، یکی از مهم‌ترین مسائل در روش خوشه‌بندی شورایی است. این تابع باید قابلیت بهبود نتایج خوشه‌بندی‌های به دست‌آمده درگام تولید را داشته باشد. یکی از مهم‌ترین رویکردها در تابع ادغام، هم‌رخدادی[18] نمونه‌ها و یا رأی اکثریت۱۹ است. در این روش تحلیل می‌شود که یک نمونهٔ داده‌ای چند بار در یک خوشه قرار گرفته یا دو نمونهٔ داده‌ای چند بار در خوشه-های مشابه قرارگرفته‌اند. شایان ذکر است در برخی موارد بر اساس تخمین دقت روش‌های خوشه‌بندی، ضریب وزنی برای نتیجهٔ هر روش در نظر گرفته می‌شود[۱۹]؛ در این مقاله نیز از روش رأی اکثریت استفاده شده است.

۱-۴-ارزیابی روش خوشه‌بندی شورایی

یکی از اصلی‌ترین نکات در خوشه‌بندی، انتخاب تعداد خوشه‌های مناسب است. بدین منظور باید نمونه‌های موجود در یک خوشه تا حد امکان شبیه به یکدیگر و نمونه‌های متعلق به خوشه‌های متفاوت تا حد امکان با یکدیگر نامتشابه باشند. به تعبیر دیگر خوشه‌ها باید بیشترین فشردگی[20] را داشته باشند و تا حد امکان جدایی[21] آنها نیز زیاد باشد. در ارزیابی روش‌های خوشه-بندی می‌توان از شاخص‌های اعتبارسنجی خوشه‌بندی استفاده کرد. بدین منظور در این مقاله از سه معیار شاخص دیویس-بولدین[۲۰]، شاخص خلوص و تجمع[22] [۲۱] و شاخص واریانس[۲۲] استفاده شده است.

شاخص اعتبار سنجی دیویس-بولدین امکان مقایسهٔ دقت خوشه‌بندی روش‌های مختلف را فراهم می‌کند. این شاخص با اندازه‌گیری فشردگی درون هر خوشه و میزان جدایی میان خوشه‌ها، کیفیت خوشه‌بندی را تعیین می‌کند. این شاخص تنها وابسته به مجموعه داده و اطلاعات عضویت آن‌ها در خوشه‌هاست و متوسط شباهت میان هر خوشه و خوشه با بیشترین شباهت به آن را

[18] co-occurrence

[19] Majority Voting

[20] Compactness

[21] Separation

[22] Purity and Gathering (P&G)

کمینه می‌سازد. این شاخص با رابطهٔ (۱) تعریف می‌شود[۲۰]:

$$DB = \frac{1}{K}\sum_{k=1}^{k} Max\left(\frac{diam\,(C_k)+diam\,(C_j)}{dist\,(C_k,C_j)}\right) \quad (1)$$

$$J= 1 ,\dots k \qquad K\neq j$$

در این رابطه، K تعداد خوشه‌ها (diam(Ck میزان پراکندگی درون خوشهٔ Ck را نشان می‌دهد که با رابطهٔ (۲) تعریف می شود:

$$diam(C_k) = \sqrt[q]{\frac{1}{T_k}\sum_{j=1}^{T_k}\left|X_j - A_k\right|^q} \quad (2)$$

در این رابطه T_k تعداد اعضای خوشه‌ی C_k ، X_j عنصر jام از خوشه C_k ، A_k مرکز خوشه C_k و $|X_j - A_k|^q$ فاصلهٔ میان عنصر j ام خوشه از مرکز آن است.($dist(C_k , C_j)$ در رابطهٔ (۱)، میزان جدایی خوشه C_k و C_j را نشان می‌دهد و با رابطهٔ (۳) محاسبه می‌شود که در آن a_{ik} عنصر i ام از خوشه‌ی k ام است:

$$dist(C_k, C_j) = \sqrt[p]{\sum_{i=1}^{n}\left|a_{ik} - a_{ij}\right|^p} \quad (3)$$

در این پژوهش، با قرار دادن $p = q = 2$ در روابط (۲) و (۳)، فاصله‌ها بر اساس فاصلهٔ اقلیدسی[۲۳] محاسبه می‌شود. بدیهی است با توجه به تعریف شاخص دیویس-بولدین، هرچه مقدار آن کمتر باشد، نشان‌دهندهٔ خوشه‌بندی بهتر است.

با استفاده از شاخص خلوص و تجمع، می‌توان کیفیت خوشه‌های ایجاد شده را بر اساس سبک یادگیری یادگیرندگان مورد ارزیابی قرار داد. به عبارت دیگر، در این شاخص از سبک یادگیری به عنوان ناظر خارجی، برای ارزیابی کیفیت گروه‌های ایجاد شده از منظر خلوص و شباهت استفاده می‌شود. این شاخص با محاسبهٔ شاخص خلوص و شاخص تجمع و ترکیب آن‌ها با یکدیگر محاسبه می‌شود. شاخص خلوص به کمک رابطهٔ (۴) محاسبه می‌شود[۲۱]:

$$Pur(C) = \sum_{c=1}^{K} \max\left\{\frac{N_{cl}}{N_c} \quad l \in L\right\} /K \quad (4)$$

که K تعداد خوشه‌ها، N_{cl} تعداد عناصر داده‌ای در خوشه C با برچسب l ؛ و N_c تعداد عناصر خوشه C است.

شاخص تجمع کیفیت خوشه‌بندی را از منظر میزان توانمندی روش خوشه‌بندی در گردآوری مناسب داده‌ها در کنار یکدیگر ارزیابی می‌کند. شاخص تجمع به صورت زیر محاسبه می‌شود[۲۱]:

$$Gat\ (c) = \sum_{l=1}^{l=m} Max\left\{\frac{N_{lc}}{N_l} \quad c \in\right\} / m \quad (5)$$

که m تعداد نشانه‌ها (اعداد سبک یادگیری ۰ تا ۱۱)، N_{cl} تعداد نشانه‌های l موجود در خوشه C و N_l تعداد داده‌هایی است که برچسب آن‌ها l است. شاخص ترکیبی خلوص و تجمع، به شرح زیر محاسبه می‌شود:

$$PG(C) = Pur(C) + Gat(C) \quad (6)$$

بیشتر بودن شاخص خلوص و تجمع ناظر به کیفیت بالاتر خوشه‌بندی است.

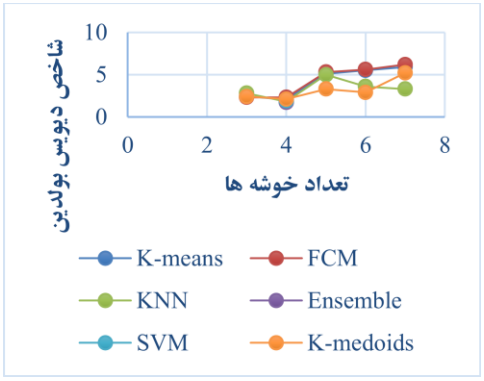
۲-روش تحقیق

۲-۱-گروه‌بندی همگن یادگیرندگان در محیط الکترونیکی به کمک روش شورایی

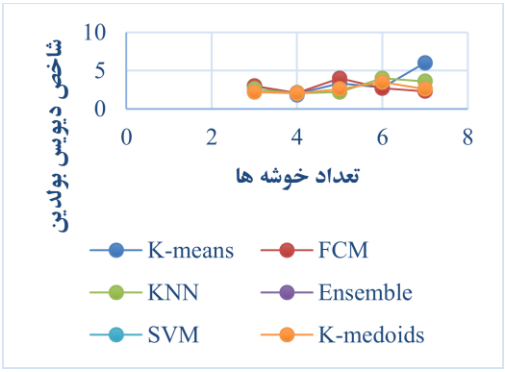
برای ارزیابی روش پیشنهادی، از اطلاعات ثبت‌شده دورهٔ آموزش الکترونیکی درس برنامه‌نویسی C++ استفاده شده است. تعداد یادگیرندگان حاضر در محیط آزمایش ۹۸ نفر بودند که در طی گذراندن این درس، شاخص‌های تعریف شده در جدول ۱ مرتبط با رفتار شبکه‌ای آنان در چهار بُعد

[23] Euclidian

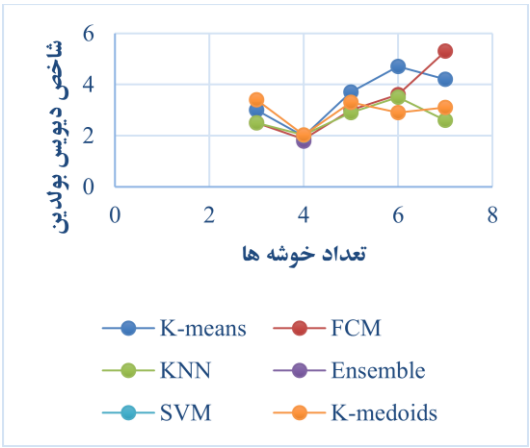
ادراک، پردازش، ورودی و فهم مدل سبک یادگیری فلدر- سیلورمن از پرونده‌های ثبت‌شده در سامانه استخراج شده است. از سوی دیگر، پرسش‌نامهٔ اندازه‌گیری سبک یادگیری فلدر- سیلورمن در اختیار دانشجویان قرارگرفت و از آنان خواسته شد تا آن را تکمیل کنند. سپس بر مبنای اطلاعات رفتاری جمع‌آوری شده، گروه‌بندی یادگیرندگان با استفاده از پنج روش K-means، FCM، KNN، SVM، K-medoids و هم‌چنین روش خوشه‌بندی شورایی انجام شده است. با توجه به اینکه روش‌های مطرح شده، قادر به تعیین خودکار تعداد بهینهٔ خوشه‌ها نیستند، خوشه‌بندی داده‌ها به ازای ۳، ۴، ۵، ۶ و ۷ خوشه توسط این روش‌ها انجام و با محاسبهٔ شاخص دیویس - بولدین تعداد خوشهٔ بهینه برای هر روش تعیین شد. در روشFCM هر داده به خوشه‌ای منتسب می‌شود که بیشترین درجهٔ تعلق به آن را داشته است. شکل ۲، شاخص دیویس- بولدین را برای خوشه‌بندی‌های انجام شده در ابعاد مختلف سبک یادگیری فلدر- سیلورمن برای تعداد خوشه‌های ۳ و ۴ و ۵ و ۶ نشان می‌دهد. با توجه به اینکه مقدار شاخص دیویس – بولدین بیانگر آن‌است که تعداد بهینهٔ خوشه‌ها برای همهٔ روش‌ها برابر با ۴ است، این شاخص برای روش خوشه‌بندی شورایی و SVM تنها در حالت تعداد خوشهٔ بهینه محاسبه شده است. همان‌گونه که در این شکل مشخص است مقدار شاخص دیویس- بولدین برای روش خوشه‌بندی شورایی نسبت به سایر روش‌ها در مقدار بهینهٔ تعداد خوشه (یعنی ۴) کمتر است که نشان‌دهندهٔ عملکرد خوشه‌بندی بهتر در این روش است.



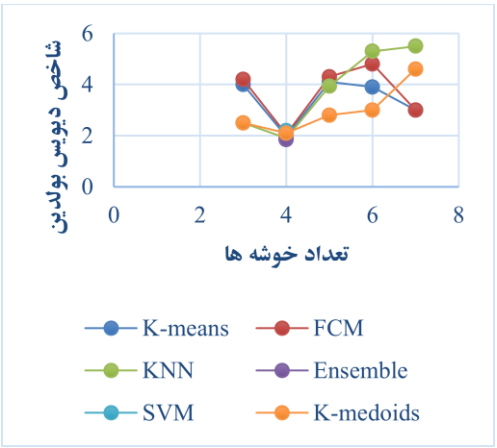
الف. بعد ادراک



ب. بعد ورودی



ج. بعد فهم



د. بعد پردازش

شکل ۴- مقایسهٔ عملکرد روش خوشه‌بندی شورایی با سایر روش‌های متداول خوشه‌بندی بر اساس شاخص دیویس- بولدین برای ابعاد سبک یادگیری

۳- نتایج و بحث

جدول ۲ نتایج محاسبهٔ معیار خلوص و تجمع برای ۶ الگوریتم را در ابعاد مختلف سبک یادگیری فلدر- سیلورمن نشان می‌دهد. همان‌گونه که در این جدول مشخص است در همهٔ ابعاد مقدار معیار خلوص و تجمع در روش شورایی بیشتر از سایر روش‌هاست که عملکرد خوشه‌بندی بهتر را نشان می‌دهد. به عبارت دیگر، در گروه‌های تشکیل شده در روش خوشه‌بندی شورایی، شباهت سبک یادگیری یادگیرندگان بیشتر بوده است.

جدول ۲- مقایسهٔ روش خوشه‌بندی شورایی با سایر روش‌های متداول خوشه‌بندی بر مبنای شاخص خلوص و تجمع

بعد	معیار	FCM	K-means	KNN	SVM	K-medoids	Ensemble
ادراک	Gat(C)	۰/۹۱۰	۰/۹۲۰	۰/۹۴۰	۰/۹۴۰	۰/۸۹۱	۰/۹۷۳
	Pur(C)	۰/۴۵۰	۰/۴۵۰	۰/۴۴۰	۰/۴۵۰	۰/۴۲۲	۰/۴۶۲
	PG(C)	۱/۳۶۰	۱/۳۷۰	۱/۳۸۰	۱/۳۹۰	۱/۳۱۳	۱/۴۳۵*
ورودی	Gat(C)	۰/۸۹۲	۰/۸۹۲	۰/۹۰۱	۰/۹۰۱	۰/۹۰۲	۰/۹۱۳
	Pur(C)	۰/۴۸۳	۰/۴۸۳	۰/۴۷۵	۰/۴۸۵	۰/۴۷۳	۰/۴۸۳
	PG(C)	۱/۳۷۵	۱/۳۷۵	۱/۳۷۶	۱/۳۸۶	۱/۳۷۵	۱/۳۹۶*
فهم	Gat(C)	۰/۸۳۰	۰/۸۳۰	۰/۸۴۹	۰/۸۷۳	۰/۸۶۱	۰/۸۹۸
	Pur(C)	۰/۳۸۹	۰/۳۸۹	۰/۳۹۶	۰/۴۱۵	۰/۳۷۳	۰/۳۹۹
	PG(C)	۱/۲۱۹	۱/۲۱۹	۱/۲۴۵	۱/۲۸۸	۱/۲۳۴	۱/۲۹۷*
پردازش	Gat(C)	۰/۸۶۱	۰/۸۹۴	۰/۸۹۴	۰/۸۵۹	۰/۸۵۴	۰/۸۸۴
	Pur(C)	۰/۴۵۴	۰/۴۴۹	۰/۴۱۵	۰/۴۷۰	۰/۴۴۵	۰/۴۷۱
	PG(C)	۱/۳۱۵	۱/۳۴۳	۱/۳۰۹	۱/۳۲۹	۱/۲۹۹	۱/۳۵۵*

علامت * بزرگترین مقدار شاخص خلوص و تجمع را نشان می‌دهد که ناظر به خوشه‌بندی بهتر است.

یکی دیگر از معیارهای اعتبار سنجی خوشه‌ها میزان فشردگی آنها است. اعضای یک خوشه باید تا حد ممکن به یکدیگر نزدیک باشند؛ یکی از معیارهای متداول فشردگی، حداقل بودن واریانس هر خوشه است[۲۲]. جدول ۳ نتایج محاسبهٔ واریانس برای ۶ الگوریتم را در ابعاد مختلف سبک یادگیری فلدر- سیلورمن نشان می‌دهد. طبق این جدول در همهٔ ابعاد، مقدار واریانس در روش شورایی کمتر از سایر روش‌هاست که عملکرد خوشه‌بندی بهتر را نشان می‌دهد. به عبارت دیگر، در گروه‌های حاصل از روش خوشه‌بندی شورایی، نزدیکی سبک یادگیری یادگیرندگان به یکدیگر بیشتر بوده است.

جدول ۳- مقایسهٔ روش خوشه‌بندی شورایی بر مبنای واریانس خوشه ها

بعد	FCM	K-means	KNN	SVM	K-medoids	Ensemble
ادراک	۳/۵۱	۳/۲۵	۳/۱۴	۳/۱۷	۳/۷۹	۳/۱۱
ورودی	۳/۱۲	۳/۱۲	۳/۳۱	۳/۱۹	۳/۶۳	۳/۰۹
فهم	۴/۳۵	۴/۳۶	۴/۵۸	۴/۳۸	۴/۹۸	۴/۳۳
پردازش	۳/۵۱	۳/۵۲	۳/۴۴	۲/۹۹	۳/۷۱	۲/۹۸

برای ارزیابی دقت روش شورایی در تشکیل گروه‌هایی با سبک یادگیری مشابه، می‌توان مقایسه‌ای بین اعداد سبک یادگیری متناظر با یادگیرندگان هرگروه انجام داد. جدول‌های ۴ تا ۷، نتایج خوشه‌بندی شورایی یادگیرندگان را در ابعاد مختلف با نمایش سبک یادگیری آن‌ها نشان می‌دهد. مقادیر این جدول‌ها بیانگر آن است که روش شورایی توانسته‌است یادگیرندگان با سبک یادگیری مشابه را با خطای بسیار ناچیز در خوشه‌های یکسان قرار دهد. با استفاده از خوشه‌های حاصل و سبک غالب در هر خوشه، می‌توان سبک یادگیری یادگیرندگان جدید را بر اساس رفتار آنان و خوشه‌ای که در آن قرار می‌گیرند، شناسایی کرد. استفاده از روش‌های خودکار، مانند روش خوشه‌بندی شورایی، که از رفتار یادگیرندگان برای شناسایی سبک یادگیری آنان استفاده می‌کند، می‌تواند تأثیر جهت گیری‌ها و اشتباهات افراد در تکمیل پرسش‌نامه را حذف کند و بر مبنای رفتارهای شبکه‌ای آن‌ها در محیط یادگیری الکترونیکی، سبک یادگیری را با دقت قابل قبولی شناسایی کرد.

جدول ۴ - نتایج خوشه بندی شورایی یادگیرندگان در بعد ادراک بر مبنای سبک یادگیری

خوشه ها	عدد سبک یادگیری یادگیرندگان	تعداد یادگیرندگان شهودی	تعداد یادگیرندگان حسی
خوشه ۱	۱۰-۹-۱۰-۱۰-۱۰-۸-۱۰-۹-۹-۸-۱۰-۱۰-۹-۹-۹-۹-۹-۹-۹-۱۱--۹-۹-۹-۱۰-۱۰-۱۱-۱۰-۱۱-۹-۹	۲۷	۰
خوشه ۲	۳-۲-۲-۰-۲-۰-۲-۳-۲-۲-۲-۰-۱-۲-۰-۳-۳-۲-۰-۲-۲-۲	۰	۲۲
خوشه ۳	۴-۴-۳-۵-۶-۴-۶-۵-۵-۵-۴-۵-۴-۵-۵-۵-۵-۳-۳-۳-۵-۴-۳-۴-۳	۲	۲۳
خوشه ۴	۸-۶-۷-۷-۵-۷-۸-۶-۸-۸-۷-۷-۸-۷-۷-۷-۷-۸-۶-۸-۶-۷-۸-۶-۸-۶	۲۳	۱

جدول ۵ - نتایج خوشه بندی شورایی یادگیرندگان در بعد پردازش بر مبنای سبک یادگیری

خوشه ها	عدد سبک یادگیری یادگیرندگان	تعداد یادگیرندگان واکنشی	تعداد یادگیرندگان تأملی
خوشه ۱	۲-۲-۴-۱-۰-۳,-۲-۳-۰-۲-۲-۲-۲-۰-۱-۲-۰-۲-۳-۲-۱-۲	۲۲	۰
خوشه ۲	۵-۴-۴-۶-۳-۵-۵-۵-۵-۵-۵-۴-۵-۵-۵-۴-۳-۴-۵-۴-۵-۵-۶-۵-۴-۵-۴-۴-۴-۴	۲۷	۲
خوشه ۳	۷-۷-۸-۷-۷-۶-۷-۶-۸-۷-۸-۷-۷-۸-۸-۸-۷-۷-۷-۷-۸-۶	۰	۲۲
خوشه ۴	۹-۱۱-۱۰-۹-۹-۱۰-۹-۱۰-۱۱-۱۰-۱۰-۱۰-۱۱-۹-۸-۹-۸-۹-۱۰-۱۰-۹-۱۰-۹	۰	۲۵

جدول ۶ - نتایج خوشه بندی شورایی یادگیرندگان در بعد ورودی بر مبنای سبک یادگیری

خوشه ها	عدد سبک یادگیری یادگیرندگان	تعداد یادگیرندگان دیداری	تعداد یادگیرندگان شنیداری
خوشه ۱	۱-۱-۳-۱-۱-۱-۲-۱-۱-۳-۲-۱--۰-۲-۲-۲-۱-۳-۲-۱	۲۰	۰
خوشه ۲	۴-۶-۳-۵-۶-۵-۶-۳-۴-۴-۵-۵-۳-۵-۴-۵-۵-۴-۵-۵-۳-۴-۵-۵-۵-۳	۲۳	۳
خوشه ۳	۶-۴-۶-۷-۷-۸-۷-۶-۷-۸-۷-۷-۸-۸-۶-۷-۷-۶-۷-۷-۷-۷-۷-۸-۸-۷-۷-۷-۷	۱	۲۸
خوشه ۴	۹-۱۰-۹-۱۱-۱۰-۱۰-۱۰-۹-۸-۱۱-۱۱-۱۱-۹-۱۰-۱۰-۱۱-۹-۹-۹-۱۱-۹-۱۰-۹-۹	۰	۲۳

جدول ۷ - نتایج خوشه بندی شورایی یادگیرندگان در بعد فهم بر مبنای سبک یادگیری

خوشه ها	عدد سبک یادگیری یادگیرندگان	تعداد یادگیرندگان متوالی	تعداد یادگیرندگان سراسری
خوشه ۱	۱-۴-۲-۱-۱-۲--۰-۱-۲-۰-۰-۰-۰-۰-۱-۰-۰-۰-۰-۰-۱-۲-۱-۱-۱-۲-۱	۲۳	۰
خوشه ۲	۵-۵-۲-۴-۲-۲-۵-۳-۵-۴-۵-۳-۵-۶-۵-۴-۲-۵-۵-۳-۳-۳-۳-۳-۵-۴-۵-۴-۴	۲۷	۱
خوشه ۳	۶-۶-۶-۶-۷-۷-۷-۷-۷-۵-۸-۷-۷-۷-۶-۶-۶-۶-۷-۸-۵-۶-۷-۶-۵-۸-۷-۶-۷	۴	۲۲
خوشه ۴	۱۱-۸-۹-۹-۱۰-۹-۱۱-۱۱-۹-۱۱-۸-۸-۱۰-۸-۸-۹-۸-۸-۸-۸-۱۰-۹-۱۰-۱۰-۸-۸-۸-۸-۹	۰	۲۲

یکی دیگر از شیوه های تلخیص داده ها، استفاده از شاخص های تمرکز است. از معروف‌ترین شاخص‌های تمرکز می‌توان به میانگین (مرکز ثقل داده‌ها)، میانه (مقدار میانی در صورتی که داده‌ها به ترتیب قرار گیرند) و نما(ویژگی با بیشترین فراوانی) اشاره کرد. تحلیل شاخص های گرایش به مرکز در کنار شاخص های پراکندگی بسیار است. همانطور که طبق جدول ۳ واریانس کمتر روش خوشه بندی شورایی بیانگر فشردگی بیشتر داده ها می باشد شاخص های تمرکز جدول ۸ نیز ناظر به خوشه بندی در محدوده دقیق هر خوشه می باشند.

جدول۸- نتایج شاخص های توصیفی در هر خوشه در خوشه بندی شورایی

ابعاد	شاخص های توصیفی	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ۴
ادراک	میانگین	۱	۴	۶	۹
	میانه	۲	۵	۷	۹
	مد	۲	۵	۷	۹
پردازش	میانگین	۱	۴	۷	۹
	میانه	۲	۵	۷	۱۰
	مد	۲	۵	۷	۹
ورودی	میانگین	۲	۴	۶	۹
	میانه	۱	۵	۷	۱۰
	مد	۱	۵	۷	۹
فهم	میانگین	۱	۳	۶	۹
	میانه	۱	۴	۷	۹
	مد	۱	۵	۷	۸

ارزیابی یک خوشه بندی میتوان از شاخص هایی نظیر Accuracy, Precision, Recall ,F-Value استفاده نمود. جدول ۹ این شاخص ها را در خوشه بندی شورایی نشان میدهد.

Precision دقت سیستم در میان داده های پیش بینی شده است. Recall نسبت تعداد داده های پیش بینی شده، به تعداد کل داده های مورد انتظار برای پیش بینی است. F-Value, هم یک نوع میانگین بین پارامتر دقت و پارامتر یادآوری است. این پارامترها بر اساس روابط ۷ الی ۱۰ نعریف

TP تعداد افرادی هستند که به درستی به کلاس مورد نظر اختصاص داده شده اند؛ FP تعداد افرادی هستند که نادرست به کلاس مورد نظر اختصاص داده شده اند؛ FN تعداد افرادی هستند که در کلاس مورد نظر دسته بندی نشده اند اما به آن کلاس اختصاص دارند و TN تعداد افرادی هستندکه در کلاس مورد نظر دسته بندی نشده اند و به آن کلاس هم اختصاص ندارند.

$$\text{Accuracy}=\frac{TN+TP}{total\ number\ of\ items} \quad (۷)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad (۸)$$

$$\text{F-Value}=2*\frac{\text{Precision}*\text{recall}}{\text{Precision}+\text{recall}} \quad (۹)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad (۱۰)$$

جدول ۹ - نتایج شاخص های دقت و صحت برای خوشه‌بندی شورایی

ابعاد	خوشه ها	Accuracy	Precision	Recall	F-value
ادراک	خوشه ۱	0.97	0.92	1	0.95
	خوشه ۲	0.95	0.81	1	0.89
	خوشه ۳	0.93	0.92	0.85	0.88
	خوشه ۴	0.94	0.95	0.85	0.88
پردازش	خوشه ۱	0.95	0.81	1	0.89
	خوشه ۲	0.93	0.93	0.87	0.88
	خوشه ۳	0.97	0.88	0.91	0.89
	خوشه ۴	0.96	0.88	1	0.93
ورودی	خوشه ۱	0.96	0.85	1	0.91
	خوشه ۲	0.92	0.88	0.85	0.85
	خوشه ۳	0.94	0.96	0.87	0.91
	خوشه ۴	0.92	0.95	1	0.97
فهم	خوشه ۱	0.95	0.95	0.88	0.90
	خوشه ۲	0.90	0.85	0.82	0.83
	خوشه ۳	0.86	0.84	0.70	0.76
	خوشه ۴	0.91	0.63	1	0.77

۴-نتیجه‌گیری

سامانهٔ یادگیری الکترونیکی، بستری برای یادگیری شخصی است که از روش‌های مختلفی برای تطبیق محتوای آموزشی با تفاوت‌های فردی یادگیرندگان، استفاده می کند. در این سامانه‌ها باید یادگیرندگان با علایق مشابه در یک گروه قرار گیرند. استفاده از سبک یادگیری افراد، یکی از روش‌هایی است که می‌توان گروه‌بندی یادگیرندگان را بر این اساس انجام داد. لذا در این پژوهش گروه‌بندی افراد بر اساس سبک یادگیری حاصل از رفتارهای شبکه‌ای افراد و با روش شورایی صورت گرفته است و نتایج عددی بر اساس سه شاخص دیویس-بولدین و واریانس و خلوص و تجمع ارزیابی شده است. در سال های اخیر پژوهش‌های مشابهی در این حوزه صورت گرفته است که در جدول ۱۰ مشاهده می کنید:

جدول ۱۰ - مقالات مشابه در حوزه خوشه‌بندی شورایی

عنوان مقاله	سال انتشار	کاربرد
خوشه بندی شورایی: چارچوب استفاده مجدد دانش برای ترکیب افرازهای متعدد[۱۶]	۲۰۰۳	ترکیب افرازهای متعدد
افزایش یادگیری در لحظه از طریق یادگیری ماشین در سنسورهای زمینه میزکار [۲۴]	۲۰۰۷	افزایش یادگیری
پیش بینی ترک تحصیل در دوره های یادگیری الکترونیکی از طریق ترکیب تکنیکهای یادگیری ماشین [۲۵]	۲۰۰۹	پیش بینی ترک تحصیل در دوره های یادگیری الکترونیکی
ارزیابی خودکار سیستم های آموزشی هوشمند [۲۶]	۲۰۰۹	ارزیابی محتوا
تکنیک شورای افزایشی طبقه بندها برای پیش بینی عملکرد دانش آموزان در آموزش از راه دور [۲۷]	۲۰۱۰	پیش بینی کارایی دانش آموزان
طبقه بندی متن با استفاده از یک طبقه بند شورایی بر اساس ماتریس همبستگی [۲۸]	۲۰۱۲	طبقه بندی متن
رویکرد شورایی در همگرایی محتواهای سیستم های مدیریت یادگیری و مدیریت دانش [۲۹]	۲۰۱۶	همگرایی محتوای سیستم مدیریت دانش و سیستم مدیریت یادگیری

در پژوهش [۲۷]پژوهشگر از سه روش بیز ساده نزدیکترین همسایه و Weighted Majority برای دسته بندی دانشجویان آموزش از راه دور استفاده نموده است و در نهایت با روش خوشه بندی شورایی و از طریق تکنیک رای

اکثریت دانشجویان را در ۴ خوشه مختلف دسته بندی نموده است و دقتی مشابه جدول ۱۱ به دست آورده است:

جدول ۱۱-دقت خوشه بندی شورایی در پژوهش[۲۷]

Accuracy	Ensemble
WRI-1	73.86
WRI-2	78.39
WRI-3	81.73
WRI-4	81.81
Average	78.95

مطابق جدول ۱۱ میانگین دقت در پژوهش فوق ۷۸,۹۵ و بر اساس جدول ۹ در پژوهش جاری ۹۳,۵ درصد می باشد؛ که این موضوع بیانگر دقت بیشتر پژوهش جاری در انتخاب الگوریتمهای پایه ای مناسب تر می باشد. که این الگوریتم ها توانسته اند به خوبی معایب یکدیگر را پوشش داده و بهترین نتیجه را حاصل نمایند.

نتایج نشان می دهد روش خوشه‌بندی شورایی توانسته خوشه‌های دقیق‌تری را از نظر فشردگی درون خوشه ای و جدایی برون خوشه ای نسبت به سایر روش های متداول خوشه‌بندی ایجاد کند. ارزیابی واریانس خوشه‌ها در ابعاد مختلف سبک یادگیری نیز بیانگر آن‌است که خوشه‌بندی شورایی توانسته خوشه‌هایی با کمترین پراکندگی را ایجاد کند. طبق شاخص خلوص و تجمع نیز، روش گروه‌بندی پیشنهادی توانسته خوشه‌هایی با کیفیت بهتر ایجاد کند و کارایی مناسب‌تری نسبت به سایر روش‌های خوشه‌بندی در شناسایی ساختار گروه‌های یادگیرندگان دارد. همچنین نتایج خوشه‌بندی یادگیرندگان با نمایش سبک یادگیرندگان در هر گروه، نشان می‌دهد روش مذکور توانایی گردآوری یادگیرندگان با سبک مشابه در هر خوشه را داراست. روش گروه‌بندی پیشنهادی علاوه بر سادگی و راحتی پیاده‌سازی، دقت و صحت بیشتری نسبت به سایر روش‌های متداول داشته است؛ با استفاده از این روش می‌توان با ترکیب روش‌های مختلف خوشه‌بندی، نگرانی انتخاب بهترین روش برای گروه‌بندی یادگیرندگان را کاهش داد و ایرادها و معایب هر روش را با روش‌های دیگر پوشش داد تا بهترین گروه‌بندی یادگیرندگان حاصل شود.

سپاسگزاری

نویسندگان از خانم دکتر فاطمه قربانی، بـه دلیـل مشـاوره و نکات تکمیلی ایشان در تدوین مقاله، سپاس‌گزاری می‌کنند.

منابع

1. Samadi.M. (2011) Study of Felder and Solomon Psychometric Properties Questionnaire of Learning Styles. Journal of New Educational Approaches. Volume 6 , Number 1 (13);pp 39-60 [in Persian]

2. Romiszowski, A. J. (2004) "How's the e-learning baby? Factors leading to success or failure of an educational technology innovation," Educational Technology-Saddle Brook Then Englewood Cliffs NJ-, vol. 44, pp. 5-27.

3. Maseleno A., Sabani N., Huda M, Ahmad. R , Jasmi . K.A and Basiron. B. (2018). Demystifying Learning Analytics in Personalised Learning. International Journal of Engineering & Technology. Vol. 7, pp. 1124-1129.

4. Chen S. and Zhang, J. (2008). "The adaptive learning system based on learning style and cognitive state," in *Knowledge Acquisition and Modeling, 2008. KAM'08. International Symposium on*, pp. 302-306.

5. Schleith, D, Widener, M.J.,Kim C ,.Liu. L (2018). Assessing the delineated commuter sheds of various clustering methods. Journal of Computers, Environment and Urban Systems. Vol 71, pp 81-87.

6. Popat S. K. and Emmanuel, M (2014). "Review and Comparative Study of Clustering Techniques," *International Journal of Computer Science and Information Technologies*, vol. 5, pp. 805-812.

. Huang, Y.-P. Lu, C.-C. Chang, T.-W. and. Horng, M.-S (2007). "An intelligent approach to detecting the bad credit card accounts," in *Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications*, pp. 220-225.

8. Shakhnarovich, G, Indyk. P, and Darrell, T. (2006). Nearest-neighbor methods in learning and vision: theory and practice.

9. Gravenhorst, R. M. (2007). "Student learning styles and academic performance in a non-traditional anatomy course," Journal of Dance Education, vol. 7, pp. 38-46.

10. Honey P. and Mumford, A. (1989). Learning styles questionnaire: Organization Design and Development, Incorporated.

11. Kolb A. and Kolb, D. A. (2012) ."Kolb's learning styles," in Encyclopedia of the Sciences of Learning, ed: Springer, pp. 1698-1703.

12. Felder R. M. and Silverman, L. K. (1988). "Learning and teaching styles in engineering education," Engineering education, vol. 78, pp. 674-681.

13. Essalmi, F. . Ayed, L. J. B ,Jemni, M. and Graf, S (2010). "A fully personalization strategy of E-learning scenarios," Computers in Human Behavior, vol. 26, pp. 581-591.

۱۴. رضایی محمدصادق ,منتظر غلامعلی. گروه بندی همگن یادگیرندگان الکترونیکی بر اساس رفتار شبکه ای آنان. نشریه فناوری اطلاعات و ارتباطات ایران، سال ششم شماره ۲۱ (پاییز و زمستان ۱۳۹۳)

15. Truong, HM (2016). Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities .Journal of Computers in Human Behavior. Vol 55, pp 1185-1193.

16. Sabitha ,A.S, Mehrotra D. and Bansal. A. (2017). An ensemble approach in converging contents of LMS and KMS. Journal of Education and Information Technologies. Vol 22, pp 1673–1694.

17. Abellán J. and .Castellano, JG. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. Journal of Expert Systems with Applications. Vol 73 , pp 1-10.

18. Fred, A. (2001). "Finding consistent clusters in data partitions," in Multiple classifier systems, ed: Springer , pp. 309-318.

19. Vega-Pons S. and Ruiz-Shulcloper, J. (2011). "A survey of clustering ensemble algorithms," International Journal of Pattern Recognition *and Artificial Intelligence*, vol. 25, pp. 337-372.

20. Davies D. L. and Bouldin, D. W. (1979). "A cluster separation measure," Pattern Analysis and Machine Intelligence, IEEE Transactions on, pp. 224-227.

21. Zheng, Q. Ding, J. Du, J. and Tian, F. (2007). "Assessing method for e-learner clustering," in Computer Supported Cooperative Work in Design, CSCWD 2007. 11th International Conference on, 2007, pp. 979-983.

22. Halkidi, M. Batistakis, Y. and Vazirgiannis, M. (2001). "On clustering validation techniques," Journal of Intelligent Information Systems, vol. 17, pp. 107-145.

23. Olson, D. L., & Delen, D. (2008). Advanced data mining techniques p. 138 (1st ed.). Springer.

24. Lokaiczyk, R , Faatz,A, Beckhaus,A and Goertz,M.(2007). Enhancing Just-in-Time E-Learning Through Machine Learning on Desktop Context Sensors. International and Interdisciplinary Conference on Modeling and Using Context.pp 330-341.

25. Lykourentzou, I *. Giannoukos,I . Nikolopoulos,V . Mpardis,G . Loumos, V. (2009). Dropout prediction in e-learning

courses through the combination of machine learning techniques. Journal of Computers & Education Vol 53. pp 950–965.
26. He ,Y . Hui, SC . Quan,TT.(2009). Automatic summary assessment for intelligent tutoring systems. Journal of Computers & Education . Vol 53 . pp 890–899.
27. Kotsiantis.S . Patriarcheas. K , Xenos.M .(2010). A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. Journal of Knowledge-Based Systems. Vol 23 .pp 529–535.
28. Moreira-Matias.L . Mendes-Moreira.J . Gama.J and Brazdil.P.(2012). Text Categorization Using an Ensemble Classifier Based on a Mean Co-association Matrix. pp. 525–539.
29. Sai.A . Sabitha1 & Mehrotra1. D & Bansal.A. (2016). An ensemble approach in converging contents of LMS and KMS. Journal of Education and Information Technologies Vol 22(4) .

فصلنامه علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال یازدهم، شماره‌های ۳۹ و ۴۰، بهار و تابستان ۱۳۹۸
صص: ۱۹-۴۸

به‌کارگیری وب‌کاوی در پیش‌بینی جهت قیمت سهام گروه محصولات شیمیایی در بورس اوراق بهادار

[*]امیر دایی [**]امید مهدی عبادتی [***]کیوان برنا
[*] کارشناسی ارشد مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه خوارزمی، تهران
[**] استادیار گروه مدیریت عملیات و فناوری اطلاعات دانشگاه خوارزمی، تهران
[***] استادیار گروه علوم کامپیوتر دانشگاه خوارزمی، تهران
تاریخ دریافت: ۱۳۹۸/۱۰/۰۹ تاریخ پذیرش: ۱۳۹۹/۰۴/۱۸

چکیده

پیش‌بینی بازارها از جمله سهام به دلیل حجم بالای معاملات و نقدینگی برای محققان و سرمایه‌گذاران دارای جذابیت بوده است. توانایی پیش‌بینی جهت قیمت ما را قادر می‌سازد با کاهش ریسک و اجتناب از ضرر و زیان مالی، به بازده بالاتری دست‌یابیم. اخبار نقش مهمی در فرایند ارزیابی قیمت فعلی سهام دارد. توسعه روش‌های داده‌کاوی، هوش محاسباتی و الگوریتم‌های یادگیری ماشین سبب ایجاد مدل‌های جدیدی در پیش‌بینی شده‌اند. هدف از این پژوهش ذخیره سازی اخبار خبرگزارها و استفاده از روش‌های متن کاوی و الگوریتم ماشین بردار پشیبان به منظور پیش‌بینی جهت قیمت روز آینده سهم است. بدین منظور خبرها منتشر شده در ۱۷ خبرگزاری با استفاده از یک خزگشر موضوعی به زبان پی‌اچپی[1] ذخیره و دسته‌بندی شده است. سپس با استفاده از روش‌های متن‌کاوی و الگوریتم ماشین بردار پشتیبان و کرنل‌های مختلف به پیش‌بینی جهت قیمت سهام گروه محصولات شیمیایی در بورس اوراق بهادار پرداخته می‌شود. دراین مطالعه از ۳۰۰ هزار خبر در دسته‌های سیاسی و اقتصادی و قیمت‌های سهام ۲۵ شرکت منتخب در بازه زمانی آبان تا اسفند ۹۷ در ۱۲۲ روز معاملاتی استفاده شده است. نتایج نشان می‌دهد با مدل ماشین بردار پشتیبان با کرنل خطی می‌توان به صورت میانگین ۸۳ درصد جهت قیمت‌ها را پیش‌بینی کرد. با استفاده از کرنل‌های غیرخطی و معادله درجه ۲ ماشین بردار پشتیبان صحت پیش‌بینی به صورت میانگین تا ۸۵ درصد افزایش می‌یابد و سایر کرنل‌ها نتایج ضعیف‌تری از خود نشان می‌دهند.

کلید واژه: متن‌کاوی، کاوش محتوای وب، خزشگر وب، پیش‌بینی بورس اوراق بهادار، ماشین بردار پشتیبان

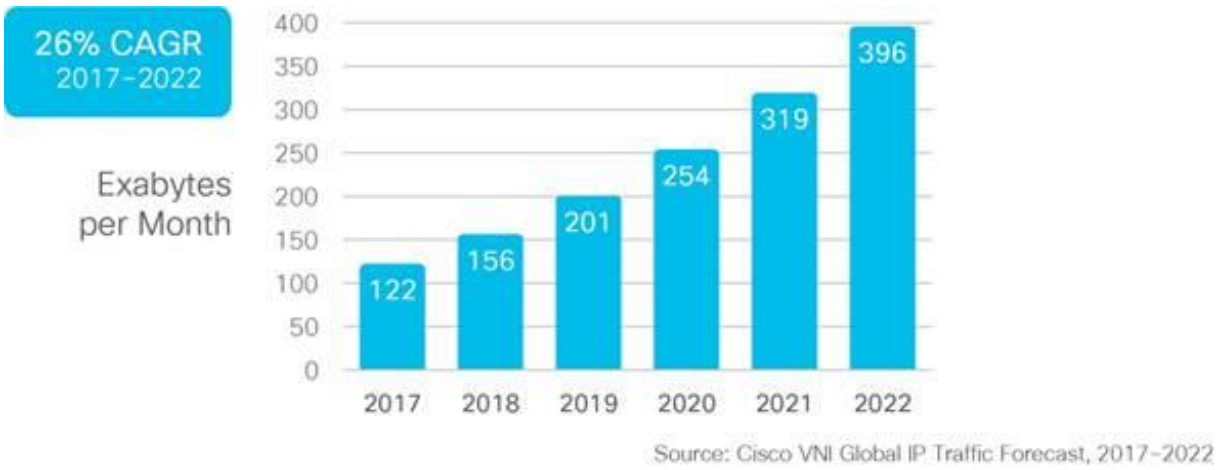
۱-مقدمه

با رشد روز افزون اینترنت و تولید محتوا در محیط وب، داده بسیار زیادی در بانک‌های اطلاعاتی ذخیره می‌شوند. این داده‌ها ممکن است از طریق صفحات وب به نمایش درآیند. اگرچه دسترسی به پایگاه داده وب سایت‌ها و گرفتن داده‌ها از مدیران سایت‌ها کار ساده‌ای نیست با این حال می‌توان با خزش وب سایت‌ها، این داده‌های ارزشمند را ذخیره کرد. این موضوع فرصتی را فراهم می‌آورد تا بتوان به مباحث داده‌کاوی و متن‌کاوی و استخراج دانش پرداخت. قبل از ظهور وب اکثر کاربران مجموعه عظیمی از اسناد نداشتند و بنابراین نیازی به سامانه‌ای پیچیده برای جستجو میان اسناد مختلف نبود. با ظهور وب، کاربران به داده زیاد در حال رشدی دسترسی پیدا کردند. مرتب‌سازی و جستجو در میان انبوهی از داده‌های وب مسائل جدیدی را ایجاد کرد که با عنوان بازیابی اطلاعات وب[1] شناخته می‌شود [1]. بر اساس شاخص cisco VNI پیش‌بینی می‌شود ترافیک آی پی جهانی از سال ۲۰۱۷ تا ۲۰۲۲ سه برابر شود. پیش‌بینی می‌شود ترافیک آی پی در سال ۲۰۲۲ به صورت ماهانه به ۳۹۶EB برسد.

ترافیک کل برای اینترنت در دو دهه اخیر رشد فوق‌العاده‌ای را تجربه کرده است. بیش از ۲۰ سال قبل، در سال ۱۹۹۲، شبکه‌های اینترنت جهانی، حدود ۱۰۰ گیگ ترافیک را در هر روز منتقل می‌کردند. در سال ۲۰۱۷، ترافیک جهانی اینترنت به ۴۵۰۰۰ گیگ در هر ثانیه رسید. [۲].

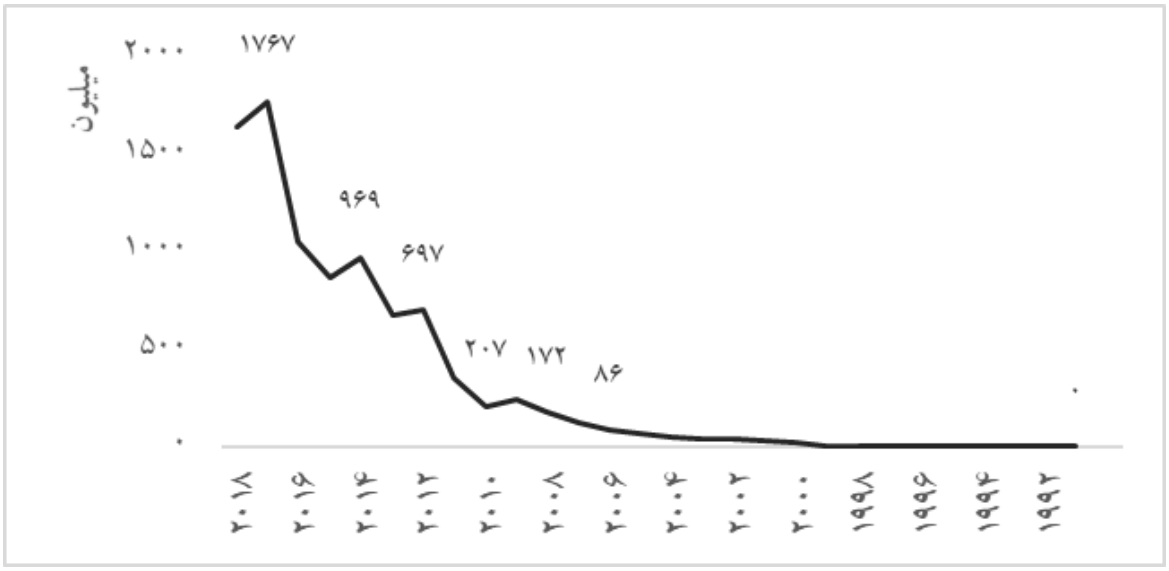
نویسنده عهده‌دار مکاتبات: امیدمهدی عبادتی ebadati@khu.ac.ir

[1] PHP



شکل ۱-پیش‌بینی شرکت سیسکو از ترافیک آی‌پی [۲]

در حال حاضر بیش از یک و نیم میلیارد وب‌سایت در شبکه جهانی اینترنت وجود دارد. کمتر از ۲۰۰ میلیون از این وب‌سایت‌ها فعال هستند. این موضوع در تحقیقات web server اکتبر ۲۰۱۴ netcraft تائید شد و برای اولین بار توسط Internet live stats برآورد و اعلام شد. شکل ۲ تعداد کل وب‌سایت‌ها را در طول تاریخ از ۱ وب‌سایت در سال ۱۹۹۱ تا سال ۲۰۱۸ نمایش می‌دهد [۳]



شکل ۲- تعداد وب‌سایت‌ها از سال ۱۹۹۲ تا ۲۰۱۸

رشد اینترنت و گسترش وب‌سایت‌ها و تولید محتواهای غیر ساختاریافته مطالعه در این حوزه را روز به روز با اهمیت‌تر می‌کند.

وب کاوی در حوزه‌های مختلف مورد مطالعه قرارگرفته است: فن‌های آماده‌سازی و پالایش داده‌های صفحات وب[۲]، استخراج و ذخیره‌سازی الگوی وب[۳]، کاوش ساختار وب[۴]، کاوش استفاده‌ی از وب[۵] و کاوش محتوای وب[۶] که هدف آن کشف دانش مفید از داده‌های موجود در اینترنت است [۱]. وب کاوی بسته به اهداف کاوش و داده‌های ذخیره‌شده به سه دسته تقسیم می‌شود: استخراج ساختار وب، استخراج کاربرد وب و استخراج محتوای وب [۴]. رشد روز افزون وب سایت‌ها و مطالب منتشرشده در محیط وب اهمیت استفاده از کاورش محتوای وب را دو چندان می‌کند.

خزش اولین قدم در وب کاوی یا ساخت یک موتور جستجو است. به صورت کلی خزشگرها به دو دسته تقسیم می‌شوند: خزشگر عمومی[۷] و خزشگر موضوعی[۸]. خزشگرهای عمومی

[2] Web Data Pre-processing and Cleaning
[3] Web Pattern Extraction and Storage
[4] Web structure mining
[5] Web usage mining
[6] Web content mining
[7] Universal crawler

همه صفحات را بدون در نظر گرفتن محتوای آنها دانلود می‌کند، اما در خزشگرهای موضوعی فقط صفحاتی با موضوعات خاص دانلود می‌شود [۵].

به‌طور کلی ابزارهای پیش‌بینی در بازار سرمایه به دو بخش تحلیل تکنیکال[۹] و بنیادی[۱۰] تقسیم می‌شوند. تفاوت این روش‌ها در داده‌های ورودی است؛ در روش تکنیکال از داده‌های تاریخی بازار استفاده می‌شود در صورتی که در تحلیل بنیادی از دیگر انواع اطلاعات یا اخبار درباره کشور، جامعه، شرکت و غیره استفاده می‌شود. بیشتر تحقیقات در گذشته بر روی رهیافت تکنیکال انجام‌شده است، که دلیل عمده آن در دسترس بودن داده‌های کمّی تاریخی در بازار و تمایل عمومی معامله‌گران برای استفاده از روش‌های کمّی تکنیکال است. داده‌های بنیادی در صورتی که بدون ساختار باشد دارای چالش‌های بیشتری برای استفاده به عنوان ورودی هستند. داده‌های بنیادی ممکن است از منابع ساختاریافته و عدد مانند داده‌های اقتصاد کلان یا گزارش صورت‌های مالی بانکی و دولتی گرفته شوند. [۶]

توسعه روش‌های داده‌کاوی، هوش محاسباتی و الگوریتم‌های یادگیری ماشین سبب ایجاد مدل‌های جدیدی در پیش‌بینی شده‌اند. به صورت کلی در مقالات متعدد نشریات مختلف دو دسته تحقیق در استفاده از روش‌های داده‌کاوی برای پیش‌بینی بازارهای مالی صورت گرفته است. روش اول استفاده از داده‌های دارای ساختار است که در بیشتر مطالعات از داده‌های دارای ساختار مانند قیمت گذشته، درآمد و سود تاریخی استفاده کرده‌اند. [۷]

همچنین در مقاله‌ای دیگر که توسط چن جانگ هانگ[۱۱] و همکاران منتشرشده است تمرکز سیستم‌های پیش‌بینی مالی را صرفاً بر روی داده‌های کمی مانند قیمت سهام و شاخص بازار می‌دانند. [۸]

اما بخش عمده‌ای از داده‌هایی که در محیط وب تولید می‌شوند بدون ساختار هستند و از این رو می‌توان از روش دوم یعنی استفاده از تکنیک‌های متن‌کاوی و داده‌کاوی بر روی داده‌های بدون ساختار استفاده کرد. متن‌کاوی شامل وظایف بسیاری مانند خوشه‌بندی اسناد، دسته‌بندی اسناد، خلاصه‌سازی متن، تحلیل احساسات، تحلیل شبکه‌های اجتماعی، تشخیص موضوع، دسته‌بندی صفحات وب، شناسایی نویسنده، تشخیص سرقت ادبی، تحلیل فیشینگ/هرزنامه و نرم‌افزارهای مخرب، تحلیل الگو، تصمیم‌گیری مالی و غیره است؛ اما اصلی‌ترین چالش در متن‌کاوی، داده‌های بدون ساختار است که قبل از شروع داده‌کاوی نیاز است آن‌ها را به فرمت ساختاریافته تبدیل کرد. [۹]

اخبار[۱۲] نقش مهمی در فرایند ارزیابی قیمت فعلی سهام، که توسط تحلیل‌گران، سرمایه‌گذاران و سرمایه‌گذاران نهادی[۱۳] صورت می‌گیرد، دارد. بر اساس یک دیدگاه تئوریک، ارزیابی کارآمد از یک شرکت باید بر ارزش فعلی و جریان وجوه نقد آینده شرکت تأثیر بگذارد. در اخبار نه تنها آمار و ارقام مالی بلکه اجزای متنی کیفی نیز قیمت سهام را تحت تأثیر قرار می‌دهند. [۱۰]

در این پژوهش فرض شده است که با استفاده از اخبار منتشر شده می‌توان جهت قیمت سهم را در روز بعد پیش‌بینی کرد. بدین منظور اخبار منتشر شده در خبرگزاری‌ها به عنوان متغییر تحقیق در نظر گرفته شده‌اند و دیتای اخبار ذخیره می‌شود. سپس با استفاده از این اخبار و روش‌های متن‌کاوی و داده‌کاوی و استفاده از الگوریتم ماشین بردار پشتیبان با کرنل‌های مختلف به دسته‌بندی اخبار برای پیش‌بینی صعودی یا نزولی بودن قیمت در روز بعد پرداخته می‌شود.

۲- ادبیات موضوع

برای مطالعات وب کاوی می‌توان از داده‌هایی که توسط خزشگرها و موتورهای جستجو ذخیره‌شده است یا با نوشتن خزشگر و ذخیره‌سازی دیتا دسترسی داشت. اگر بخواهیم از داده‌های خزشگرها و موتورهای جستجو استفاده کنیم این دیتا در مدت زمان کوتاه و به راحتی ممکن است در اختیار قرار گیرد. بدین منظور می‌توانید از مجموعه داده‌های آماده استفاده کنیم، اما این دیتا بیشتر برای آزمایش یک الگوریتم مفید است. شاید در این زمینه بهترین مجموعه داده ترک (trec.org) باشد که شامل مجموعه از صفحات وب با ساختار html است [۱۱]. ممکن است داده‌ها به روز نباشد، داده‌هایی مورد نیاز ذخیره نشده باشد، دیتا ساختار مناسب نداشته باشد، دیتا بیش از نیازها باشد یا متحمل هزینه شود. برای دستیابی به اطلاعات دقیق، به روز و مورد اعتماد نیاز است که از خزشگر مورد نیاز نوشته شود [۱۱]. اما این روش چالش‌هایی مانند نیاز به دانش برنامه‌نویسی، زمان‌بر بودن نوشتن و آزمایش خزشگر و در نهایت فرایند ذخیره‌سازی داده دارد. در

[8] Topic crawlers
[9] Technical
[10] Fundamental
[11] Chenn-Jung Huang
[12] News
[13] Institutional Traders

این پژوهش ابزار قابل توسعه‌ای طراحی و ارائه‌شده است تا بتوان محتوای مورد نیاز را از وب ذخیره کرد.

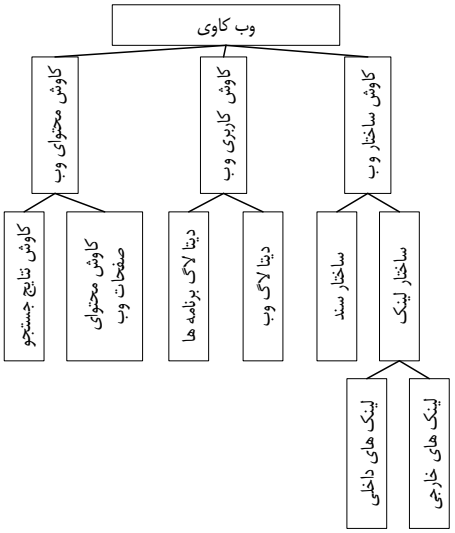
۲-۱-وب‌کاوی

در طی سال‌ها شبکه وب از طریق انتقال سنتی و به اشتراک گذاشتن رایانه‌ها و اسناد به عنوان «وب داده»، به اتصال فعلی مردم به عنوان «وب افراد» و به اتصال در حال ظهور میلیاردها اشیاء به عنوان «وب اشیاء» تغییر کرده است. [۱۲]

وب کاوی به معنی استفاده از فن داده‌کاوی به‌منظور خودکار کردن جستجو و استخراج اطلاعات از اسناد و خدمات وب است. با توجه به حجم عظیمی اطلاعاتی که در محیط اینترنت، شبکه جهانی وب حوزه بکری برای مطالعات داده‌کاوی است. این داده‌های عظیم باعث ایجاد مسائلی از قبیل پیدا کردن اطلاعات مرتبط، ایجاد دانش جدید از اطلاعات در محیط وب، شخصی‌سازی اطلاعات و شناخت رفتار مصرف کنند و کاربران شده است [۱۳]

۲-۲-روش‌های وب کاوی

وب کاوی بر اساس نوع دیتا به سه دسته کاوش کاربری وب، کاوش ساختار وب و کاوش محتوای وب تقسیم می‌شود. هر سه دیدگاه بر روی کشف دانش ضمنی، اطلاعات ناشناخته و بالقوه تمرکز دارند. [۱۴]



شکل ۳-روش‌های وب کاوی [۱۵]

۲-۳-انواع روش‌های متن‌کاوی

الگوریتم‌های متن‌کاوی به‌طور کلی به دو دسته الگوریتم یادگیری با نظارت و الگوریتم یادگیری بدون نظارت تقسیم می‌شوند. برای متن‌کاوی از ابزارها و روش‌های مختلفی استفاده می‌شود که به سه گروه اصلی تقسیم می‌شوند. این ابزارها و روش‌ها در زیر بیان‌شده است [۱۶]:

۲-۳-۱-کلاس‌بندی یا دسته‌بندی[14]

کلاس‌بندی یا دسته‌بندی یکی از روش‌هایی است که برای گروه‌بندی متن‌ها در متن‌کاوی مورد استفاده قرار می‌گیرد . هدف از کلاس‌بندی متون، نسبت دادن کلاس‌های از پیش تعریف‌شده به اسناد متنی موجود، مانند گروه اخبار، گروه کالاها، جناح سیاسی است. [۱۷, ۱۸].

۲-۳-۲-خوشه‌بندی[15]

خوشه‌بندی یکی از روش‌های متن‌کاوی است که برای شناسایی گروه‌های داده‌ها بر اساس مشخصه‌ها یا ویژگی‌های آن‌ها به کار برده می‌شود که در آن هیچ گروه از پیش تعیین شده‌ای وجود ندارد [۱۷].

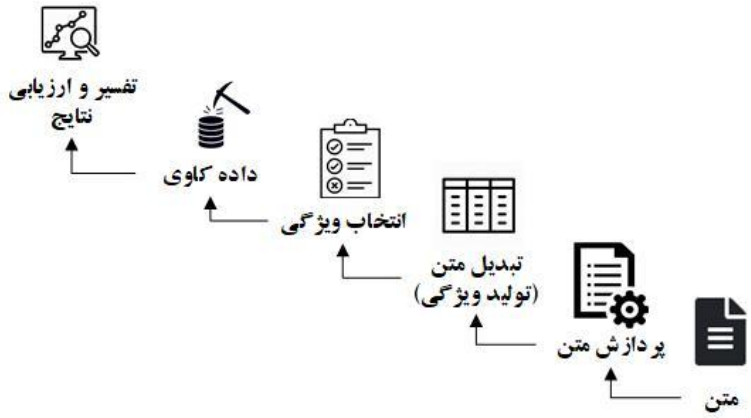
۲-۳-۳- کاوش قوانین وابستگی

کاوش قوانین وابستگی روشی است که با کشف و یافتن قوانینی در رخدادهای دیگر وقوع یک مورد را در آینده پیش‌بینی می‌کند. کاوش قوانین وابستگی در دسته الگوریتم‌های یادگیری بدون نظارت جای دارد زیرا هیچ برچسبی از قبل برای آموزش الگوریتم وجود ندارد [۱۶, ۱۹].

۲-۴-مراحل متن‌کاوی

فرایند متن‌کاوی مراحل مختلفی دارد که در پژوهش هاشیمی[16] و همکاران (۲۰۱۵)، پیش‌پردازش، استخراج ویژگی و انتخاب ویژگی از مراحل اصلی این فرایند عنوان‌شده است. مراحل کلی این فرایند در شکل ۴ نشان داده‌شده است. [۲۰] کومر و راوی[17] به‌طور کلی فرایند متن‌کاوی شامل دو فاز اساسی می‌باشد: پیش‌پردازش متن، استخراج دانش [۱۶].

[14] Classification
[15] Clustering
[16] Hashimi
[17] Kumar & Ravi



شکل ۴-مراحل متن‌کاوی

۴-۲-۱-پیش‌پردازش

برای کشف دانش از حجم قابل توجهی از اسناد، لازم است که بر روی اسناد پیش‌پردازش انجام شود؛ یعنی اطلاعات در یک ساختار داده‌ای مناسب برای پردازش‌های بعدی آماده و ذخیره‌سازی شود. در این مرحله داده‌های ورودی که در دسترس است باید برای ورود به الگوریتم یادگیری ماشین آماده شود، یعنی از حالت غیرساخت‌یافته به فرمت ساختاریافته و قابل تشخیص برای ماشین تبدیل شود. [۱۸]

۴-۲-۲-تولید و استخراج ویژگی

اگرچه برنامه‌های کاربردی زیادی درزمینه بازیابی اطلاعات مانند پالایش و جست‌وجوی اطلاعات مرتبط می‌توانند از تحقیقات درزمینه رده‌بندی متن سود ببرند، مشکل اصلی رده‌بندی متن، ابعاد بالای فضای ویژگی با توجه به تعداد زیاد لغات است. راه‌حل این مشکل استفاده از روش‌های استخراج و انتخاب ویژگی است [۲۰].

۴-۲-۳-انتخاب ویژگی

انتخاب ویژگی فرایندی است که زیرمجموعه‌ای از ویژگی‌های اصلی را با توجه به برخی از معیارها و یا اهمیت ویژگی‌ها انتخاب می‌کند. الگوریتم‌های انتخاب ویژگی به دو دسته زیر تقسیم می‌شوند:

۱- الگوریتم‌های رتبه‌بندی ویژگی

۲- الگوریتم‌های انتخاب زیر مجموعه ویژگی

۴-۲-۴-کشف دانش

گام بعدی استخراج و کشف دانش از فرم‌های میانی است که بر اساس نحوه نمایش هر سند می‌تواند متفاوت باشد. نمایش مبتنی بر سند برای گروه‌بندی، طبقه‌بندی و تجسم‌سازی استفاده می‌شود، در حالی‌که نمایش مبتنی بر مفهوم برای یافتن روابط میان مفاهیم، ساخت اتوماتیک آنتولوژی و غیره به‌کار می‌رود [۲۱].

۴-۲-۵-تفسیر و ارزیابی

در این مرحله، خروجی مراحل قبل مورد ارزیابی قرار می‌گیرد تا مشخص شود که دانشی کشف‌شده است و اینکه دانش کشف‌شده اهمیت دارد یا خیر. با اجرای الگوریتم‌ها، داده/ متن استخراج‌شده به فنون مختلفی تحویل داده می‌شود که امکان استفاده مستقیم از اطلاعات استخراج‌شده را از طریق ابزار کشف پیوند یا مصورسازی فراهم می‌کنند [۲۱].

۲-۵-یادگیری ماشین

یادگیری ماشین ارتباط نزدیکی با آمار محاسباتی دارد و اغلب با آن همپوشانی دارد. تمرکز این شاخه پیش‌بینی کردن با رایانه است و پیوند محکمی با بهینه‌سازی ریاضی دارد. مدل‌های تحلیلی به محققان، پژوهشگران علم داده و تحلیلگران اجازه می‌دهد تصمیمات و نتایج قابل اطمینان و تکرارپذیر به دست آورند و با یادگیری از روابط و روندهای مربوط به گذشته، از الگوهای پنهان پرده‌برداری کنند [۲۲].

۱-۲-۵-یادگیری بدون نظارت: خوشه‌بندی

الگوریتم خوشه‌بندی کا-مینز[18] و کا-مد[19]

الگوریتم کا مینز که در سال ۱۹۶۷ توسط مک کویین[20] مطرح شد، یکی از محبوب‌ترین الگوریتم‌های خوشه‌بندی است که در زمینه‌های مختلف مورد استفاده قرار می‌گیرد. هدف الگوریتم کا-مینز، بهینه‌سازی تابع هدف[21] می‌باشد که پاسخ‌های حاصل از خوشه‌بندی به کمینه‌سازی یا بیشینه‌سازی تابع هدف منجر می‌شود. در این الگوریتم باید

[18] K-means
[19] K-mode
[20] Mac Queen
[21] Object Function

تعداد خوشه‌ها از قبل مشخص‌شده باشد. این الگوریتم بر روی داده‌های پیوسته تعریف می‌شود. [۲۳].

۱-۱-۲-۵-الگوریتم DBSCAN

یکی از پرکاربردترین الگوریتم‌های خوشه‌بندی، تکنیک DBSCAN است که خوشه‌ها را بر اساس تراکم و غلظت آن‌ها تعیین کند. در این الگوریتم نیازی به مشخص کردن تعداد خوشه‌ها توسط کاربر نیست. این الگوریتم قادر است با اثربخشی بالایی خوشه‌هایی به شکل‌های دلخواه ایجاد کند. برای خوشه‌بندی تعدادی از نقاط توسط این الگوریتم، از دو پارامتر شعاع همسایگی[۲۲] و حداقل تعداد نقاط موجود[۲۳] در همسایگی، استفاده می‌شود و مجموعه نقاط را به نقاط مرکزی[۲۴]، نقاط مرزی[۲۵] و داده‌های پرت[۲۶] تقسیم می‌کند. [۲۴].

الگوریتم سلسه‌مراتبی[۲۷]

این الگوریتم به دو دسته از بالا به پایین[۲۸] و از پایین به بالا[۲۹] تقسیم می‌شود. در روش از بالا به پایین، تمامی داده‌ها با هم به‌عنوان یک خوشه بزرگ در نظر گرفته می‌شوند و در مرحله بعد به خوشه‌های کوچک‌تر تقسیم‌شده تا جایی که هر داده به‌عنوان یک خوشه در نظر گرفته شود. اما در روش پایین به بالا هر داده در ابتدا یک خوشه است، به ترتیب در هر مرحله داده‌هایی که بیشترین شباهت (کمترین فاصله) را به هم دارند به خوشه می‌پیوندند تا خوشه بزرگ‌تر شود و در نهایت همه داده‌ها با هم یک خوشه شوند. برای اندازه‌گیری فاصله بین دو خوشه از روش‌های پیوند[۳۰] استفاده می‌شود. انواع روش‌های پیوند عبارت‌اند از: نزدیک‌ترین همسایه[۳۱]، پیوند تکی[۳۲]، دورترین همسایه[۳۳]، پیوند کامل[۳۴] و یا پیوند میانگین[۳۵] [۲۵].

مدل ترکیبی گاوسی[۳۶]

منظور از نمودار گوسی این است که یک مقدار مشخص وجود دارد که حداکثر نمونه‌ها در آن قرار دارند و هرچه از این مقدار دورتر می‌شویم تعداد داده‌ها کمتر می‌شود. به این نوع پخش‌شدگی توزیع گوسی می‌گویند که بیشتر داده‌های جهان از این نوع توزیع پیروی می‌کنند. شکل توزیع گوسی همیشه متوازن نیست و ممکن است به سمت چپ یا راست چولگی[۳۷] داشته باشد. این الگوریتم از روش بیشینه‌سازی انتظار[۳۸] استفاده می‌کند.

۲-۲-۵-الگوریتم یادگیری بانظارت: کلاس‌بندی

الگوریتم درخت تصمیم[۳۹]

درخت تصمیم راه‌حلی سریع و مفید برای کلاس‌بندی مجموعه داده‌های بزرگ با تعداد زیادی از متغیرها را فراهم می‌کند. این الگوریتم که متغیرهای کمی و کیفی را پیش‌بینی می‌کند، اولین بار توسط برمان[۴۰] مطرح شد. نتیجه این الگوریتم مجموعه‌ای از شرط‌های منطقی با ساختار درختی است که برای پیش‌بینی یک ویژگی به‌کار می‌رود. الگوریتم درخت تصمیم به‌گونه‌ای عمل می‌کند که گوناگونی یا تنوع در گره‌ها را به حداقل برساند. ۴ نوع الگوریتم درخت تصمیم CART، QUEST، CHAID و C۰.۵ وجود دارد که تفاوت آن‌ها در معیار اندازه‌گیری عدم خلوص، شیوه شاخه‌بندی و هرس‌کردن گره‌های درخت می‌باشد [۲۶].

الگوریتم جنگل تصادفی[۴۱]

الگوریتم جنگل تصادفی یکی از الگوریتم‌های یادگیری ماشین و ابزاری برای کلاس‌بندی و رگرسیون است [۲۷]. جنگل تصادفی الگوریتمی ترکیبی است که بر اساس مدل درخت تصمیم شکل‌گرفته است. در این الگوریتم ابتدا با استفاده از روش نمونه‌گیری بوت استرپ[۴۲] تعداد K زیرمجموعه آموزش[۴۳] را از بین مجموعه داده‌های اصلی استخراج می‌کند و سپس با آزمایش کردن این زیرمجموعه‌ها تعداد K درخت تصمیم ایجاد می‌شود. در نهایت یک جنگل تصادفی از این درختان تصمیم ایجاد می‌شود. [۲۸].

الگوریتم شبکه عصبی[۴۴]

شبکه عصبی ابزار قدرتمندی برای حل مشکلات پیچیده است که با پردازش داده‌ها، دانش پنهان آن‌ها را به ساختار شبکه

[22] Epsilon
[23] MinPoints
[24] Core points
[25] Border point
[26] Outlier
[27] Hierarchical
[28] Partitioning
[29] Agglomerative
[30] Linkage
[31] Nearest Neighbor
[32] Single Linkage
[33] Furthest Neighbor
[34] Complete Linkage
[35] Average Linkage
[36] Gaussian Mixture Model (GMM)
[37] Skew
[38] Expectation Maximization (EM)
[39] Decision Tree
[40] Breman
[41] Random Forest
[42] Bootstrap
[43] Train
[44] Neural Network

منتقل می‌کند. نورون یک تابع غیرخطی، پارامتری و محدود است که برای راحتی، به این تابع نورون گفته می‌شود. به متغیرهای نورون ورودی نورون گفته می‌شود و مقدار آن در خروجی آن نشان داده می‌شود. نورون‌ها می‌توانند به‌راحتی به‌صورت گرافیکی نمایش داده شوند. ترکیب توابع غیرخطی دو یا چند نورون، شبکه‌ای از نورون‌ها را تشکیل می‌دهد. به‌طور کلی شبکه‌های عصبی دو نوع هستند: شبکه پیش‌خور[45] و شبکه بازخور[46].[۲۹].

ماشین بردار پشتیبان[47]

مسأله طبقه بندی یکی از مسائل اصلی مطر شده در یادگیری ماشین است. بسیاری از مسائل را می توان بصورت یک مسأله کلاسه بندی مطرح کرده و حل نمود. روش های طبقه بندی خطی، سعی دارند که با ساختن یک ابر سطح (که عبارت است از یک معادله خط)، داده ها را از هم تفکیک کنند. چندین تفکیک کننده خطی می تواند داده دو کلاس را از هم جدا کند. یکی از روش هایی که بصورت گسترده برای اینگونه مسائل استفاده می شود، ماشین بردار پشتیبان است. ماشین بردار پشتیبان در سال ۱۹۹۳ توسط ولادیمیر واپنیک پیشنهاد شد. SVM بهترین ابر سطحی را پیدا می کند که با حداکثر فاصله، داده های مربوط به دو طبقه را از هم تفکیک کند. [۳۰]

۲-۶-ارزیابی مدل در یادگیری با ناظر

در یادگیری با ناظر چندین راه برای ارزیابی عملکرد مدل یادگیری و نتایج دسته‌بندی وجود دارد. شاخص‌های ارزیابی کیفیت کلاس‌بندی بر اساس ماتریس درهم ریختگی تعیین می‌شوند. در ماتریس درهم ریختگی تعیین می‌شود چه تعداد از پیش‌بینی‌های مدل صحیح و غلط بوده‌اند. جدول ۱ ماتریس درهم ریختگی برای یک مدل دوتایی را نمایش می‌دهد.

جدول ۱- ماتریس در هم ریختگی

کلاس / تشخیص	تشخیص مثبت	تشخیص منفی
مثبت	Tp	Fn
منفی	Fp	Tn

Tp (true positive): کلاس مثبتی که به درستی مثبت تشخیص داده‌شده است.

Fn (false negative): کلاس مثبتی که به اشتباه منفی تشخیص داده‌شده است.

Fp (false positive): کلاس منفی که به اشتباه مثبت تشخیص داده‌شده است.

Tn (true negative): کلاس منفی که به درستی مثبت تشخیص داده‌شده است.

بر اساس این ماتریس شاخص‌های دقت، صحت و امتیاز-F محاسبه می‌شود.

۱-۲-۶-دقت[48]

precision درستی پیش‌بینی‌ها را نسبت به کل موارد کلاس نشان می‌دهد. فرمول محاسبه به صورت زیر است:

$$\text{precision} = \frac{Tp}{Tp + Fp}$$

۲-۲-۶-صحت[49]

در شاخص صحت نسبت مقدار موارد صحیح طبقه‌بندی‌شده توسط الگوریتم از یک کلاس به کل اعضای پیش‌بینی شده در آن گروه را محاسبه می‌کنیم. فرمول محاسبه به صورت زیر است:

$$\text{Recall} = \frac{Tp}{Tp + Fn}$$

۳-۲-۶-امتیاز-F1[50]

این معیار دقت و صحت را با هم در نظر می‌گیرد. معیار F1 در بهترین حالت، یک و در بدترین حالت صفر است. این معیار توصیف‌کننده میانگین وزن‌دار مابین دو کمیت دقت و صحت است. این معیار میزان دقت و صحت یک مدل را به صورت هم‌زمان بررسی می‌کند و میزان کیفیت کلاس‌بندی را تعیین می‌کند. برای محاسبه امتیاز F1 از فرمول زیر استفاده می‌شود. برای محاسبه این شاخص از فرمول زیر استفاده می‌شود:

$$\text{F1 Score} = \frac{\text{precision} * \text{Recall}}{\text{precision} + \text{Recall}}$$

با استفاده از این شاخص‌های می‌توان مدل‌های یادگیری دوتایی با ناظر را ارزیابی کرد. [۳۱]

۷-۲-پیش‌بینی بازار بر اساس اخبار منتشرشده

مدیریت ریسک مالی یکی از کارهای بسیار پر چالش در واحدهای مالی است. در دو دهه گذشته روش‌ها و مدل‌های کمی[51] بسیاری جهت مشخص کردن تأثیر بازارهای بی‌ثبات مالی بر تجارت، توسعه و گسترش یافته‌اند. اکثر این مطالعات

[45] Feedforward
[46] Feedback
[47] Support vector machine
[48] Precision
[49] Recall
[50] F1 Score
[51] Quantitative

بر روی داده‌های ساختاریافته مانند سری قیمت‌های تاریخی انجام‌شده است و توجه اندکی به داده‌های غیر ساختار یافته (متنی[۵۲]) شده است، در حالی که بیشتر حجم منابع اطلاعاتی موجود از این نوع محتوا هستند. تحقیقات تجربی که در گذشته صورت گرفته است نشان می‌دهد که اخبار خاص، مانند افشای شرکت‌های بزرگ می‌توانند سبب تغییر رفتار غیرطبیعی قیمت پس از انتشار شود. [۳۲]

فرضیه بازار کارا و نظریه گام تصادفی دو نظریه هستند که تأثیر قابل توجهی بر پیش‌بینی بازار داشته است. در فرضیه بازار کارا، قیمت سهام بازتابی از اطلاعات بازار کامل است و زمانی که اطلاعات جدیدی ارائه شود، بلافاصله در قیمت سهام منعکس می‌شود. در نظریه گام تصادفی اعتقاد بر این است که بازارها کارآمد هستند و اصلاح قیمت‌ها بلافاصله رخ می‌دهد و پیش‌بینی قیمت از داده بازار غیرممکن است. این نظریه‌ها بیان می‌کنند که قیمت‌ها به اطلاعات پیوند خورده‌اند و از اطلاعات امروز نمی‌توان برای پیش‌بینی قیمت‌ها در آینده استفاده کرد. [۳۳]

همچنین در پیش‌بینی قیمت سهام دو فلسفه کاملاً مخالف وجود دارد. فن‌های تجزیه و تحلیل بنیادی و تکنیکال. در حالی که تحلیلگران بنیادی به دنبال استفاده از داده‌های نسبی، نسبت‌ها و درآمد نسبی سهام هستند، تحلیلگران تکنیکال از نمودارها، فن‌های مدل‌سازی بر اساس حجم معاملات تاریخی و قیمت‌ها برای تحلیل خود استفاده می‌کنند. [۳۳]

در حالی که رفتار بازار تحت تأثیر اتفاقات محلی و جهانی است، جزئیات این اتفاقات در داده‌های ساختاریافته مشاهده نمی‌شود. انتظار می‌رود که داده‌کاوی نقش مهمی در طراحی راهبردهای پیش‌بینی رفتار بازار داشته باشد، زیرا می‌تواند با موفقیت برای تجزیه و تحلیل مقالات اخبار مالی و گزارش‌های همراه داده‌های سری زمانی بازار مورد استفاده قرار گیرد. متن‌کاوی گزارش‌های خبری مالی می‌تواند برای استخراج اطلاعات مهم در مورد وقایع مختلف سیاسی و اقتصادی که به طور کلی بازار مالی یک منطقه را تعیین می‌کند و نیز علل عملکرد ضعیف یا افزایش ناگهانی در بازار را توضیح دهد. امروزه با توجه به حجم اخبار موجود در اینترنت، نیاز روزافزونی به فناوری قابل اطمینانی وجود دارد که بتواند برای تجزیه و تحلیل خودکار گزارش‌های خبری و کشف اطلاعات کلیدی از طرف تحلیلگران و سرمایه‌گذاران استفاده شود. [۳۴]

اخبار[۵۳] نقش مهمی در فرایند ارزیابی قیمت فعلی سهام، که توسط تحلیل‌گران، سرمایه‌گذاران و سرمایه‌گذاران نهادی[۵۴] صورت می‌گیرد، دارد. بر اساس یک دیدگاه تئوریک، ارزیابی کارآمد از یک شرکت باید بر ارزش فعلی و جریان وجوه نقد آینده شرکت تأثیر بگذارد. در اخبار نه تنها آمار و ارقام مالی بلکه اجزای متنی کیفی نیز قیمت سهام را تحت تأثیر قرار می‌دهند [۱۰]

در سال‌های اخیر روزنامه‌های سنتی به دلیل افزایش فراگیر شدن شبکه جهانی وب، به گسترش سرویس‌های خبری بر خط را در محیط وب پرداخته‌اند. برای سرمایه‌گذاران، خبرهای بلادرنگ مالی[۵۵] در تصمیم‌گیری‌ها سرمایه‌گذاری بسیار مهم است، زیرا در محیط وب خبرها به طور مکرر در حال به روزرسانی هستند. اطلاعات بیش از حد یک مشکل قابل توجه است. برای سرمایه‌گذاران غیرممکن است که همه آنها را یکی یکی بخوانند. اگر چه تحقیقات رو به افزایشی در متن‌کاوی اسناد در حال انجام است، اما دقت کم و پایین بودن نرخ بازگشت سبب می‌شود سرمایه‌گذاران زمان زیادی را برای به دست آوردن اطلاعات معنی‌دار اندکی، در گشت و گذار وب از دست بدهند. [۸]

برای بررسی تأثیر اخبار بر روی قیمت سهام مطالعات مختلفی صورت گرفته است. مهاجان و همکاران به بررسی اخبار برای شناسایی رویدادهای مهم و تأثیر آن بر بازار سهام هند پرداختند. [۳۵] هوانگ و همکاران [۸] تأثیر تیترهای خبری منتشرشده توسط روزنامه‌های الکترونیکی بر روی شاخص قیمت مالی بورس تایوان را مورد بررسی قرار دادند. هاگونا[۵۶] و همکاران سعی کرده‌اند با استفاده از روش‌های متن‌کاوی و انتخاب ویژگی‌ها از طریق بازخورد بازار به بهبود پیش‌بینی قیمت سهام بپردازند [۱۰]. مدل دیگری که بر روی پیش‌بینی قیمت سهام بر اساس اخبار منتشرشده می‌پردازد سیستم متن مالی آریزونا[۵۷] است که توسط شوماکر[۵۸] و همکاران طراحی‌شده است [۳۳]. در شرکت مشاوره‌ای تاتا در هند صورت گرفته است به دنبال ایجاد سامانه‌ای جهت تحلیل بازار

[52] Nextual

[53] News
[54] Institutional traders
[55] Real-time financial news
[56] Hagenau
[57] Arizona Financial Text
[58] Robert P. Schumaker

سهام با استفاده از بررسی اخبار مالی منتشرشده به‌منظور شناسایی و تشخیص اتفاقات مهمی که بر روی بازار تأثیر می‌گذارند هستند [۳۴]. در تحقیقی که در دانشگاه میسوری[۵۹] در سال ۲۰۱۷ انجام شد، فرایند جامع داده‌کاوی را برای پیش‌بینی جهت روزانه شاخص اس‌اندپی ۵۰۰ بر اساس ۶۰ ویژگی مالی و اقتصادی ارائه گردید [۳۶]. در مدل دیگری که توسط آیمن و همکاران ارائه شد، به دنبال پیش‌بینی مبتنی بر تحلیل احساسات از اخبار مالی و قیمت‌های بازار سهام بودند [۳۷].

از آنجایی که سهامی که در یک گروه بورسی قرار می‌گیرند داری رفتارهای نسبتاً مشابهی هستند از این رو تحقیقات نشان می‌دهد برای کاهش ریسک سبد سهام، تمام سهم‌ها از یک گروه انتخاب نشود و با خرید متنوع از سهم در گروه‌های مختلف ریسک خود را کاهش دهند. [۳۸] از این رو بهتر است برای تحلیل، سهم‌هایی انتخاب شود که در یک گروه قرار دارند. برای مثال به نظر می‌رسد عوامل موثر بر قیمت سهم‌های گروه شیمیایی متفاوت از عوامل گروه غذایی باشد. گروه محصولات شیمیایی بزرگترین گروه سهام از نظر ارزش سهام در بورس اوراق بهادار است، این گروه به عنوان نمونه برای مطالعه انتخاب شده است.

۲-۸-یادگیری ماشین و پیش‌بینی قیمت‌ها

مطالعات انجام شده خارجی در این حوزه بر خلاف منابع فارسی بسیار غنی است. از جمله دلایل عدم تمایل محققین ایرانی در این حوزه پیچیده و دشوار بودن فرایند انجام مراحل متن‌کاوی و عدم دسترسی مناسب به منابع سخت‌افزاری است. از انجایی که اجرای الگوریتم‌ها بر روی داده‌های ساختار یافته تاریخی بسیار ساده است، اکثر مطالعات داخلی بر روی این حوزه انجام شده است.

در مطالعه‌ای که توسط تای وو[۶۰] و همکاران در سال ۲۰۲۰ با عنوان «یک مدل شبکه عصبی حلقه‌ای جدید برای پیش بینی قیمت سهام» بر روی سهام بورس تایوان انجام شده است، توضیح داده می‌شود که استفاده از شبکه عصبی در یادگیری عمیق بر اساس ویژگی‌ها بسیار موثر است. همچنین از چارچوب شبکه عصبی حلقوی می‌توان برای انتخاب ویژگی و پیش‌بینی قیمت سهام با استفاده از دیتای تاریخی استفاده کرد. در این مقاله از ترکیب این دو روش برای پیش‌بینی قیمت سهام استفاده شده است. [۳۹]

در تحقیق دیگری با عنوان «روش یادگیری کارآمد ماشین هیبریدی برای پیش بینی بازار سهام سری زمانی« توسط عبادتی و مرتضوی، با استفاده از یک روش ترکیبی از الگوریتم ژنتیک و تکنیک شبکه عصبی مصنوعی برای تهیه روشی برای پیش بینی قیمت سهام و سری های زمانی استفاده نموده اند. در این روش مقادیر خروجی الگوریتم ژنتیک وارد الگوریتم توسعه یافته تکنیک شبکه عصبی مصنوعی می شوند تا خطاها را در نقطه دقیق برطرف کنند. تجزیه و تحلیل نشان می دهد که روش ترکیبی الگوریتم ژنتیک و تکنیک شبکه عصبی مصنوعی می توانند دقت را در تکرارهای کمتری افزایش دهند. این تجزیه و تحلیل بر روی شاخص اصلی ۲۰۰ روزه و همچنین بر روی پنج شرکت ذکر شده در NASDAQ انجام شده است. [۴۰]

در مطالعه‌ای دیگر توسط رانجا سناپاتی[۶۱] و همکاران در سال ۲۰۱۸ با عنوان «یک مدل جدید برای پیش بینی قیمت سهام با استفاده از شبکه عصبی ترکیبی» با استفاده از دیتای تاریخی قیمت سهم بازار بومبای و با استفاده از شبکه عصبی مصنوعی به پیش‌بینی قیمت پرداخته شده است. همچنین از الگوریتم بهینه‌سازی ازدحام ذرات برای بهینه کردن وزن ورودی استفاده شده است. در نهایت نتایج بدست آمده با نتایج مدل‌های اندازه‌گیری فاصله و شبکه عصبی مصنوعی بیزین مقایسه شد و نتایج مطلوب‌تری از این روش بدست آمد. [۴۱]

در مقاله‌ای با عنوان «یک روش جامع خوشه‌بندی و طبقه‌بندی برای پیش‌بینی بازده روزانه بازار سهام» توسط ژونگ[۶۲] و انکه[۶۳] در سال ۲۰۱۷ از روش fuzzy-means برای پاک‌سازی و PCA برای کاهش ابعاد داده‌ها استفاده شده است. خوشه‌بندی و دسته‌بندی با استفاده از روش‌های شبکه عصبی و رگرسیون لجستیک انجام و نشان داده شد شبکه عصبی نتایج مطلوب‌تری به همراه دارد. [۴۲]

در مقاله‌ای با عنوان «پیش‌بینی رفتار بازار سهام با استفاده از تکنیک داده‌کاوی و تجزیه و تحلیل احساسات اخبار« توسط آیمن خدر و همکاران در سال ۲۰۱۷ بر روی پیش‌بینی قیمت سهام انجام شد با استفاده از الگوریتم ناوی بایاس جهت‌گیری (مثبت یا منفی) اخبار را مشخص کردن بو با ترکیب جهت‌گیری‌ها با قیمت تاریخی و سپس استفاده از الگوریتم

[59] Missouri University of Science and Technology
[60] Jimmy Ming-Tai Wu
[61] Manas Ranjan Senapati
[62] Xiao Zhong
[63] David Enke

کاامین نزدیکترین همسایه جهت قیمت سهام (مثبت یا منفی) را مشخص کردن. [۴۳]

در تحقیقی دیگر توسط شوماکر[۶۴] و همکاران در سال ۲۰۱۲ با عنوان «ارزیابی احساسات در مقالات اخبار مالی» پیش‌بینی بازار سهام با استفاده از ابزار تحلیل احساسات بر اساس اخبار منتشرشده و روش استفاده از روش رگرسیون بردار پشتیبان پرداخته شد. [۳۳]

در مقاله‌ای که توسط هانگ[۶۵] و همکاران در سال ۲۰۱۰ با عنوان «درک عامل انتشار اخبار بر اساس قوانین انجمنی و تکنیک‌های داده‌کاوی» منتشر شد، با استفاده از سیستم پردازش اطلاعات و دانش چینی برای جداسازی کلمات و استفاده از الگوریتم قوانین انجمنی وزن دهی برای تشخیص دو یا چند عبارت تأثیرگذار در تیتر خبرها بر روی شاخص بورس تایوان استفاده شد. [۱۱]

در تحقیقی دیگر توسط گروس[۶۶] و مونترمن[۶۷] در سال ۲۰۱۱ با عنوان «یک رویکرد مدیریت ریسک در بازار روزانه بر مبنای تحلیل متنی» قبل از شروع فرایند داده‌کاوی اسناد از طریق سه قدم، شناسایی ویژگی‌ها، انتخاب ویژگی‌ها و نمایندگی ویژگی به داده‌های ساختاریافته و عددی تبدیل می‌شوند. سپس از طریق چهار روش ناوی بایاس، کاامین نزدیکترین همسایه، شبکه عصبی و ماشین بردار به دسته‌بندی داده‌ها می‌پردازد. در این مطالعه استفاده از روش ماشین بردار پشتیبان به نتایج بهتری منجر شده است. [۳۲]

در مطالعه‌ای دیگر توسط هانگنوا[۶۸] و همکاران با عنوان «خواندن خودکار اخبار: پیش‌بینی قیمت سهام بر اساس اخبار مالی با استفاده از ویژگی‌ها» در سال ۲۰۱۳ در مدلی چهار مرحله‌ای به استخراج ویژگی‌ها، انتخاب ویژگی‌ها و نمایندگی ویژگی‌ها با استفاده از بازخورد بازار و سپس طبقه‌بندی با استفاده از روش ماشین بردار پشتیبان پرداختند. [۴۴]

در مطالعه ای دیگر با عنوان «کاوش اخبار مالی برای وقایع مهم و و تأثیر آنها بر بازار» توسط ماهان جان[۶۹] و همکاران در سال ۲۰۰۸ روش تخصیص درکله پنهان برای شناسایی موضوعات و کلمات مرتبط مورد استفاده قرار گرفت. سپس با استفاده از روش ترکیبی شامل درخت تصمیم و ماشین بردار پشتیبان با هسته سیگموئید برای دسته‌بندی تأثیرگذاری اخبار بر بازار استفاده شد. [۴۱]

در مقاله‌ای دیگر با عنوان «خوشه‌بندی اسناد برای شناسایی رویدادها و تحلیل روند در اخبار بازار» توسط دی[۷۰] و همکاران در سال ۲۰۰۸ با استفاده از روش تخصیص دریکله پنهان برای استخراج اتفاقات مهم و استفاده از کرنل کاامین میانه برای خوشه‌بندی موضوعات استفاده شده است. [۴۱]

در این پژوهش به دلیل موجود بودن دیتای آموزش و تست می‌توان از الگوریتم‌های یادگیری با ناظر استفاده کرد. از آنجایی که می‌توان تغییر قیمت سهام را به صورت بزرگتر مساوی صفر و کوچکتر از صفر دسته بندی کرد (دسته‌بندی باینری) می‌توان از مدل ماشین بردار پشیبان سود برد. در سایر مطالعات انجام شده نیز این الگوریتم به عنوان یک الگوریتم بهینه با نتایج قابل قبول معرفی شده است. از مزایای الگوریتم ماشین بردار پشتیبان به دلیل سادگی در محاسبات نیاز کمتری به منابع سخت‌افزاری دارد.[45] به دلیل حجم بالای دیتای پردازش شده در این مطالعه و محدودیت‌های سخت‌افزاری این الگوریتم به عنوان مناسب‌ترین گزینه انتخاب شد.

۳-فرآیند انجام پژوهش

به‌منظور ورود داده‌ها به مدل، در قدم اول سیستمی با زبان برنامه‌نویسی PHP و محیط توسعه PHPStorm طراحی شده است که با خزش وب‌سایت‌های خبری پربازدید لینک آخرین خبرهای منتشر شده به‌صورت ۱۵ دقیقه یک‌بار در دیتابیس Mysql به‌عنوان وظیفه‌ای که در آینده باید انجام شود، ذخیره می‌شود. سپس لینک‌های ذخیره‌شده خوانده می‌شود و اطلاعاتی مانند عنوان خبر، متن خبر و دسته خبر در دیتابیس ذخیره می‌شود. پس از پیش‌پردازش متن‌ها، داده‌ها آماده ورود به الگوریتم یادگیری ماشین شود.

در قدم دوم داده‌های مربوط به معامله سهم از سایت بورس اوراق بهادار استخراج می‌شود. برای ذخیره‌سازی این داده‌ها ابتدا پیش‌پردازش بر روی دیتا انجام می‌شود و داده‌های سهم در دیتابیس ذخیره می‌شود.

در مرحله سوم سه ماتریس ایجاد می‌شود. ماتریس کلمات خبری که نشان دهنده تکرار کلمات در هرروز است. ماتریس TF-IDF که امتیاز هر کلمه را برای مرحله انتخاب ویژگی

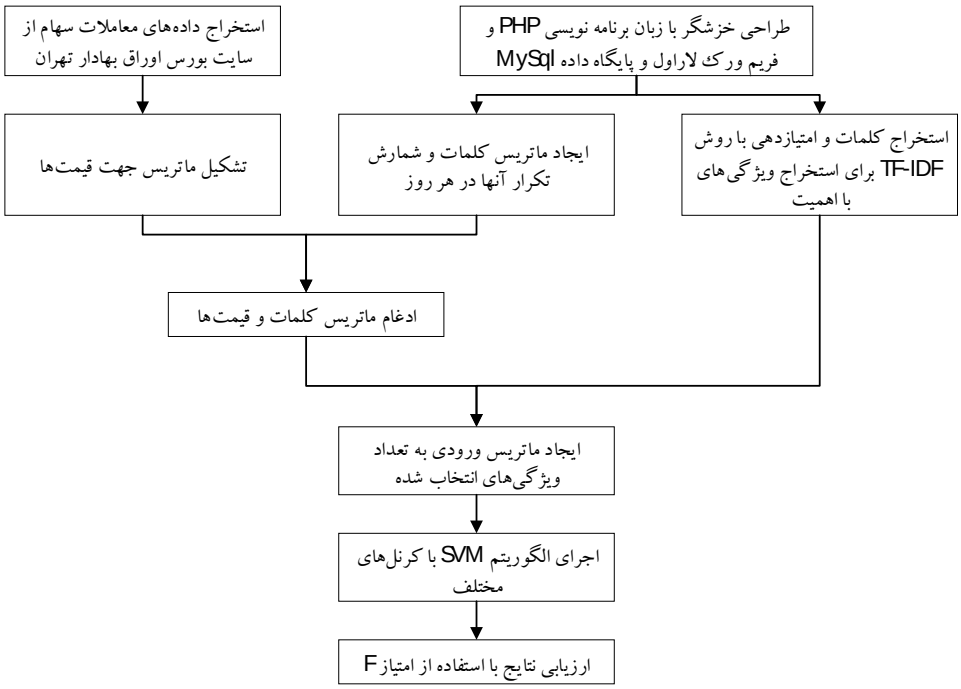
[64] Robert P. Schumaker
[65] Chenn-Jung Huang
[66] Sven S. Groth
[67] Jan Muntermann
[68] Michael Hagenau
[69] Anuj Mahajan
[70] Lipika Dey

مشخص می‌کند. ماتریس جهت قیمت که نشان دهنده جهت قیمت در هرروز است.

در مرحله چهارم پس از انتخاب کلماتی که بیشترین امتیاز را دارند ویژگی‌ها مشخص می‌شوند و از ترکیب ماتریس کلمات و جهت قیمت سهام، ماتریس ورودی مدل تشکیل می‌شود.

در قدم پنجم با استفاده از الگوریتم SVM مدل آموزش داده می‌شود و نتایج حاصل از پیش‌بینی مدل مورد ارزیابی قرار می‌گیرد. به‌غیراز مرحله خزش تمام مراحل با استفاده از زبان برنامه‌نویسی Python و محیط توسعه PyCharm انجام می‌شود.

برای انتخاب الگوریتم مناسب برای پروژه و تحقیق داده‌کاوی باید به ۵ عامل، دقت، مدت زمان آموزش الگوریتم، خطی یا غیرخطی بودن مدل، تعداد متغیرهای مسئله و تعداد ویژگی‌های انتخابی دقت کنید.



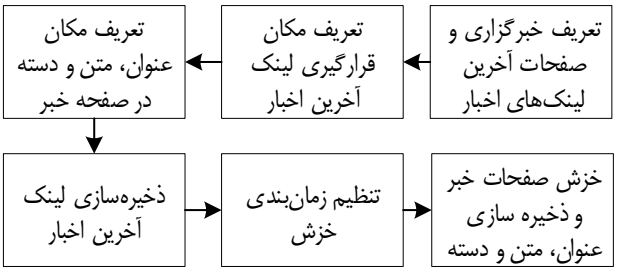
شکل ۵- روش جمع‌آوری و تجزیه و تحلیل داده‌ها

۳-۱-سیستم خزش و ذخیره‌سازی اخبار روزانه

۱-۳-۱-ذخیره‌سازی اخبار خبرگزاری‌ها

یک برنامه خزشگر برای جمع‌آوری دیتای مورد نیاز از سطح وب یکی از ضروری‌ترین بخش مطالعات وب‌کاوی است. در اینجا به بررسی یک خزشگر که در این مطالعه با زبان پی‌اچ‌پی نوشته شده است معرفی می‌شود. در شکل (۴) فرایند کار این برنامه نمایش داده‌شده است.

ین سیستم با استفاده از زبان برنامه‌نویسی پی‌اچ‌پی و چهارچوب[71] لاراول[72] نوشته شد. برای دانلود صفحات خبر از پکیج گازل و برای ذخیره‌سازی عنوان، متن و دسته خبر بر اساس آدرس مکان سی‌اس‌اس[73] از پکیج دام کراولر استفاده می‌شود. در فاز اول ۱۷ سایت خبرگزاری پربازدید ایران برای ذخیره‌سازی تعیین شده‌اند. همچنین ۲ دسته اصلی اخبار سیاسی و اخبار اقتصادی و ۱۶۰ زیر دسته برای آن‌ها تعریف‌شده است.



شکل ۶- نحوه کار پی‌اچ‌پی کراولر

[71] Framework
[72] Laravel
[73] Cascading Style Sheets

اجدول ۲- سایت‌های خبرگزاری خزش‌شده

نام خبر گذاری	نام خبر گذاری
ایسنا	شبکه خبر
عصر ایران	اقتصاد آنلاین
باشگاه خبرنگاران	خبر آنلاین
تسنیم	مشرق
خبرگزاری فارس	مهر
افکار نیوز	تابناک
موج	دنیای اقتصاد
فردا	برتر ین‌ها
اقتصاد نیوز	

الگوریتم کار این خزشگر در نمودار زیر ارائه شده است. در قدم اول داده‌های اولیه یک وب‌سایت جهت خزش ذخیره می‌شود. این داده‌ها شامل عنوان خبرگزاری، لینک اصلی خبرگزاری، صفحه‌ای که آخرین خبرها در آن منتشر می‌شود، مکان درج لینک آخرین خبرها براساس انتخابگر سی‌اس‌اس، مکان درج عنوان خبر بر اساس انتخابگر سی‌اس‌اس و مکان درج دسته خبر بر اساس انتخابگر سی‌اس‌اس ذخیره می‌شود.

id	task_url	rand_number	complete	error	created_at
1	https://www.isna.ir/news/98060402201/ترامپ-از-قران...	45	2	NULL	2019-08-26 22:24:15
2	https://www.isna.ir/news/98060402200/سیاوشی-اداره-...	152	2	NULL	2019-08-26 22:24:15
3	https://www.isna.ir/news/98060402199/فولادگر-در-جه...	26	2	NULL	2019-08-26 22:24:15
4	https://www.isna.ir/news/98060402198/سند-نهایی-گرو...	42	2	NULL	2019-08-26 22:24:15

شکل ۷- نمونه دیتای جدول crawl_tasks و لینک‌های ذخیره‌شده جهت خزش

در قدم دوم لینک آخرین خبرها با استفاده از کران جاب‌ها هر ۱۰ دقیقه به عنوان یک وظیفه ذخیره می‌شوند. این امر به این دلیل است که در طول روز بار سرور زیاد است و کار ذخیره‌سازی اخبار در انتهای روز و قبل از صبح انجام می‌شود. شبه‌کد این مرحله در زیر آورده شده است:

```
Select all agents
For each agent in agents do
   Download page for last news links
   Remove JavaScript code
   For each link in last news links do
      If link has base_agent_url
         task_link = link
      Else
         task_link = base_agent_url  + link
      If task_link not in tasks then
         Create new task
         Add random number to task
         Add task_link to task
         Insert task to database
```

در قدم سوم بررسی می‌شود آیا لینکی برای ذخیره‌سازی خبر وجود دارد یا خیر. اگر وظیفه وجود داشته باشد، لینک‌هایی که به عنوان یک وظیفه ذخیره شده بودند به صورت تصادفی انتخاب می‌شوند و اخبار آنها ذخیره می‌شود. این فرایند به خاطر جلوگیری از ارسال درخواست‌های زیاد پشت سر هم به یه خبرگزاری است.

شبه‌کد فرایند ذخیره‌سازی اخبار به صورت زیر است:

```
Select 12 tasks randomly that not crawled
Change tasks status to is crawling
If number of tasks is 0 then
   Select 3 tasks randomly that had error in crawl
For each task in task do
   Download task page
   Remove JavaScript code
   Remove tab and newline character from text
   Create new news
   Select and add title to news
   Select and add content to news
   Select news category
   Insert news to database
If news page has category then
   If news category is defined then
      Add category to news
   Else
      If category not defined in suggested category then
```

Add category to suggestion

id	crawl_task_id	title	content
1	839	اولین جایزه ادبی شهید اندرزگو برگزیدگانش را شناخت	به گزارش مشرق، نخستین دوره جایزه ادبی شهید سیدعلی ...
2	193	روزنامه پیروزی - ۵ شهریور	
3	53	راز بزرگ دوچرخه‌سوار مشهور ایتالیایی چه بود؟	انتخاب: دوچرخه سوار مشهور ایتالیایی، جینو بارتالی ...
4	58	نخست‌وزیر انگلیس: ایران نباید به سلاح هسته‌ای دست ...	نخست‌وزیر انگلیس در کفرانسی خبری اعلام کرد که اعضا...
5	581	ماجرای اصرار هاشمی بر حذف شعار مرگ بر آمریکا	خبرگزاری تسنیم: آیت‌الله اکبر هاشمی‌رفسنجانی از آ...

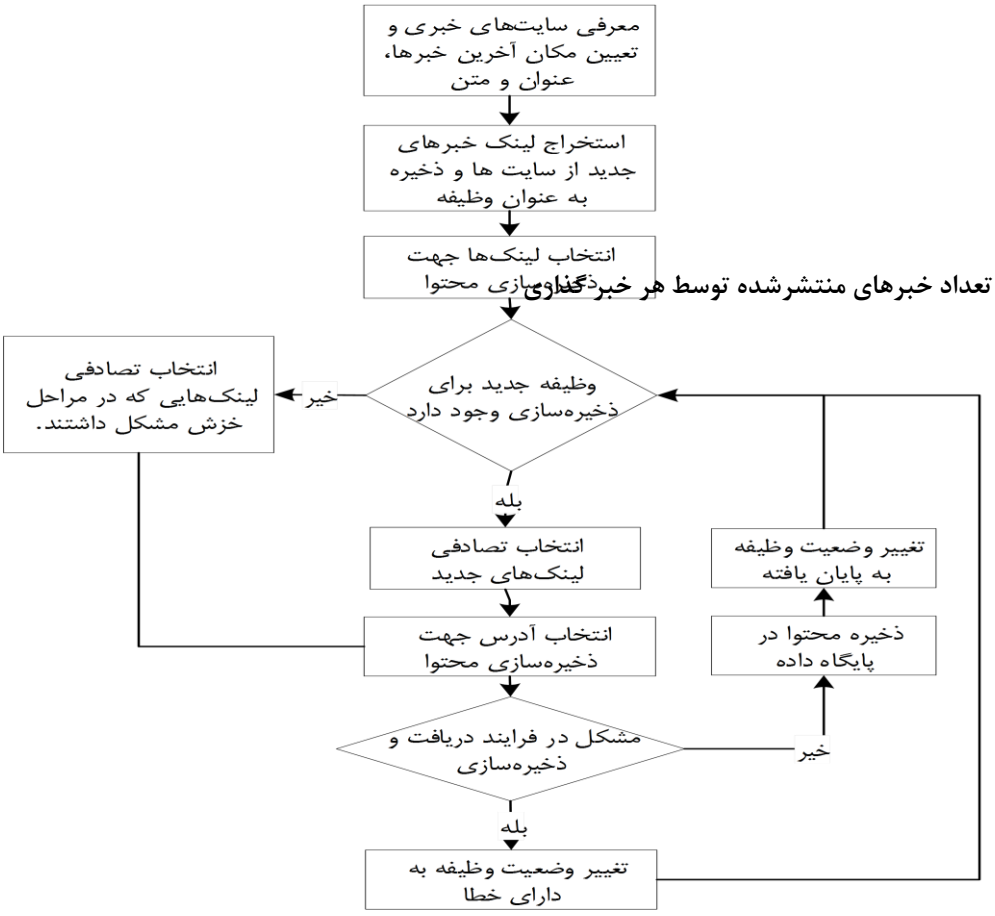
شکل ۸- نمونه دیتای جدول news (خبرهای منتشرشده)

در قدم چهارم عنوان خبر و دسته خبر ذخیره می‌شود و اگر این کار با موفقیت انجام شود وضعیت ذخیره‌سازی به پایان یافته تغییر می‌کند. اگر سرور خبرگزاری پاسخگو نباشد، وضعیت به دارای خطا تغییر می‌کند. اگر لینکی دارای خطا باشد یکبار دیگر پس از ذخیره همه خبرها خوانده می‌شود.

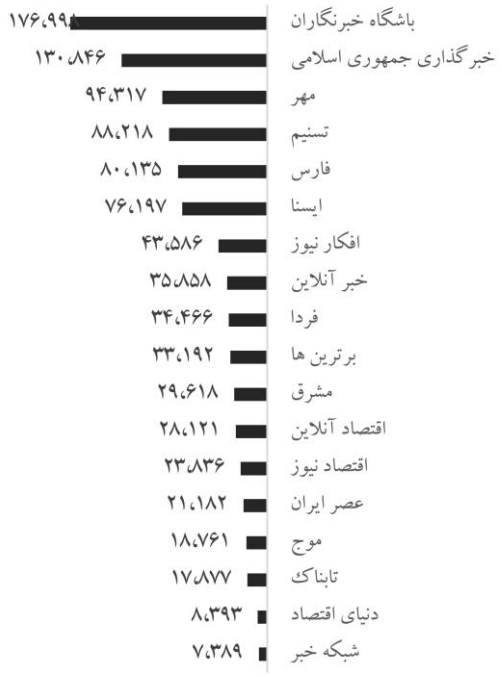
وجود کدهای جاوا اسکریپت به دلیل داشتن برخی از تگ‌های اچ‌تی‌ام‌ال در صفحات وب باعث ایجاد خطا در پکیج دام کراولر در تشخیص ساختار صفحه و در نتیجه ذخیره‌سازی می‌شود. به این دلیل پس از دانلود محتوای صفحه وب تگ‌های <script> و محتوای داخل آن از صفحه حذف می‌شود.

همچنین وجود کاراکترهای خط بعد (\n) و تب (\t) در صفحات دانلود شده، کاراکترهای زائدی هستند که برای کاهش حجم دیتابیس و مرتب شدن نمایش محتوای ذخیره‌شده، حذف می‌شوند.

شکل ۱۰- تعداد خبرهای منتشرشده توسط هر خبرگزاری



شکل ۹- الگوریتم پیشنهادی کار خزشگر



شکل ۱۱- تعداد خبرهای منتشرشده توسط هر خبر گذاری

پس از مراجعه به لینک آرشیو معاملات نقد در سایت بورس اوراق بهادار تهران به نشانی tse.ir/archive.html پس از انتخاب تاریخ موردنظر می‌توانید داده‌های سهام معامله‌شده در آن تاریخ را ذخیره کنید. لینک دانلود به‌صورت زیر است.

http://tse.ir/archive/Trade/Cash/TradeOneDay/TradeOneDay_1398_5_14.xls

حال با داشتن تاریخ می‌توان قسمت آخر لینک موردنظر را تولید و تمام فایل‌های معاملات را دانلود کرد. برای گرفتن تاریخ‌های معاملاتی کافی است در قسمت آرشیو سایت بورس اوراق بهادار فایل خلاصه معاملات برای سال ۹۷ را دانلود کرد و سپس از ستون تاریخ برای دانلود فایل ریز معاملات استفاده کرد. بدین منظور ستون تاریخ در یک فایل csv ذخیره‌شده است و در کد از سطرهای این ستون برای ساخت لینک استفاده می‌شود.

ازآنجایی‌که فایل‌های دانلود جدول‌های HTML هستند، اگر با فرمت html به‌جای xls ذخیره شوند، می‌توان توسط کتابخانه Pandas خوانده و یکپارچه شوند و سپس در دیتابیس ذخیره شوند.

۳-۲-ذخیره‌سازی داده‌های سهام

id	stock_namad	stock_name	stock_volume	stock_v	stock_transa
1	1		142594	3179589	40
2	1		508973	4336016	47
3	1		2105211	9110653	231
4	1		5666758	4785844	439
5	1		187658	2169823	38
6	1		3746245	9874156	177
7	1		6686	2496984	10

شکل ۱۲- نمونه دیتای جدول transaction (تراکنش‌های سهام)

در مرحله بعد داده مربوط به شرکت‌های بورس اوراق بهادار را ذخیره می‌شود تا بتوانیم بر اساس دسته تعیین‌شده برای هر سهم، سهام مربوط به شرکت‌های گروه پتروشیمی را تعیین و الگوریتم را بر روی آن‌ها اجرا کنیم. برای این کار از لینک زیر استفاده می‌کنیم و داده‌ها را آماده ورود به سیستم می‌کنیم.

http://www.tsetmc.com/Loader.aspx?ParTree=111C1417

در این لینک داده‌های مربوط به هر سهم موجود است و می‌توان از داده‌های ذخیره شده در مرحله قبل، سهم های مربوط به گروه محصولات شیمیایی را ذخیره کرد.

stock_code	stock_group	stock_industry
IRB5IKCO8751	N2	خودرو و ساخت قطعات
IRO1APPE0001	N2	رایانه و فعالیت‌های وابسته به آن
IRO1ASIA0001	N1	ه وصندوق بازنشستگی به جزتامین اجتماعی
IRO1CONT0001	N1	ابزاریزشکی، اپتیکی و اندازه‌گیری
IRR1CONT0101	N1	ابزاریزشکی، اپتیکی و اندازه‌گیری
IRO1MKRT0001	N1	

شکل ۱۳- نمونه دیتا جدول company (داده‌های مربوط به شرکت‌ها)

۳-۳-پیش‌پردازش متن

پیش‌پردازش متن از مراحل بسیار مهم در متن‌کاوی است. ازاین‌رو در این پژوهش چندین مرحله جهت پیش‌پردازش و حذف داده‌های اضافه و یکسان‌سازی استفاده شده است. به دلیل حجم بالای خبرها، قبل از پیش‌پردازش خبرهایی انتخاب شدند که فقط در دو دسته سیاسی و اقتصادی دسته‌بندی‌شده بودند. دیتای جدید در جدول news_back ذخیره شد. در این مرحله ۸۰۶۱۸ خبر انتخاب شد.

۱-۳-۳-نرمال‌سازی متن

در این مرحله با استفاده از کتابخانه هضم عنوان و متن خبرها نرمال‌سازی شد. به دلیل حجم داده‌ها برای استفاده از حداکثر ظرفیت سیستم، از روش multi thread در کدهای پایتون استفاده شد. یعنی شناسه خبرها در دسته‌های ۱۰۰هزارتایی در یک حلقه قرار می‌گرفتند و کد نرمال‌سازی به‌صورت موازی برای خبرها اجرا شد.

در حالت ساده عملیات نرمال‌سازی متن با مراحل زیر انجام‌پذیر است:

اصلاح انواع حرف «ک» به معادل فارسی آنان.

اصلاح انواع حرف «ی» به معادل فارسی آنان.

بررسی همزه و انواع مختلف املاهای موجود و اصلاح هرکدام (به‌عنوان‌مثال تبدیل ؤ به و ، ئ به ی ، أ به ا ، إ به ا و...)

حذف شناسه‌ی همزه از انتهای واژه‌هایی مثل شهداء

حذف شناسه «آ» به «ا» مانند: آب به اب

اصلاح نویسه‌ی «طور» در واژه‌هایی مانند به‌طور، آن‌طور، این‌طور و ...

بررسی وجود حرف «ی» در انتهای لغاتی مانند خانه‌ی ما و اصلاح آنان

حذف تشدید از واژه‌ها

تبدیل ارقام عربی و انگلیسی به معادل فارسی.

اصلاح نویسه‌ی نیم‌فاصله

اصلاح اعراب و حذف فتحه، کسره و ضمه و همچنین تنوین‌ها

حذف نیم‌فاصله‌های تکراری

حذف نویسه‌ی «ـ» که برای کشش نویسه‌های چسبان مورداستفاده قرار می‌گیرد. مانند تبدیل«بـــــــر» و «بـــر» به «بر»

چسباندن پسوندهای «تر»، «ترین» و ... به آخر واژه‌ها

اصلاح فاصله‌گذاری «ها» در انتهای واژه‌ها و همچنین پسوندهای «های»، «هایی»، «هایم»، «هایت»، «هایش» و ...

اصلاح فاصله‌گذاری «می»، «نمی»، «درمی»، «برمی»، «بی» در ابتدای واژه‌ها

تبدیل «ه» به «ه‌ی»

تبدیل «ب» متصل به ابتدای واژه‌ها به «به»

اصلاح فاصله‌گذاری پسوندها

حذف فاصله‌ها و نیم‌فاصله‌های اضافه بکار رفته در متن

تصحیح فاصله‌گذاری در مورد علائم سجاوندی بدین‌صورت که علائم سجاوندی به لغات قبل از خود می‌چسبند و با لغت بعد از خود فاصله خواهند داشت. [۴۶]

۲-۳-۳-حذف کلمات توقف

در مرحله دوم پیش‌پردازش متن تمامی کلمات توقف از متن حذف شدند. برای جلوگیری افزایش حجم جدول‌ها داده‌های هر مرحله در جدول جدیدی ذخیره شدند. در این مرحله ابتدا داده‌های مرحله قبل در جدول news_stop ذخیره و سپس مرحله حذف کاراکترهای اضافه و کلمات توقف بر روی دیتا اجرا شد و ستون جدید content_nomalize_stop به دیتا اضافه شد.

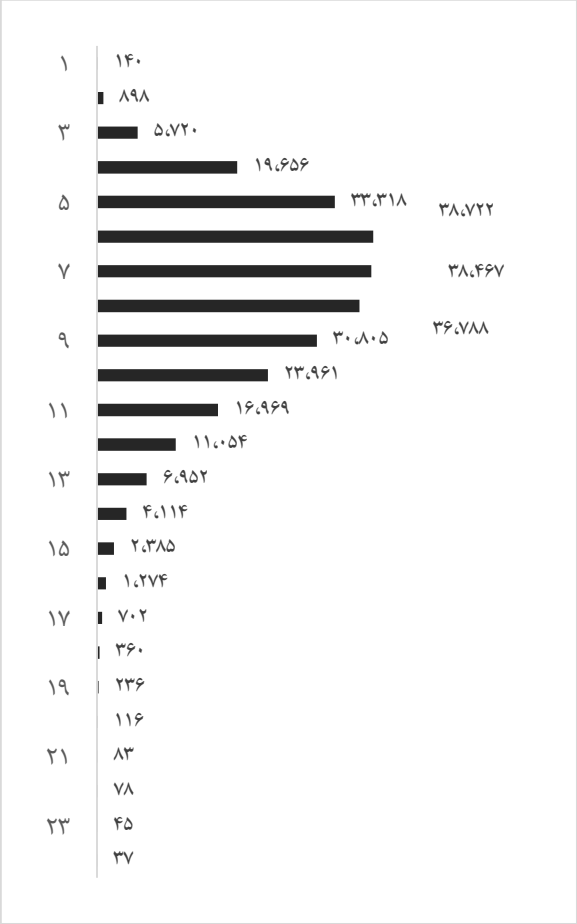
۳-۳-۳-ذخیره تمامی کلمات و حذف کلمات و کاراکترهای زائد

برای تشخیص کاراکترها و کلمات زائد ابتدا تمام کلمات در جدول جدید words ذخیره شدند. در این مرحله ابتدا با متد tokenize در کتابخانه هضم خبرها به کلمات تجزیه شدند، سپس تمامی کاراکترهای اضافه که لیست آن در زیر آمده است و همچنین تمامی کاراکترهای انگلیسی از کلمات حذف شدند.

{}⁎$#@!s-\،«»<>:؛/()

_ 《》 €﴾﴿±@؟?&×+۴۳۵۲۷;=

بعدازاین مرحله اگر کلمه موردنظر به‌صورت کامل حذف نشد و در جدول کلمات موجود نباشد در جدول ذخیره می‌شود. درمجموع ۲۷۲۸۸۰ کلمه یکتا ذخیره شد.



شکل ۱۴-تعداد کلمات بر اساس تعداد کاراکترها

از این تعداد کلمات ۱۰۹۷ کلمه به‌عنوان کلمات زائد شناسایی و به‌عنوان کلماتی که نباید در محاسبات شمارش شوند علامت‌گذاری شدند. که عمدتاً کلمات یک و دو کاراکتری هستند که شامل علامت‌های خاص و نگارشی، حروف، حرف‌های اضافه و غیره هستند. در این مرحله اگر لغات نامه استانداردی در اختیار باشد که بتوان کلمات استخراج‌شده را با آن مقایسه کرد و کلماتی که در لغت‌نامه نیستند را مشخص کرد می‌توان ابعاد مسئله را کاهش داد. همچنین اگر برنامه‌ای باشد تا غلط‌های املایی را اصلاح کرد می‌تواند به کاهش ابعاد کمک کند.

۴-۳-۳-ایجاد جدول کلمات و شمارش کلمات در هرروز

پس از مرحله پیش‌پردازش متن، ماتریس کلمات ورودی به مدل ایجاد شد. در این بخش مراحل زیر بر روی دیتا انجام شد. در قدم اول تعداد تکرارهای هر کلمه در روز موردنظر شمارش شد و شناسه کلمه به همراه تعداد تکرار در جدول extract_word ذخیره شد. ازآنجایی‌که ساعت معاملات در بورس اوراق بهادار تهران از ساعت ۸:۳۰ تا ۱۲:۳۰ دقیقه است، خبرهایی که در ساعت بعد از ۱۲:۳۰ منتشر می‌شوند جزء اخبار روز بعد تلقی می‌شوند و بر روی قیمت‌های روز آینده تأثیرگذار هستند. بدین‌جهت کلمات آن‌ها در کلمات روز آینده شمارش می‌شوند. تعداد کلمات به تفکیک روزهای انتشار خبر در این مرحله ۳۴۳۳۰۸۳۸ است.

۵-۳-۳-انتخاب ویژگی‌ها با استفاده از روش TF-IDF

روش tf × idf (تکرار کلمه × تکرار سند معکوس)، یک روش نمایندگی است که معمولاً در روش‌های استخراج ویژگی‌ها از متن استفاده می‌شود. به‌صورت پیش‌فرض، تکرار کلمات[74] (tf) در سند نشان دهنده میزان اهمیت اصطلاح در این سند است و تکرار سند در این اصطلاح[75] (df، درصد اسناد حاوی این اصطلاح) نشان دهنده میزان اهمیت این اصطلاح در کل متن‌ها است. مقدار df پایین نشان می‌دهد که این اصطلاح در بسیاری از اسناد ظاهر نمی‌شود و این نشان دهنده منحصربه‌فرد بودن این اصطلاح در اسناد است. بنابراین، به‌جای استفاده از df، idf (معکوس df) را به‌عنوان برنامه وزن دهی انتخاب می‌شود. بنابراین وزن بالا در روش tf × idf نشان دهنده تکرار زیاد یک کلمه در یک سند و اسناد کم حاوی این متن است. [۴۷]

در این بخش با استفاده از Pandas ماتریس کلمات جهت محاسبه TF-IDF تشکیل شد. تمام خبرهایی که در یک روز منتشرشده‌اند به‌عنوان یک سند در نظر گرفته می‌شوند. پس باید برای محاسبه TF تعداد تکرار کلمات در هرروز را به دست آورد. سپس با تقسیم تکرار هر کلمه در هر روز TF محاسبه می‌شود. برای محاسبه IDF از فرمول $IDF = log(\frac{c_d}{i_d})$ استفاده شده است که در آن c_d نشان‌دهنده کل سندها است. که در پژوهش ما برابر تعداد روزهای تولید خبر و برابر ۱۳۶ است. i_d نیز نشان‌دهنده تعداد اسناد است که ما هرروز را برابر یک سند گرفتیم. سپس با ضرب TF در IDF شاخص TF-IDF برای هر کلمه محاسبه می‌شود. سپس ماتریس کلمات تشکیل‌شده و براساس تعداد ویژگی‌های موردنظر فیلتر می‌شود. مثلاً ابتدا ۱۰۰۰ ویژگی که بیشترین امتیاز TF-IDF را گرفته‌اند انتخاب می‌شود و از بین ویژگی‌ها فقط ۱۰۰۰ ستون انتخاب می‌شود.

۴-۳-آماده‌سازی دیتای بورس

ازآنجایی‌که نوع مدل انتخابی یادگیری با ناظر بر اساس جهت قیمت سهام است، باید دیتای سهم موردنظر را بر اساس اینکه

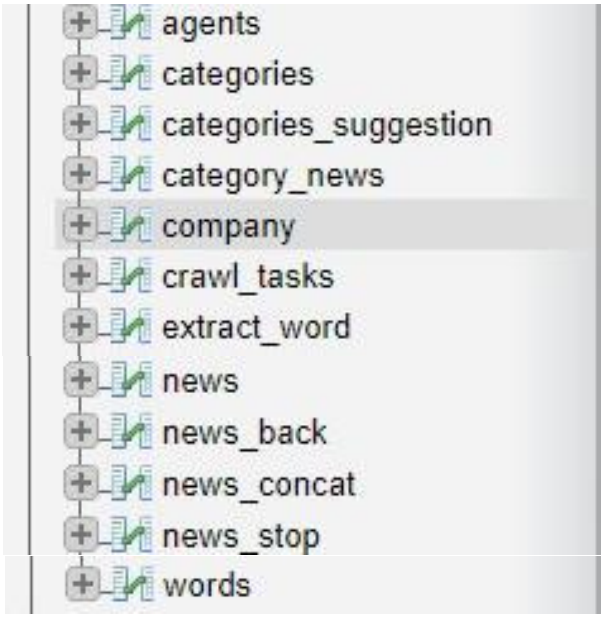
[74] Term frequency
[75] Document frequency

کاهشی یا افزایشی بوده است برچسب‌گذاری کنیم. سپس ماتریس کلمات و جهت قیمت را ادغام می‌کنیم تا دیتا جهت ورود به الگوریتم آماده شود. قیمت منفی و صفر نسبت به‌روز قبل را با برچسب صفر و قیمت مثبت را با برچسب یک برای یک سهم نمایش داده می‌شود.

created_at	change_end_price
2018-03-25	1
2018-03-26	0
2018-03-27	0
2018-03-28	0
2018-04-03	0
2018-04-04	0
2018-04-07	0
2018-04-08	1
2018-04-09	1

شکل ۱۵- برچسب‌گذاری قیمت‌های سهم در هرروز

از ادغام ماتریس کلمات و جهت قیمت یک سهم خاص در روزهای معاملاتی، ماتریس ورودی به الگوریتم ماشین بردار پشیتبان تشکیل می‌شود. به دلیل حجم بالای دیتا همان‌طور که در شکل می‌بینید انجام این فرایند حداقل به ۲۴ گیگ حافظه موقت جهت ایجاد ماتریس کلمات نیاز است.



شکل ۱۶-جدول‌های ذخیره‌سازی دیتا در سیستم خزشگر

agent: جدول ذخیره‌سازی داده‌های خبرگزاری (عنوان، لینک صفحه اصلی و غیره)

Crawl_task: جدول ذخیره‌سازی لینک خبرها به عنوان وظیفه

news: اخبار ذخیره شده هر لینک

news_back: متن خبرهای پیش پردازش شده

news_stop: حذف کلمات توقف از متن خبرها

news_concat: تجمع متن خبرهای یک روز

categories: جدول ذخیره‌سازی دسته‌بندی‌ها

categories suggestion: جدول شناسایی سایر دسته‌ها به غیر از سیاسی و اقتصادی

category_news: جدول تعیین دسته خبرها

company: جدول ذخیره‌سازی داده‌های شرکت‌های بورسی

transaction: جدول ذخیره‌سازی داده‌های بورسی سهم

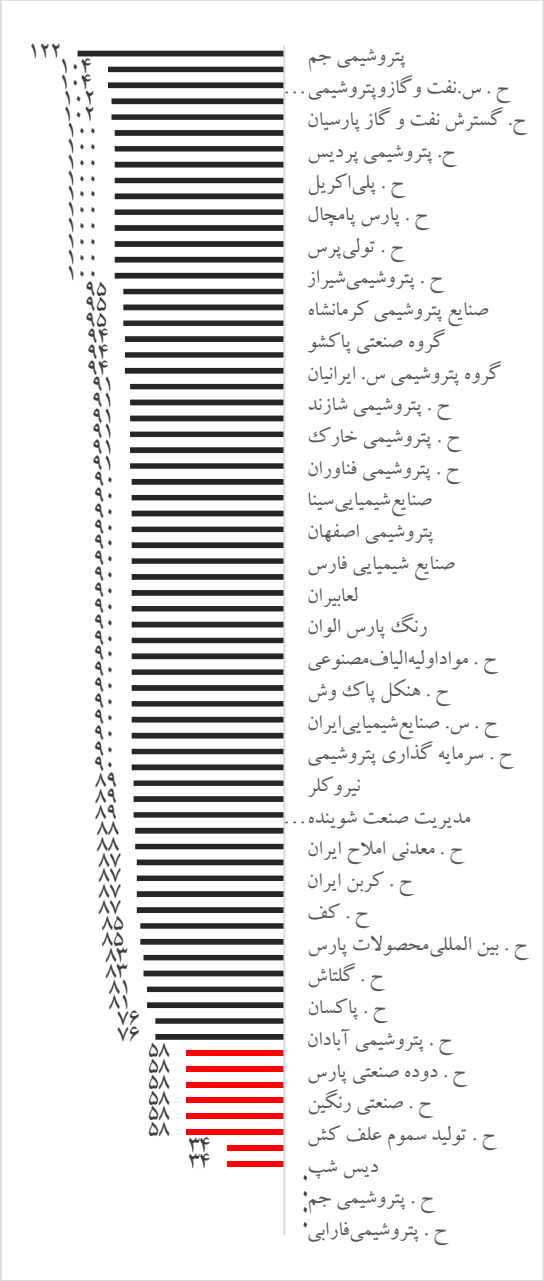
words: کلمات استخراج شده یکتا از خبرها بعد از پاکسازی

Extract_word: کلمات موجود در هر خبر بر اساس جدول words

۳-۵-اجرای مدل ماشین بردار پشتیبان با کرنل‌های خطی و غیرخطی

در این پژوهش از ۴ الگوریتم ماشین بردار پشتیبان با تنظیمات مختلف استفاده شده است که درنهایت به بررسی ۷ مدل خواهیم پرداخت. مدل‌های بررسی‌شده شامل: کرنل خطی، کرنل چندجمله‌ای درجه ۲ با گاما auto، کرنل چندجمله‌ای درجه ۲ با گاما scale، کرنل شعاعی با گاما auto ، کرنل شعاعی با گاما scale، کرنل سیگموئید با گاما auto، کرنل سیگموئید با گاما scale.

برای آموزش مدل از دو پارامتر متغیر برای انجام محاسبات مختلف استفاده شده است. پارامتر اول ویژگی‌های انتخابی و پارامتر دوم تعداد داده‌های آزمون است. برای هر سهم پارامتر ویژگی‌های انتخابی از ۱۰۰۰ تا ۱۹۰۰۰ ویژگی است. مدل برای هر ویژگی انتخابی با تعداد داده‌های آزمون ۱۰ درصد تا ۳۵ درصد اجرا می‌شود. سهم‌هایی انتخاب‌شده‌اند که بیش از ۷۰ روز در بازه زمانی موردنظر معامله‌شده‌اند.



شکل ۱۷- تعداد روزهای معاملاتی هر سهم

پس از آموزش مدل نتایج ارزیابی پیش‌بینی‌ها در شاخص‌های دقت، صحت و امتیاز و میانگین ساده و وزنی هر سهم برای هر پارامتر محاسبه می‌شود.

```
number of future:1000
[[9 0]
 [3 4]]
              precision    recall  f1-score   support

         0.0       0.75      1.00      0.86         9
         1.0       1.00      0.57      0.73         7

   micro avg       0.81      0.81      0.81        16
weighted avg       0.86      0.81      0.80        16
```

شکل ۱۸- نمونه‌ای از نتایج ارزیابی مدل

در شکل ۱۸ خط اول: نشان‌دهنده تعداد ویژگی‌های انتخابی
خط دوم و سوم: ماتریس در هم ریختگی
خط چهارم: شاخص‌های ارزیابی
خط ششم: نتایج ارزیابی پیش‌بینی جهت‌های منفی
خط هفتم: نتایج ارزیابی پیش‌بینی جهت‌های مثبت
خط هشتم: میانگین ساده هر شاخص
خط نهم: میانگین وزنی هر شاخص

ازآنجایی‌که امتیاز F معیار دقت و صحت را به صورت همزمان در نظر می‌گیرد شاخص مناسبی برای ارزیابی یک مدل دسته‌بندی است، برای تحلیل نتایج این پژوهش از این شاخص استفاده شده است.

نتیجه نهایی هر الگوریتم برای هر سهم در یک فایل CSV برای تحلیل در نرم‌افزار اکسل ذخیره می‌شود. نتایج حاصل از هر مدل در یک پوشه به اسم کرنل مورداستفاده شده ذخیره می‌شود. برای تهیه گزارش‌ها علاوه بر نتایج ارزیابی تعداد ویژگی‌های انتخاب‌شده و اندازه نمونه‌های آزمون نیست در فایل ذخیره می‌شود.

	A	B	C	D	E	F	G
1	index	f1-score	precision	recall	support	number_of	test_size
2	0	0.6	0.5	0.75	4	1000	0.1
3	1	0.5	0.666667	0.4	5	1000	0.1
4	micro avg	0.555556	0.555556	0.555556	9	1000	0.1
5	macro avg	0.55	0.583333	0.575	9	1000	0.1
6	weighted a	0.544444	0.592593	0.555556	9	1000	0.1
7	0	0.285714	0.25	0.333333	3	2000	0.1
8	1	0.545455	0.6	0.5	6	2000	0.1
9	micro avg	0.444444	0.444444	0.444444	9	2000	0.1
10	macro avg	0.415584	0.425	0.416667	9	2000	0.1
11	weighted a	0.458874	0.483333	0.444444	9	2000	0.1
12	0	0.222222	1	0.125	8	3000	0.1
13	1	0.222222	0.125	1	1	3000	0.1

شکل ۱۹- نمونه فایل ارزیابی پیش‌بینی مدل

۳-۶-نتایج ارزیابی مدل برای هر سهم

در این بخش به بررسی نتایج ارزیابی هر سهم خواهیم پرداخت و برای هر سهم بهترین کرنل با پارامترهای مختلف را انتخاب خواهیم کرد. سهم‌ها به ترتیب بیشترین روز معاملاتی موردبررسی قرار می‌گیرند. برای تحلیل نتایج از نرم‌افزار اکسل استفاده شده است. در قدم اول فایل‌های ارزیابی مدل توسط ابزار پاور کوئری بارگذاری و آماده شدند. سپس توسط ابزار پیوت تیبل خروجی‌ها و نمودارها ایجاد شد. ازآنجایی‌که برای ارزیابی از میانگین وزنی امتیاز F استفاده شد، تعداد سطرهای بارگذاری شده توسط اکسل ۲۰۰۶۴ سطر است.

تعداد روش‌ها: ۷

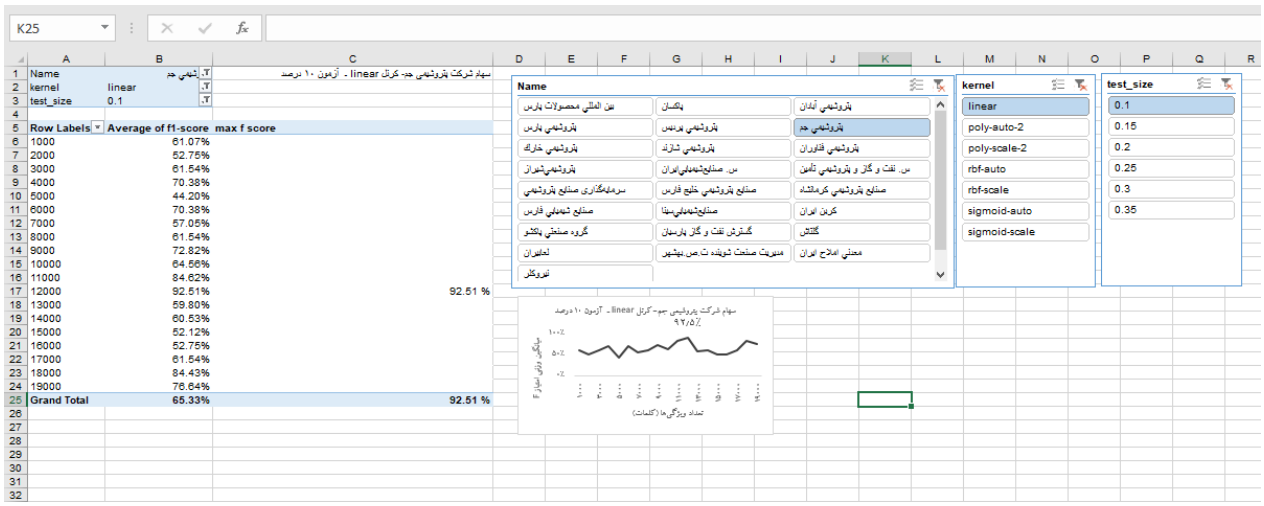
تعداد انتخاب شاخص‌ها: ۱۹

تعداد انتخاب نمونه آزمایش: ۶

تعداد سهم تحلیل‌شده: ۲۵

پس تعداد نتایج برابر است:

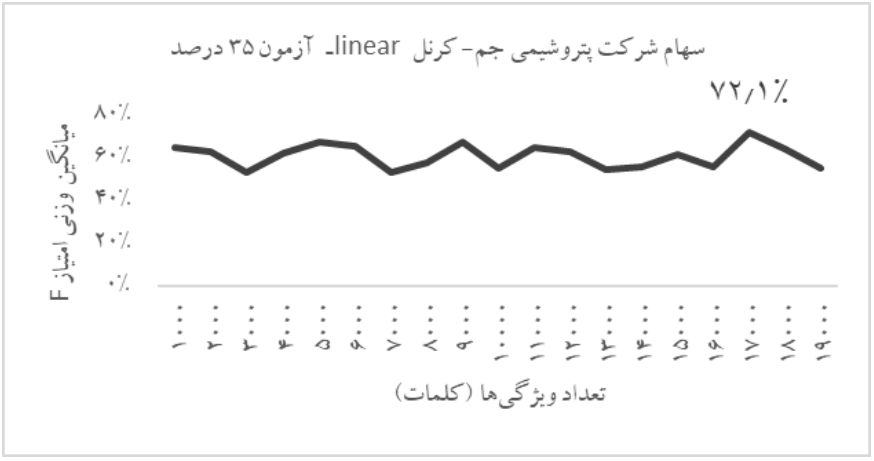
۷×۱۹×۶×۲۵=۲۰۰۶۴

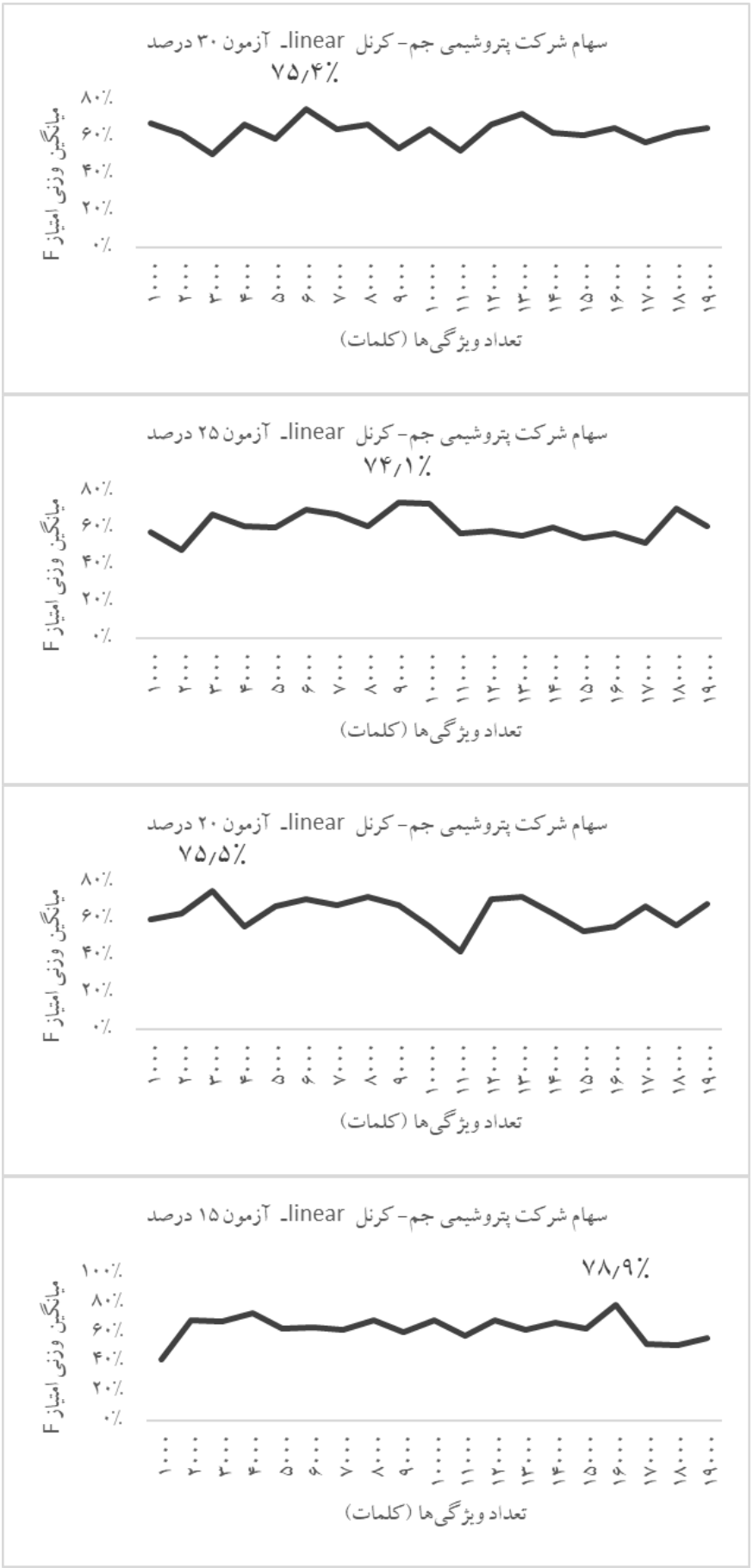


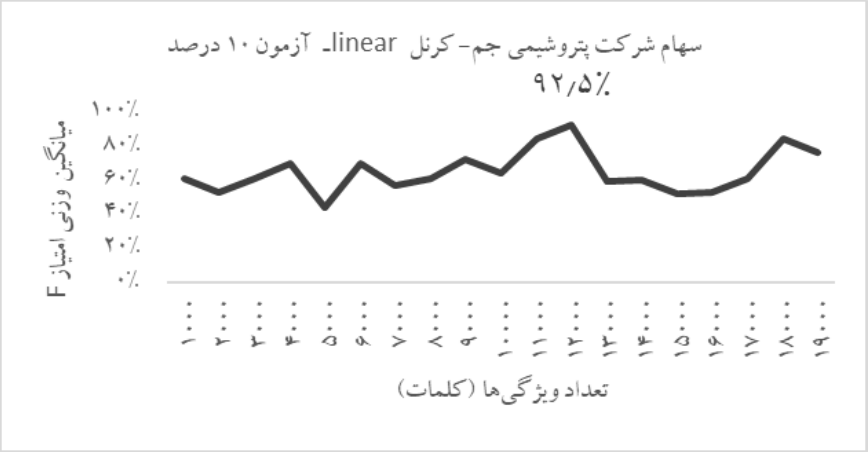
شکل ۲۰- تحلیل نتایج در اکسل

۱-۳-۶-نمونه ارزیابی نتایج برای سهم شرکت پتروشیمی جم

نتایج آزمون با کرنل خطی برای سهم پتروشیمی جم به‌صورت زیر است.




سهام شرکت پتروشیمی جم- کرنل linear- آزمون ۳۰ درصد
۷۵/۴٪
میانگین وزنی امتیاز F
تعداد ویژگی‌ها (کلمات)
سهام شرکت پتروشیمی جم- کرنل linear- آزمون ۲۵ درصد
۷۴/۱٪
میانگین وزنی امتیاز F
تعداد ویژگی‌ها (کلمات)
سهام شرکت پتروشیمی جم- کرنل linear- آزمون ۲۰ درصد
۷۵/۵٪
میانگین وزنی امتیاز F
تعداد ویژگی‌ها (کلمات)
سهام شرکت پتروشیمی جم- کرنل linear- آزمون ۱۵ درصد
۷۸/۹٪
میانگین وزنی امتیاز F
تعداد ویژگی‌ها (کلمات)



۱-۳-۶-ارزیابی کلی نتایج

بررسی ۲۰۰۶۴ حالتی که برای این ۲۵ سهم اتفاق می‌افتد به‌صورت تک‌به‌تک بسیار کار دشواری است. بدین منظور برای بررسی بهترین حالت از پیوت تیبل در اکسل استفاده شده است و برای هر کرنل بهترین حالت انتخاب می‌شود. برای کرنل خطی برای هر سهم تنظیمات زیر بهترین نتیجه را خواهد داشت. اطلاعات این جداول به شرح زیر است:

شرکت: نام شرکت در گروه محصولات شیمیایی

ویژگی‌ها: تعداد کلمات انتخاب شده بر اساس بیشترین امتیاز TF-IDF

درصد آزمایش: درصد روزهای معاملاتی که به عنوان بخش تست در نظر گرفته شده اند.

کرنل: امتیاز F بدست آماده از این کرنل با توجه به تعداد ویژگی‌های انتخاب شده و درصد آزمایش

جدول ۳- ارزیابی نتایج کرنل خطی

شرکت	ویژگی‌ها	درصد آزمایش	کرنل خطی
بین‌المللی محصولات پارس	۹۰۰۰	۰.۲۵	۷۶.۹۱%
پاکسان	۱۱۰۰۰	۰.۱	۸۹.۵۷%
پتروشیمی آبادان	۱۰۰۰	۰.۳	۸۶.۹۶%
پتروشیمی پارس	۱۷۰۰۰	۰.۳	۸۱.۱۹%
پتروشیمی پردیس	۱۸۰۰۰	۰.۱۵	۷۳.۳۳%
پتروشیمی جم	۱۲۰۰۰	۰.۱	۹۲.۵۱%
پتروشیمی خارک	۵۰۰۰	۰.۱۵	۷۱.۴۳%
پتروشیمی شازند	۹۰۰۰	۰.۱۵	۷۱.۴۳%
پتروشیمی فناوران	۹۰۰۰	۰.۱	۱۰۰.۰۰%
پتروشیمی شیراز	۳۰۰۰	۰.۱۵	۸۶.۰۰%
س. صنایع‌شیمیایی‌ایران	۱۵۰۰۰	۰.۲۵	۷۳.۸۱%
س. نفت و گاز و پتروشیمی تأمین	۸۰۰۰	۰.۱	۱۰۰.۰۰%
سرمایه‌گذاری صنایع پتروشیمی	۱۷۰۰۰	۰.۱	۸۸.۸۹%
صنایع پتروشیمی خلیج فارس	۷۰۰۰	۰.۱	۹۰.۳۳%
صنایع پتروشیمی کرمانشاه	۷۰۰۰	۰.۱	۹۱.۳۷%
صنایع شیمیایی فارس	۶۰۰۰	۰.۱	۷۷.۷۸%
صنایع‌شیمیایی سینا	۱۱۰۰۰	۰.۱	۸۹.۱۸%
کربن ایران	۱۸۰۰۰	۰.۱	۸۱.۷۵%
گروه صنعتی پاکشو	۱۵۰۰۰	۰.۲	۸۴.۶۴%
گسترش نفت و گاز پارسیان	۶۰۰۰	۰.۱	۹۱.۰۶%
گلتاش	۳۰۰۰	۰.۱	۸۹.۵۷%
لعابیران	۱۷۰۰۰	۰.۱۵	۷۸.۴۶%
مدیریت صنعت شوینده ت.ص.بهشهر	۷۰۰۰	۰.۱	۸۸.۸۹%
معدنی املاح ایران	۷۰۰۰	۰.۱	۸۸.۲۱%
نیروکلر	۶۰۰۰	۰.۱	۸۹.۵۷%
میانگین			۸۵.۳۱%

با توجه به داده‌های بالا مدل خطی می‌تواند به‌صورت میانگین تا ۸۵,۳۱ درصد جهت قیمت‌ها را درست پیش‌بینی کند.

جدول ۴-ارزیابی نتایج کرنل چندجمله‌ای درجه ۲ با گاما auto

شرکت	تعداد ویژگی‌ها	درصد آزمایش	معادله درجه ۲ با گاما auto
بین‌المللی محصولات پارس	۱۳۰۰۰	۰.۲	۸۱.۵۸%
پاکسان	۵۰۰۰	۰.۱	۱۰۰.۰۰%
پتروشیمی آبادان	۱۳۰۰۰	۰.۱	۸۷.۷۰%
پتروشیمی پارس	۱۳۰۰۰	۰.۱	۸۰.۰۰%
پتروشیمی پردیس	۳۰۰۰	۰.۱	۸۹.۳۳%
پتروشیمی جم	۳۰۰۰	۰.۱	۸۴.۴۳%
پتروشیمی خارک	۶۰۰۰	۰.۱۵	۷۰.۸۳%
پتروشیمی شازند	۴۰۰۰	۰.۱	۷۰.۹۹%
پتروشیمی فناوران	۱۱۰۰۰	۰.۱	۱۰۰.۰۰%
پتروشیمی شیراز	۶۰۰۰	۰.۱	۱۰۰.۰۰%
س. صنایع‌شیمیایی‌ایران	۱۶۰۰۰	۰.۱	۸۷.۵۰%
س. نفت و گاز و پتروشیمی تأمین	۴۰۰۰	۰.۱	۱۰۰.۰۰%
سرمایه‌گذاری صنایع پتروشیمی	۱۴۰۰۰	۰.۱	۸۷.۴۱%
صنایع پتروشیمی خلیج فارس	۵۰۰۰	۰.۱	۸۰.۰۰%
صنایع پتروشیمی کرمانشاه	۵۰۰۰	۰.۱	۸۰.۰۰%
صنایع شیمیایی فارس	۱۶۰۰۰	۰.۱	۷۸.۳۳%
صنایع شیمیایی سینا	۱۸۰۰۰	۰.۱	۸۹.۱۸%
کربن ایران	۱۴۰۰۰	۰.۱	۹۰.۳۷%
گروه صنعتی پاکشو	۱۸۰۰۰	۰.۱	۹۰.۱۰%
گسترش نفت و گاز پارسیان	۱۱۰۰۰	۰.۱	۸۲.۱۲%
گلتاش	۱۲۰۰۰	۰.۱۵	۸۴.۶۲%
لعابیران	۱۵۰۰۰	۰.۱	۸۹.۱۸%
مدیریت صنعت شوینده ت.ص.بهشهر	۸۰۰۰	۰.۱	۸۹.۱۸%
معدنی املاح ایران	۹۰۰۰	۰.۱	۱۰۰.۰۰%
نیروکلر	۲۰۰۰	۰.۱	۸۸.۸۹%
میانگین			۸۷.۲۷%

با توجه به داده‌های بالا کرنل معادله درجه ۲ با گاما auto می‌تواند به‌صورت میانگین تا ۸۷,۲۷ درصد جهت قیمت‌ها را درست پیش‌بینی کند.

جدول ۵-ارزیابی نتایج کرنل چندجمله‌ای درجه ۲ با گاما scale

شرکت	تعداد ویژگی‌ها	تعداد آزمایش	معادل درجه ۲ با گاما scale
بین‌المللی محصولات پارس	۴۰۰۰	۰.۱۵	۸۴.۶۲%
پاکسان	۱۹۰۰۰	۰.۱۵	۷۷.۵۶%
پتروشیمی آبادان	۱۳۰۰۰	۰.۱	۷۵.۰۰%
پتروشیمی پارس	۱۹۰۰۰	۰.۱۵	۶۹.۱۴%
پتروشیمی پردیس	۱۰۰۰	۰.۲	۷۸.۱۱%
پتروشیمی جم	۴۰۰۰	۰.۱۵	۷۳.۲۵%
پتروشیمی خارک	۱۵۰۰۰	۰.۱۵	۶۲.۵۷%
پتروشیمی شازند	۴۰۰۰	۰.۱	۷۱.۱۱%
پتروشیمی فناوران	۳۰۰۰	۰.۱۵	۸۵.۷۱%
پتروشیمی شیراز	۱۲۰۰۰	۰.۱	۷۱.۱۱%
س. صنایع شیمیایی ایران	۱۵۰۰۰	۰.۱	۸۳.۶۶%
س. نفت و گاز و پتروشیمی تأمین	۷۰۰۰	۰.۱	۷۳.۶۴%
سرمایه‌گذاری صنایع پتروشیمی	۵۰۰۰	۰.۱	۸۷.۵۰%
صنایع پتروشیمی خلیج فارس	۲۰۰۰	۰.۱	۸۵.۲۶%
صنایع پتروشیمی کرمانشاه	۱۴۰۰۰	۰.۱	۱۰۰.۰۰%
صنایع شیمیایی فارس	۹۰۰۰	۰.۲۵	۶۰.۸۷%
صنایع‌شیمیایی‌سینا	۸۰۰۰	۰.۱۵	۷۲.۰۲%
کربن ایران	۱۶۰۰۰	۰.۱۵	۸۹.۴۲%
گروه صنعتی پاکشو	۱۴۰۰۰	۰.۲	۷۶.۹۹%
گسترش نفت و گاز پارسیان	۱۷۰۰۰	۰.۲۵	۶۵.۶۴%
گلتاش	۵۰۰۰	۰.۱	۶۸.۰۶%

لعابیران	۸۰۰۰	۰.۱	۸۸.۸۹%
مدیریت صنعت شوینده ت.ص.بهشهر	۹۰۰۰	۰.۱	۸۸.۲۱%
معدنی املاح ایران	۳۰۰۰	۰.۱۵	۷۲.۰۲%
نیروکلر	۱۰۰۰	۰.۱	۸۷.۴۱%
میانگین			۷۷.۹۱%

با توجه به داده‌های بالا کرنل معادله درجه ۲ با گاما scale می‌تواند به‌صورت میانگین تا ۷۷,۹۱ درصد جهت قیمت‌ها را درست پیش‌بینی کند.

جدول ۶-ارزیابی نتایج کرنل شعاعی با گاما auto

شرکت	تعداد ویژگی‌ها	تعداد آزمایش	کرنل شعاعی با گاما auto
بین‌المللی محصولات پارس	۱۲۰۰۰	۰.۱	۶۸.۰۶%
پاکسان	۱۰۰۰	۰.۱	۸۳.۶۶%
پتروشیمی آبادان	۹۰۰۰	۰.۱	۶۴.۲۹%
پتروشیمی پارس	۲۰۰۰	۰.۱	۶۱.۶۷%
پتروشیمی پردیس	۱۹۰۰۰	۰.۱	۱۰۰.۰۰%
پتروشیمی جم	۱۸۰۰۰	۰.۱	۹۱.۷۲%
پتروشیمی خارک	۱۶۰۰۰	۰.۳	۵۳.۳۹%
پتروشیمی شازند	۱۱۰۰۰	۰.۱	۷۱.۱۱%
پتروشیمی فناوران	۵۰۰۰	۰.۱	۷۱.۱۱%
پتروشیمی شیراز	۱۲۰۰۰	۰.۱۵	۷۸.۴۰%
س. صنایع‌شیمیایی‌ایران	۷۰۰۰	۰.۱	۸۷.۴۱%
س. نفت و گاز و پتروشیمی تأمین	۱۷۰۰۰	۰.۲۵	۸۲.۲۷%
سرمایه‌گذاری صنایع پتروشیمی	۵۰۰۰	۰.۱	۸۳.۶۶%
صنایع پتروشیمی خلیج فارس	۵۰۰۰	۰.۱	۸۸.۶۳%
صنایع پتروشیمی کرمانشاه	۱۰۰۰۰	۰.۱	۸۸.۶۳%
صنایع شیمیایی فارس	۱۷۰۰۰	۰.۲	۶۴.۰۷%
صنایع‌شیمیایی‌سینا	۳۰۰۰	۰.۱	۶۸.۰۶%
کربن ایران	۱۰۰۰۰	۰.۱	۸۷.۴۱%
گروه صنعتی پاکشو	۱۶۰۰۰	۰.۱	۸۰.۰۰%
گسترش نفت و گاز پارسیان	۱۱۰۰۰	۰.۱	۸۱.۸۲%
گلتاش	۸۰۰۰	۰.۱۵	۷۷.۵۶%
لعابیران	۸۰۰۰	۰.۱۵	۷۸.۲۳%
مدیریت صنعت شوینده ت.ص.بهشهر	۱۹۰۰۰	۰.۲۵	۶۱.۰۲%
معدنی املاح ایران	۱۶۰۰۰	۰.۲۵	۶۳.۳۳%
نیروکلر	۱۱۰۰۰	۰.۱	۸۳.۶۶%
میانگین			۷۶.۱۹%

با توجه به داده‌های بالا کرنل با گاما auto می‌تواند به‌صورت میانگین تا ۷۶,۱۹ درصد جهت قیمت‌ها را درست پیش‌بینی کند.

جدول ۷-ارزیابی نتایج کرنل شعاعی با گاما scale

شرکت	تعداد ویژگی‌ها	تعداد آزمایش	کرنل شعاعی با گاما scale
بین‌المللی محصولات پارس	۱۲۰۰۰	۰.۱	۷۷.۷۸%
پاکسان	۱۰۰۰۰	۰.۱	۸۳.۶۶%
پتروشیمی آبادان	۵۰۰۰	۰.۱۵	۷۶.۰۵%
پتروشیمی پارس	۸۰۰۰	۰.۱	۸۳.۶۶%
پتروشیمی پردیس	۱۱۰۰۰	۰.۱	۸۵.۲۶%
پتروشیمی جم	۱۸۰۰۰	۰.۱	۶۶.۸۹%
پتروشیمی خارک	۹۰۰۰	۰.۱	۶۹.۷۰%
پتروشیمی شازند	۱۸۰۰۰	۰.۱۵	۶۹.۱۴%
پتروشیمی فناوران	۱۰۰۰۰	۰.۱۵	۷۵.۵۱%
پتروشیمی شیراز	۵۰۰۰	۰.۱	۸۵.۲۶%
س. صنایع‌شیمیایی‌ایران	۱۰۰۰	۰.۱	۶۸.۰۶%
س. نفت و گاز و پتروشیمی تأمین	۹۰۰۰	۰.۱	۷۳.۶۴%
سرمایه‌گذاری صنایع پتروشیمی	۹۰۰۰	۰.۱	۷۳.۸۱%
صنایع پتروشیمی خلیج فارس	۴۰۰۰	۰.۱	۸۵.۲۶%
صنایع پتروشیمی کرمانشاه	۱۱۰۰۰	۰.۲	۷۴.۳۰%
صنایع شیمیایی فارس	۱۰۰۰	۰.۱	۵۵.۵۶%
صنایع‌شیمیایی‌سینا	۱۷۰۰۰	۰.۱	۶۸.۰۶%
کربن ایران	۱۳۰۰۰	۰.۱۵	۶۹.۱۴%
گروه صنعتی پاکشو	۱۰۰۰۰	۰.۱	۸۵.۲۶%
گسترش نفت و گاز	۱۴۰۰۰	۰.۱	۷۳.۶۴%

پارسیان			
گلتاش	۱۶۰۰۰	۰.۲	۷۴.۳۸%
لعابیران	۳۰۰۰	۰.۱۵	۵۵.۳۶%
مدیریت صنعت شوینده ت.ص.بهشهر	۱۰۰۰	۰.۱	۶۵.۸۰%
معدنی املاح ایران	۱۷۰۰۰	۰.۱	۷۷.۷۸%
نیروکلر	۴۰۰۰	۰.۲	۶۹.۴۴%
میانگین			۷۳.۷۰%

با توجه به داده‌های بالا کرنل شعاعی با گاما scale می‌تواند به‌صورت میانگین تا ۷۳.۷۰%درصد جهت قیمت‌ها را درست پیش‌بینی کند که نسبت به سایر کرنل‌ها نتیجه مناسبی محسوب نمی‌شود.

جدول ۸-ارزیابی نتایج کرنل سیگموئید با گاما auto

شرکت	تعداد ویژگی‌ها	تعداد آزمایش	کرنل سیگموئید با گاما auto
بین‌المللی محصولات پارس	۱۶۰۰۰	۰.۱	۸۳.۶۶%
پاکسان	۳۰۰۰	۰.۱۵	۸۸.۶۲%
پتروشیمی آبادان	۷۰۰۰	۰.۱	۶۴.۲۹%
پتروشیمی پارس	۷۰۰۰	۰.۱	۱۰۰.۰۰%
پتروشیمی پردیس	۱۷۰۰۰	۰.۱	۸۵.۲۶%
پتروشیمی جم	۱۳۰۰۰	۰.۱	۶۶.۸۹%
پتروشیمی خارک	۳۰۰۰	۰.۲	۳۶.۳۰%
پتروشیمی شازند	۶۰۰۰	۰.۱	۷۱.۱۱%
پتروشیمی فناوران	۸۰۰۰	۰.۱	۵۷.۶۵%
پتروشیمی شیراز	۲۰۰۰	۰.۱۵	۷۱.۱۱%
س. صنایع‌شیمیایی‌ایران	۲۰۰۰	۰.۱	۶۸.۰۶%
س. نفت و گاز و پتروشیمی تأمین	۱۱۰۰۰	۰.۱	۷۳.۶۴%
سرمایه‌گذاری صنایع پتروشیمی	۷۰۰۰	۰.۱۵	۷۹.۱۲%
صنایع پتروشیمی خلیج فارس	۱۶۰۰۰	۰.۱۵	۹۰.۱۱%
صنایع پتروشیمی کرمانشاه	۱۰۰۰	۰.۱۵	۷۱.۱۱%
صنایع شیمیایی فارس	۸۰۰۰	۰.۱	۳۹.۶۸%
صنایع‌شیمیایی‌سینا	۱۰۰۰	۰.۱	۶۸.۰۶%
کربن ایران	۶۰۰۰	۰.۱	۸۳.۶۶%
گروه صنعتی پاکشو	۵۰۰۰	۰.۱	۸۵.۲۶%
گلتاش	۴۰۰۰	۰.۱	۸۳.۶۶%
لعابیران	۳۰۰۰	۰.۱۵	۴۱.۵۶%
مدیریت صنعت شوینده ت.ص.بهشهر	۲۰۰۰	۰.۱۵	۴۱.۵۶%
معدنی املاح ایران	۱۰۰۰	۰.۲	۳۳.۳۳%
نیروکلر	۱۴۰۰۰	۰.۱	۸۳.۶۶%
میانگین			۶۹.۴۷%

با توجه به داده‌های بالا کرنل سیگموئید با گاما auto می‌تواند به‌صورت میانگین تا ۶۹,۴۷ درصد جهت قیمت‌ها را درست پیش‌بینی کند که نسبت به سایر الگوریتم‌ها نتایج نامناسب‌تری است.

جدول ۹-ارزیابی نتایج کرنل سیگموئید با گاما scale

شرکت	تعداد ویژگی‌ها	تعداد آزمایش	کرنل سیگموئید با گاما scale
بین‌المللی محصولات پارس	۵۰۰۰	۰.۱	۶۸.۰۶%
پاکسان	۷۰۰۰	۰.۱	۱۰۰.۰۰%
پتروشیمی آبادان	۱۰۰۰	۰.۱	۶۴.۲۹%
پتروشیمی پارس	۱۷۰۰۰	۰.۱۵	۶۹.۱۰%
پتروشیمی پردیس	۱۸۰۰۰	۰.۱	۸۵.۲۶%
پتروشیمی جم	۱۰۰۰	۰.۱۵	۶۶.۴۱%
پتروشیمی خارک	۱۷۰۰۰	۰.۲۵	۵۰.۳۱%
پتروشیمی شازند	۱۵۰۰۰	۰.۱	۷۱.۱۱%
پتروشیمی فناوران	۳۰۰۰	۰.۱	۵۷.۶۵%
پتروشیمی شیراز	۱۵۰۰۰	۰.۱	۸۵.۲۶%
س. صنایع‌شیمیایی‌ایران	۴۰۰۰	۰.۱	۶۸.۰۶%
س. نفت و گاز و پتروشیمی تأمین	۸۰۰۰	۰.۱	۷۳.۶۴%
سرمایه‌گذاری صنایع پتروشیمی	۱۰۰۰	۰.۱	۶۸.۰۶%
صنایع پتروشیمی خلیج فارس	۱۰۰۰	۰.۱	۸۵.۲۶%
صنایع پتروشیمی کرمانشاه	۱۰۰۰	۰.۱	۷۱.۱۱%

صنایع شیمیایی فارس	۱۸۰۰۰	۰.۲	۷۷.۲۰%
صنایع‌شیمیایی‌سینا	۴۰۰۰	۰.۱	۶۸.۰۶%
کربن ایران	۲۰۰۰	۰.۱	۸۳.۶۶%
گروه صنعتی پاکشو	۱۱۰۰۰	۰.۱	۸۵.۲۶%
گسترش نفت و گاز پارسیان	۶۰۰۰	۰.۱۵	۷۷.۰۸%
گلتاش	۱۰۰۰	۰.۱	۶۸.۰۶%
لعابیران	۱۹۰۰۰	۰.۲	۵۴.۴۰%
مدیریت صنعت شوینده ت.ص.بهشهر	۱۲۰۰۰	۰.۲۵	۵۶.۶۹%
معدنی املاح ایران	۶۰۰۰	۰.۱۵	۴۷.۵۰%
نیروکلر	۱۰۰۰	۰.۱۵	۷۹.۱۲%
میانگین			۷۱.۱۴%

با توجه به داده‌های بالا کرنل سیگموئید با گاما scale می‌تواند به‌صورت میانگین تا ۷۱,۱۴ درصد جهت قیمت‌ها را درست پیش‌بینی کند.

۴-نتیجه‌گیری و پیشنهادات آتی

گسترش روز افزون محیط وب و رشد تولید محتوای غیرساختار یافته شامل متن، صوت و فیلم و عدم توانایی انسان در بررسی همه آنها و تصمیم‌گیری، روز به روز اهمیت متن‌کاوی افزایش می‌یابد. هر چند نتایج حاصل از داده‌کاوی با درصدی خطا همراه است، اما در تعداد تصمیم‌گیری زیاد، مانند تشخیص نامه‌های الکترونیک اسپم استفاده از این روش‌ها ضروری به نظر می‌رسد. هر چند استفاده از الگوریتم‌های یادگیری ماشین در حجم دیتای زیاد نیاز به قدرت پردازش بالایی است، اما با گسترش تکنولوژی و تولید کامپیوتری‌ها با قدرت پردازش بالا استفاده از روش‌های داده‌کاوی عمومیت بیشتری پیدا می‌کند.

یکی از کاربردهای متن‌کاوی بررسی تأثیر اخبار بر قیمت سهام در بورس است. از آنجایی که اخبار زیادی هر روز توسط خبرگذاری‌ها منتشر می‌شود بررسی همه خبرها توسط انسان کار دشواری به نظر می‌رسد. خرید و فروش سهام همواره با سود و ضرر همراه است و ریسک وجود دارد، توانایی پیش‌بینی قیمت‌ها می‌تواند تأثیر قابل توجهی در میزان سود ما داشته باشد. در این پژوهش سعی شده است با ذخیره‌سازی اخبار و استفاده از الگوریتم ماشین بردار پشتیبان با کرنل‌های مختلف میزان دقت پیش‌بینی سهم ارزیابی شود.

با توجه به الگوریتم‌ها، کرنل‌ها و پارامترهای مختلفی که وجود دارد، نمی‌توان یک الگوریتم خاص با تنظیمات خاص را برای همه سهم‌ها انتخاب کرد، بلکه باید با توجه به نتایج حاصل‌شده از هر مدل، بهترین مدل را برای سهم موردنظر انتخاب کرد. با گذشت زمان و تغییر خبرها و شرایط یک سهم، ممکن است پارامترها و کرنل‌های نتایج مختلفی را ایجاد کنند. بهتر است هرروز مدل‌ها آموزش داده شوند و بهترین مدل برای پیش‌بینی جهت قیمت فردا انتخاب شود. از آنجایی که تعداد حالت‌های زیادی که اتفاق می‌افتد، تحلیل نتایج به‌صورت دستی بسیار زمان‌بر خواهد بود و برای انجام این وظایف بهتر است از کدنویسی و اتوماتیک کردن فرایند استفاده کرد.

در این پژوهش اخبار ۱۷ ذخیره شد و داده‌های معاملات شرکت‌های بورس در گروه محصولات شیمیایی ذخیره شد و ۲۵ شرکت در این پژوهش مورد بررسی قرار گرفت که در بازه آبان تا اسفند ۹۷ بیش از ۷۰ روز معاملاتی داشتند.

شکل ۲۱ - میانگین نتایج کرنل‌ها

با توجه به نتایج حاصل از ارزیابی مدل‌ها به‌صورت شکل ۲۱ خواهد بود. بهترین میانگین مربوط به کرنل معادله درجه ۲ با گاما auto است که با دقت تقریبی ۸۷ درصد می‌توان جهت قیمت سهم را در شرکت‌های گروه محصولات شیمیایی پیش‌بینی کند. همچنین مشاهده می‌شود در این پژوهش کرنل خطی هرچند ساده‌تر است و زمان پردازش کمتری را به خود اختصاص می‌دهد و سرعت بالاتری دارد، نتایج قابل قبولی در پیش‌بینی دارد و می‌تواند به صورت میانگین ۸۵٪ پیش‌بینی درستی انجام دهد و کرنل‌های غیرخطی هر چند سرعت پایین‌تری دارند ولی نتایج مناسبی را ارائه نمی‌کنند.

به دلیل محدودیت‌های زمانی و مکانی موجود این تحقیقات می‌تواند با انجام سایر مطالعات تکمیل‌تر شود که در زیر به آنها اشاره می‌شود.

۱- استفاده از بازه زمانی بیشتر برای ذخیره‌سازی اخبار و بررسی روزهای معاملاتی بیشتر در الگوریتم ماشین بردار پشتیبان و بررسی نتایج حاصل از روزهای معاملاتی بیشتر می‌تواند مورد بررسی قرار گیرد.

۲- در تحلیل‌های تکنیکال اخبار جدیدتر می‌توانند تاثیرگذاری بیشتری بر روی قیمت داشته باشند در صورتی که در این تحقیق فقط از خبرهای منتشر شده در روز قبل برای پیش‌بینی جهت قیمت استفاده شده است، در مطالعات آتی می‌توان از اخبار چند روز قبل و وزن دهی به روزهای گذشته نیز استفاده کرد.

۳- استفاده از سایر الگوریتم‌های کلاس‌بندی مانند درخت تصمیم، جنگل تصادفی و بیزین برای دسته و نتایج حاصل با مدل ماشین بردار پشتیبان مقایسه گردد.

۴- اجرای مدل پژوهش برای سایر شرکت‌ها و نتایج حاصل در گروه‌های محصولات مختلف بررسی و مقایسه شود.

۵- در این مطالعه فقط به بررسی جهت قیمت (افزایش یا کاهش قیمت) پرداخت شده است. در معاملات مقدار افزایش قیمت نیز بسیار مهم است و معامله‌گران تمایل دارند سهمی را خریداری کنند که قیمت آن درصد بیشتری افزایش پیدا کند. می‌توان در مطالعات آتی بر روی میزان افزایش قیمت نیز تحقیقاتی صورت بگیرد و نتایج الگوریتم‌ها با هم مقایسه شود.

منابع

1.J. D. Velásquez, V. Palade, and L. C. Jain, *Advanced techniques in web intelligence*: Springer, 2013.

2. Cisco. (2019). *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper*. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

3 .internetlivestats. (2019). *Total number of Websites*. Available: https://www.internetlivestats.com/total-number-of-websites/

4 .Z. Markov and D. T. Larose, *Data mining the Web :uncovering patterns in Web content, structure, and usage*: John Wiley & Sons, 2007.

5.B. Liu, *Web data mining: exploring hyperlinks, contents, and usage data*: Springer Science & Business Media, 2007.

6 A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining for market prediction: A systematic review," *Expert Systems with Applications*, vol. 41, pp. 7653-7670, 11/15/ 2014.

7. M.-A. Mittermayer and G. Knolmayer, *Text mining systems for market response to news: A survey*: Institut für Wirtschaftsinformatik der Universität Bern, 2006.

8 .C.-J. Huang, J.-J. Liao, D.-X. Yang, T.-Y. Chang, and Y.-C. Luo, "Realization of a news dissemination agent based on weighted association rules and text mining techniques," *Expert Systems with Applications*, vol. 37, pp. 6409-6413, 2010.

.۹B. S. Kumar and V. Ravi, "A survey of the applications of text mining in financial domain," *Knowledge-Based Systems*, vol. 114, pp. 128-147, 12/15/ 2016.

10. M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decision Support Systems*, vol. 55, pp. 685-697, 6// 2013.

11 .M. Thelwall, "Data cleansing and validation for multiple site link structure analysis," in *Web mining: Applications and techniques*, ed: IGI Global, 2005, pp. 208-227.

12 .M. Sheng, Y. Qin, L. Yao, and B. Benatallah, *Managing the web of things: linking the real world to the web*: Morgan Kaufmann, 2017.

13.R. Kosala and H. Blockeel, "Web mining research :A survey," *ACM Sigkdd Explorations Newsletter*, vol. 2, pp. 1-15, 2000.

14 .M. G. Da Costa and Z. Gong, "Web structure mining: an introduction," in *Information Acquisition, 2005 IEEE International Conference on*, 2005, p. 6 pp.

15 .F. Johnson and S. K. Gupta, "Web content mining techniques: a survey," *International Journal of Computer Applications*, vol. 47, 2012.

16 .Kumar and Ravi, "A survey of the applications of text mining in financial domain," vol. 114, pp. 128-147, 2016.

17.A. Hotho, A. Nürnberger ,and G. Paaß, "A brief survey of text mining," in *Ldv Forum*, 2005, pp. 19-62.

18.Gupta and Lehal, "A survey of text mining techniques and applications," vol. 1, pp. 60-76, 2009.
19.Y. Zhang, M. Chen, and L. Liu, "A review on text mining," in *Software Engineering and Service Science (ICSESS), 2015 6th IEEE International Conference on*, 2015, pp. 681-685.
20 .H. Hashimi, A. Hafez, and H. Mathkour, "Selection criteria for text mining approaches," *Computers in Human Behavior*, vol. 51, pp. 729-733, 2015.
21 .K .Javed, S. Maruf, and H. A. Babri, "A two-stage Markov blanket based feature selection algorithm for text classification," *Neurocomputing*, vol. 157, pp. 91-104, 2015.
22 .G. Hackeling, *Mastering Machine Learning with scikit-learn*: Packt Publishing Ltd, 2.۰۱۷
23 .Khan and A. Ahmad, "Cluster center initialization algorithm for K-means clustering," *Pattern recognition letters*, vol. 25, pp. 1293-1302, 2004.
24.J. Hou, H. Gao, and X. Li, "DSets-DBSCAN: a parameter-free clustering algorithm," *IEEE Transactions on Image Processing*, vol. 25, pp. 3182-3193, 2016.
25.Zhang and Z. Xu, "Hesitant fuzzy agglomerative hierarchical clustering algorithms," *International Journal of Systems Science*, vol. 46, pp. 562-576, 2015.
26 .D. M. Farid, L. Zhang, C. M. Rahman, M .A. Hossain, and R. Strachan, "Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks," *Expert Systems with Applications*, vol. 41, pp. 1937-1946, 2014.
27.V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B .P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling," *Journal of chemical information computer sciences*, vol. 43, pp. 1947-1958, 2003.
28 .Chen, K. Li, Z. Tang, K. Bilal, S. Yu, C. Weng, *et al*" ,.A parallel random forest algorithm for big data in a spark cloud computing environment," *IEEE Transactions on Parallel Distributed Systems*, pp. 1-1, 2017.
29.G. Dreyfus, *Neural networks: methodology and applications*: Springer Science & Business Media, 2.۰۰۵
30.C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining knowledge discovery*, vol. 2, pp. 121-167, 1998.
31 .M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation," in *Australasian joint conference on artificial intelligence*, 2006, pp. 1015-1021.
32.S. S. Groth and J. Muntermann, "An intraday market risk management approach

based on textual analysis," *Decision Support Systems* ,vol. 50, pp. 680-691, 2011.
33 .R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen, "Evaluating sentiment in financial news articles," *Decision Support Systems*, vol. 53, pp. 458-464, 6// 2012.
34 .L. Dey, A. Mahajan, and S. M. Haque, "Document clustering for event identification and trend analysis in market news," in *Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on*, 2009, pp. 103-106.
35.A. Mahajan, L. Dey, and S. M. Haque, "Mining Financial News for Major Events and Their Impacts on the Market," in *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2008, pp. 423-426.
36 .X. Zhong and D. Enke, "A comprehensive cluster and classification mining procedure for daily stock market return forecasting," *Neurocomputing*, vol. 267, pp. 152-168, 2017/12/06/ 2017.
37 .A. E. Khedr, S. Salama, and N. Yaseen, "Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis," *International Journal of Intelligent Systems and Applications (IJISA)*, vol. 9, pp. 22-30, 2017.
38 .H. Levy and M. Sarnat, "International Diversification of Investment Portfolios," *The American Economic Review*, vol. 60, pp. 668-675, 1970.
39 .J. M.-T. Wu, Z. Li, C.-W. Lin, and M. Pirouz, "A New Convolution Neural Network Model for Stock Price Prediction," ed, 2020, pp. 581-585.
40 .O. M. Ebadati E and M. Mortazavi T, "An efficient hybrid machine learning method for time series stock market forecasting," *Neural Network World*, vol. 28, pp. 41.۲۰۱۸ ,۵۵-
41.A. Mahajan, L. Dey, and S. M. Haque, "Mining financial news for major events and their impacts on the market," in *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on*, 2008, pp. 423-42.۶
42.X. Zhong and D. J. N. Enke, "A comprehensive cluster and classification mining procedure for daily stock market return forecasting," vol. 267, pp. 152-168, 2017.
43.A. E. Khedr and N. Yaseen, "Predicting stock market behavior using data mining technique and news sentiment analysis," *International Journal of Intelligent Systems Applications*, vol. 9, p. 22, 2017.
44 .M. Hagenau, M. Liebmann, and D. J. D. S. S. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," vol. 55, pp. 685-697, 2013.
45 .scikit-learn.org. *Choosing the right estimator*. Available: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html
۴۶. ع. نوریان. (۲۰۱۸). هضم برای پردازش زبان فارسی در پایتون. Available: http://www.sobhe.ir/hazm/
47.W.-H. Chen, S.-H. Hsu, and H.-P. Shen, "Application of SVM and ANN for intrusion detection," *Computers Operations Research*, vol. 32, pp. 2617-2634, 2005.

سال یازدهم، شماره‌های ۳۹و۴۰، بهار و تابستان ۱۳۹۸
صص: ۴۹-۷۲

فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

بهبود مسیریابی جهت کنترل ازدحام در شبکه‌هاب مبتنی بر نرم‌افزار با استفاده از کنترلرهای توزیع‌شده

[*] سعید بختیاری [**] اردشیر آذرنژاد
* استادیار دانشگاه علوم انتظامی امین، تهران
**کارشناسی‌ارشد فناوری اطلاعات دانشگاه آزاد اسلامی واحد تهران مرکزی
تاریخ دریافت: ۱۳۹۸ /۰۴/۱۳ تاریخ پذیرش: ۱۳۹۹/۰۳/۲۴

چکیده

شبکه‌های مبتنی بر نرم‌افزار (SDN) برای استفاده در تعیین مسیریابی ترافیک شبکه قابل انعطاف هستند، زیرا سطح داده‌ای و سطح کنترلی را از یکدیگر تفکیک می‌کنند. یکی از چالش‌های بزرگی که پیش روی شبکه‌های مبتنی بر نرم‌افزار قرار گرفته است، انتخاب مکان‌هایی مناسب برای قراردادن و توزیع کنترلرها (کنترل‌کننده‌ها) است؛ به گونه‌ای که بتوان تأخیر بین کنترلرها و سوئیچ‌ها را در شبکه‌های گسترده کاهش داد. در همین راستا اغلب روش‌های ارائه شده بر روی کاهش تأخیر متمرکز بوده‌اند. ولی تأخیر تنها یکی از عواملی است که در کارائی شبکه و کاهش هزینه‌ی کلی بین کنترلرها و سوئیچ‌های مرتبط با آن‌ها نقش دارد. این مقاله به بررسی عوامل بیشتری برای کاهش هزینه بین کنترلرها و سوئیچ‌ها نظیر ترافیک لینک‌های ارتباطی می‌پردازد. به همین منظور یک الگوریتم مبتنی برخوشه‌بندی برای بخش‌بندی شبکه ارائه می‌شود. با بهره گیری از این الگوریتم می‌توان تضمین کرد که هر بخش از شبکه می‌تواند حداکثر هزینه (شامل تأخیر و ترافیک موجود روی لینک‌ها) را در بین کنترلر و سوئیچ‌های مربوط به آن کاهش دهد. در این مقاله، با بکارگیری از Topology Zoo، شبیه‌سازی‌های گسترده‌ای تحت توپولوژی‌های واقعی شبکه انجام شده است. نتایج شبیه‌سازی‌ها نشان می‌دهد در شرایطی که احتمال ازدحام در شبکه بالا می‌رود، الگوریتم پیشنهادی با شناسایی لینک‌های گلوگاه در مسیرهای ارتباطی هر گره با سایر گره‌ها، توانسته به خوبی ازدحام را در شبکه کنترل نماید. لذا، با در نظر گرفتن دو معیار تأخیر و میزان مشغول بودن لینک‌ها، فرآیند قرارگیری و توزیع کنترلرها را در عمل خوشه‌بندی با دقت بالاتری انجام می‌دهد. با این کار، میانگین حداکثر هزینه‌ی انتها به انتها بین هر کنترلر و سوئیچ‌های مربوط به آن به ترتیب در توپولوژی‌های Chinanet کشور چین، Uunet کشور آمریکا، DFN کشور آلمان، و Rediris کشور اسپانیا به اندازه‌ی ۴۱/۴۶۹۴، ۲۹/۲۸۵۳، ۲۱/۳۸۰۵ و ۴۶/۴۸۲۹ درصد کاهش یافته است.

واژه‌های کلیدی: شبکه‌های مبتنی بر نرم‌افزار، کنترلرهای توزیع‌شده، قراردادن کنترلرها، خوشه‌بندی، کنترل ازدحام.

نویسنده عهده‌دار مکاتبات : سعید بختیاری saeid.bakhtiarii@chmail.ir

۱- مقدمه

افزایش فناوری‌های رایانشی جدید اعم از کلان داده‌ها[1]، اینترنت اشیاء[2]، و رایانش ابری[3] باعث شده تا تغییرات قابل ملاحظه‌ای در روش ذخیره‌سازی، گردآوری و انتقال اطلاعات و داده‌ها صورت گیرد. در کنار این روندهای رایانشی جدید، چالش‌های جدیدی اعم از مدیریت و ارتقای شبکه، استفاده‌ی بهینه از منابع، انتقال سریع داده‌ها نیز پدید آمده است. به منظور غلبه بر این چالش ها، از شبکه-های مبتنی بر نرم‌افزار (SDN)[4] به عنوان یک الگوی مطلوب برای شبکه‌های نسل آینده استفاده می‌شود. شبکه-های نرم‌افزار محور در مقایسه با شبکه‌های متعارف از اصل تفکیک سطح کنترلی[5] و سطح داده‌ای[6] بهره می‌برد. این بدین معنا است که سطح کنترلی در این شبکه‌ها به وسیله‌ی یک مجموعه از کنترلرهای[7] اختصاصی شکل گرفته و هر کدام از آن‌ها می‌توانند یک یا چند سوئیچ که وظیفه‌ی هدایت بسته‌های داده‌ای به سمت جلو را بر عهده دارند را مدیریت نمایند. در نتیجه توابع کنترلی و مدیریتی در این شبکه‌ها به گونه‌ای طراحی می‌شوند که خدمات شبکه و برنامه‌های کاربردی، از زیرساختار زیرین آن مجزا باشند.

در شبکه‌های نرم‌افزار محور، کنترلرِ مبتنی بر نرم‌افزار نقش یک واسط هوشمند شبکه را بازی می‌کند و سوئیچ‌های شبکه نیز نقش دستگاه‌های ساده‌ای برای هدایت و فوروارد بسته‌های داده‌ای به سمت جلو را بر عهده دارند که البته می‌توان این فرآیند فوروارد را از طریق واسط‌های بازی مانند OpenFlow برنامه نویسی کرد [۱]. در صورتی که یک بسته وارد یک سوئیچ شده ولی جدول جریان در این سوئیچ هیچ تطابقی با الگوی این بسته نداشته باشد، سوئیچ اقدام به ایجاد یک بسته‌ی درخواست نموده و آن را برای کنترلر مربوطه ارسال می‌کند. کنترلر پس از دریافت این درخواست اقدام به ارسال یک سیاست فوروارد جدید نموده و بر همین اساس سوئیچ اقدام به بروز رسانی جدول جریان خود می‌کند. پس از آن، بسته‌ی داده‌ای بر مبنای جدول جریان که بروز شده است تحویل داده می‌شود.

همان‌طور که در این مدل تشریح شده است، همه‌ی توابع موجود در SDN از طریق مبادله‌ی مکرر پیام ها در بین کنترلر ها و سوئیچ‌ها صورت می‌گیرد. بنابراین، موقعیت کنترلرها نقش قابل ملاحظه‌ای در مبادله‌ی پیام داشته و از این رو بر روی کارائی SDN تأثیر می‌گذارد. مسأله‌ی تعیین موقعیت‌هایی مناسب برای قرار دادن و توزیع کنترلرها را مسأله‌ی گمارش کنترلرها[8] می‌گویند. در این مسأله به دنبال آن هستیم که چطور می‌توان کنترلرها را در یک شبکه‌ی مبتنی بر نرم‌افزار قرار داده و سوئیچ‌های مربوطه را به گونه‌ای به این کنترلرها تخصیص داد تا به هدف مد نظر برسیم. گمارش کنترلرها اهمیت بسیاری در شبکه‌های گسترده[9] دارد که دلیل آن را می‌توان ناشی از توپولوژی نامتعارف و تأخیر بالا در انتشار بسته‌های داده‌ای دانست.

به منظور حل مسأله‌ی گمارش کنترلرها در شبکه‌های نرم‌افزار محور، معیارهای مختلفی ارائه شده است. در بین این معیارها، تأخیر در بین کنترلر و شبکه نقش بسیار مهمی را بازی می‌کند چرا که تأثیر قابل ملاحظه‌ای بر روی کارائی کلی SDN دارد [۲] [۳]. برای مثال در صورتی که یک بسته‌ی داده‌ای هیچ تطابقی با الگوهای تعریف شده در جدول جریان یک سوئیچ نداشته باشد، نیاز است تا این سوئیچ برای یک بازه‌ی زمانی منتظر مانده تا جدول‌های جریان را قبل از پردازش این بسته‌ی داده‌ای دریافت نماید. بدیهی است که یک زمان انتظار طولانی می‌تواند افت کلی بازدهی SDN را به همراه داشته باشد، یعنی برنامه‌های کاربردی[10] که نسبت به زمان حساس هستند با کندی روبرو شده و وظایفی که باید به صورت بلادرنگ صورت گیرند، به هیچ وجه انجام نخواهند شد. به طور معمول، لینک‌های ارتباطی که پهنای باند محدودی دارند و ترافیک بالای شبکه باعث شده تا شبکه دچار ازدحام شده و تأخیر بالایی پیش روی شبکه قرار گیرد. با توجه به ویژگی انحصاری که در SDN وجود دارد، تأخیر ناشی از ازدحام بین کنترلرها و سوئیچ‌ها در شبکه‌های SDN بسیار محدود خواهد بود. چرا که سطح کنترلی از سطح داده‌ای جدا بوده و از این رو پیام‌های کنترلی که بین کنترلرها و سوئیچ‌ها مبادله می‌شود از طریق یک کانال اختصاصی (مد برون باند) منتقل می‌شود [۴] [۵]. علاوه بر این، پیام‌های کنترلی، شامل یک سری

[1] Big Data
[2] Internet of Things (IoT)
[3] Cloud Computing
[4] Software Defined Networking (SDN)
[5] Control Plane
[6] Data Plane
[7] Controller
[8] Controller Placement Problem (CPP)
[9] Wide Area Networks (WAN)
[10] Application

جریان داده‌ای سبک وزن در مقایسه با بارهای داده‌ای موجود در سطح داده‌ای می‌باشند [۶]. در کنار معیار تأخیر، باید بار موجود و درصد مشغول بودن لینک‌های ارتباطی را هم در نظر بگیریم تا از به وجود آمدن نقاط گلوگاه در شبکه و در نتیجه به وجود آمدن ازدحام جلوگیری به عمل آوریم.

در این مقاله به بررسی هزینه انتها به انتها (شامل تأخیر انتها به انتها و درصد مشغول بودن لینک‌ها) در بین کنترلرها و سوئیچ‌ها می‌پردازیم و به دنبال راهکاری برای کاهش هزینه و کنترل ازدحام در شبکه‌های گسترده خواهیم بود. هر کدام از این ها را به صورت مجزا مورد بحث قرار می‌دهیم. در ابتدا باید یک شبکه را به چندین زیر شبکه تقسیم بندی کرده تا بتوان هزینه انتها به انتها در بین کنترلرها و سوئیچ‌ها را کاهش داد. با در نظر گرفتن اهمیت تأخیر و هم‌چنین بار موجود روی لینک‌های ارتباطی، روش پیشنهادی در این مقاله ارائه شده است. باید حتماً در نظر داشته باشیم که تأخیر انتها به انتها تنها یکی از مؤلفه‌های تأثیر گذار بر روی کارائی شبکه می‌باشد. علاوه بر تأخیر عواملی هم‌چون بار موجود روی لینک‌های ارتباطی بین کنترلرها و سوئیچ‌های مربوط به آن‌ها نیز در کارائی کلی شبکه تاثیرگذار می‌باشد.

برای این منظور، یک الگوریتم مبتنی بر خوشه‌بندی برای تقسیم‌بندی شبکه ارائه شده است. در این الگوریتم این اطمینان داده می‌شود که هر بخش از شبکه می‌تواند هزینه حداکثری در بین کنترلر و سوئیچ‌های مربوطه را کاهش دهد. در ادامه، در بخش دوم به بررسی مطالعات مربوطه و تعاریف پایه خواهیم پرداخت. در بخش سوم به بیان ریاضی مسأله‌ی گمارش و توزیع کنترلرها می‌پردازیم. در بخش چهارم، راهکار جدید پیشنهادی را به صورت کامل تشریح خواهیم نمود و به تحلیل کارائی نتایج خواهیم پرداخت.

۲- شبکه‌های مبتنی بر نرم‌افزار

SDN یک معماری جدید برای شبکه‌های کامپیوتری است که طی آن کنترل اطلاعات از خود اطلاعات و انتقال اطلاعات، مجزا می‌شود. روترها و سوئیچ‌های کنونی شبکه‌ها هرچقدر که پیشرفته و قدرتمند باشند عملیات انتقال و کنترل اطلاعات را با هم انجام می‌دهند. در معماری SDN کنترل اطلاعات از سخت‌افزار سوئیچ و روتر مجزا شده و به یک لایه بالاتر رفته و توسط نرم‌افزار انجام می‌شود.

با مجزا شدن سطح کنترلی از سطح داده‌ای در سخت‌افزارها، شرکت‌ها می‌توانند نرم‌افزارها و ابزارهای زیادی برای کنترل اطلاعات نوشته و در نتیجه سرعت، انعطاف‌پذیری[11]، مقیاس‌پذیری[12]، دسترس‌پذیری[13]، و قابلیت اعتماد[14] شبکه را بیشتر کنند. شرکت‌های مختلف می‌توانند برای سخت افزارهایی با برندهای مختلف رابط‌های برنامه‌نویسی کاربردی[15] (API) بنویسند که قابلیت ها و امکانات بیشتری برای شبکه به همراه دارند و مدیریت شبکه را متمرکز و یکپارچه می‌کنند و همچنین امنیت شبکه بالاتر می رود زیرا کاربران با نوشتن نرم‌افزارهایی می‌توانند مدیریت و مانیتورینگ بهتر و بیشتری روی اطلاعات داشته باشند و بر اساس نیازهای شبکه و تهدیداتی که متوجه شبکه آنها است، فایروال ها و سیستم های کشف فیلترینگ را برنامه ریزی و سیاستگذاری کنند. مزیت دیگر، پیکربندی مجدد شبکه و سخت‌افزار بدون نیاز به شرکت سازنده آن سخت‌افزار است. در شبکه‌های کنونی کاربران محدود به استفاده از فناوری و معماری ارائه شده توسط شرکت‌های سازنده سخت افزار هستند و نمی‌توانند خودشان دست به توسعه شبکه بزنند. برنامه‌نویسی رابط شبکه در SDN توسط خود کاربر صورت می‌گیرد و مطابق با نیازهای او می‌تواند بومی‌سازی شود. استاندارد SDN به گونه ای طراحی شده است که فرآیند ارسال اطلاعات در شبکه ها را آسان تر و انعطاف پذیری شبکه‌ها را تحت فضای برنامه‌ریزی شده هوشمند بیشتر می-کند.

به منظور بهبود مسیریابی در شبکه‌های مبتنی بر نرم‌افزار با استفاده از کنترلرهای توزیع شده، ابتدا برخی از مفاهیم مربوط به آن را بیان می‌نماییم و سپس به بررسی پژوهش-های اخیر صورت پذیرفته در این حوزه می‌پردازیم.

مفهوم کنترلر در شبکه‌های مبتنی بر نرم‌افزار: شبکه‌های مبتنی بر نرم‌افزار سعی دارد هوشمندی شبکه‌ها را بیشتر کرده و با انتقال بخش کنترل داده‌ها از سوئیچ و روتر سخت‌افزاری به لایه‌های نرم‌افزاریِ مجازی شبکه و بهره‌گیری از یک واحد نرم‌افزاری متمرکز، قابلیت‌هایی مانند برنامه‌ریزی، مقیاس‌پذیری، انعطاف‌پذیری، خودکارسازی، هوشمندی و توسعه نرم‌افزاری شبکه توسط سازمان‌ها را

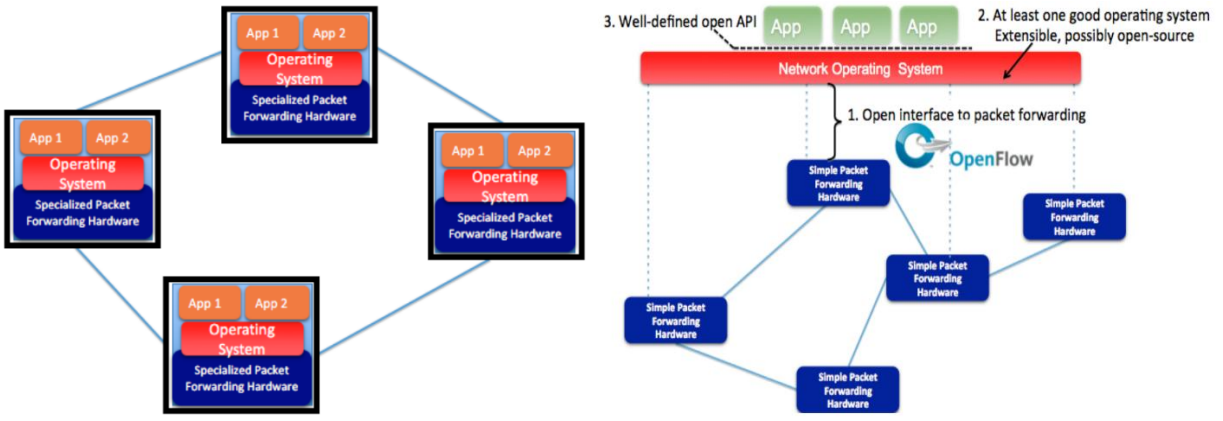
[11] Flexibility
[12] Scalability
[13] Availability
[14] Reliability
[15] Application Programming Interface

فراهم کند [۷]. این واحد متمرکز، کنترلر یا همان کنترل-کننده نام دارد و باید حتماً تأکید کنیم که منظور از متمرکز بودن به معنی منطقاً متمرکز می‌باشد، به این معنی که کنترلرها را می‌توانیم به صورت فیزیکی و آن هم توزیع‌شده، در سطح شبکه قرار دهیم.

معماری شبکه‌های مبتنی بر نرم‌افزار: مدیریت و کنترل شبکه‌های بزرگ همیشه دردسرهای مخصوص به خود را دارد. یکی از آسان‌ترین روش‌های پیشگیری از بروز مشکلات و پیچیدگی‌های مدیریت شبکه‌های بزرگ استفاده از محصولات یک تولیدکننده در تمامی قسمت‌های شبکه مورد نظر است. اتکا به یک تولیدکننده، علاوه بر تحمیل هزینه‌های بیشتر (به خاطر محدودیت‌های مربوط به لایسنس، حق نام، و...) می‌تواند خلاقیت را از سازمان‌ها و شرکت‌ها دور کند.

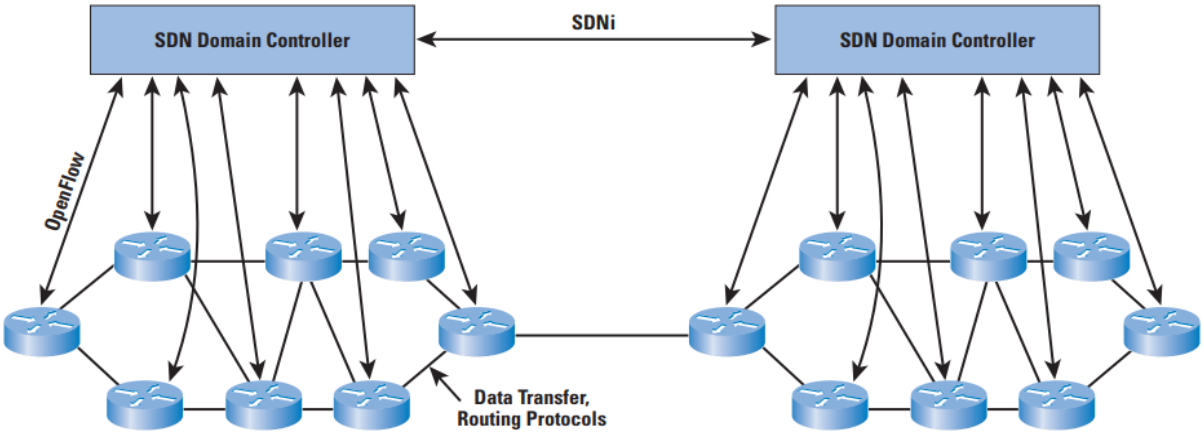
شبکه‌های امروزی شامل کاربرانی است که بوسیله سوئیچ ها و روترها با یکدیگر ارتباط یافته‌اند. این تجهیزات به صورت دستگاه‌هایی عرضه می‌شود که سخت‌افزار، سیستم‌عامل و نرم‌افزار توسط تولیدکننده به صورت یکپارچه در آنها تعبیه شده و تغییر در سیستم‌عامل تقریبا امکان‌پذیر نیست. از این‌رو منطق معماری این تجهیزات را عمودی می‌نامند. در واقع در ساختارهای فعلی، در شبکه‌های بزرگ سوئیچ ها، روترها و سایر تجهیزات شبکه، هم داده و هم اطلاعات کنترلی را در بر دارند که کار بهینه‌سازی ساختار شبکه را بسیار مشکل می‌سازد.

اما تجهیزاتی که برای SDN و استفاده از OpenFlow تولید می‌شوند، از منطق معماری افقی پیروی می‌کنند. در این معماری، دیگر از دستگاه‌هایی یکپارچه خبری نیست و تولیدکننده امکان استفاده از سیستم‌عامل و نرم‌افزار دلخواه مشتری را روی سخت‌افزار تولید شده فراهم می‌کند تا بتوان به‌طور سفارشی از سخت‌افزار بهره جست. در واقع از دیدگاه شبکه می‌توان گفت، قابلیت مدیریت دلخواه چند Control Plane مختلف و استفاده از نرم‌افزارهای کاربردی مجزا روی این تجهیزات فراهم می‌شود. در SDN، داده‌ها و اطلاعات کنترلی تجهیزات شبکه مانند سوئیچ ها و روترها، توسط یک API جدا می‌شوند. در شکل ۱ مقایسه‌ای از معماری عمودی تجهیزات فعلی شبکه در مقابل معماری افقی تجهیزات شبکه SDN نمایش داده شده است.



شکل ۱- معماری عمودی تجهیزات فعلی شبکه در مقابل معماری افقی تجهیزات شبکه SDN [۸]

دامنه‌های موجود در شبکه‌های مبتنی بر نرم‌افزار: در یک شبکه بزرگ، استقرار یک کنترلر واحد برای مدیریت تمام دستگاه‌های شبکه کار مناسبی نیست. یک سناریو محتمل‌تر این است که مدیر شبکه، شبکه را همانطور که در شکل ۲ نشان داده شده است، به تعدادی دامنه SDN که همپوشانی ندارند تقسیم کند.



شکل ۲-تقسیم‌بندی شبکه SDN به دامنه‌های مختلف [۹]

وجود چندین دامنه نیازمند ایجاد کنترلرهای تکی برای برقراری ارتباط با یکدیگر از طریق یک پروتکل استاندارد برای تبادل اطلاعات مسیریابی می‌باشد. برای این منظور IETF اخیراً بر روی توسعه و گسترش پروتکلی به نام SDNi[۱۶] کار می‌کند.

۳- کارهای مرتبط

در این بخش به بررسی تحقیقاتی که بر روی مسأله‌ی گمارش کنترلرها صورت گرفته است می‌پردازیم. تحقیقات موجود را می‌توان بر حسب اهداف زیر به چهار بخش تقسیم کرد [۱۰]: کاهش تأخیر شبکه در بین کنترلرها و سوئیچ‌ها، افزایش قابلیت اطمینان و اعتماد، کاهش هزینه‌ی توسعه و مصرف انرژی و روش چند هدفی.

کاهش تأخیر شبکه در بین کنترلرها و سوئیچ‌ها: آقای هِلر و همکارانش [۱۱] به مطالعه‌ی مسأله‌ی گمارش کنترلرها در SDN پرداختند و خاطر نشان کردند که تأخیر در انتشار (تأخیر میانگین و تأخیر در بدترین حالت)، ملاحظه‌ی اصلی آن‌ها در این مطالعه می‌باشد. این مسأله به صورت یک مسأله‌ی سهولت در مکان‌یابی شبیه‌سازی شده است و از الگوریتم K-center نیز برای حل این مسأله استفاده شده است. آقای یائو و همکارانش [۱۲] نیز هر دوی تأخیر در انتظار و ظرفیت کنترلر را در نظر گرفتند. بنابراین فرآیند گمارش را می‌توان به عنوان یک مسأله‌ی K-center در نظر گرفت [۱۳]. نتایج شبیه‌سازی‌های آن‌ها نشان می‌دهد که راهکار پیشنهادی آن‌ها می‌تواند تعداد کنترلرها و بار کنترلرهای درگیر را کاهش دهد. در [۱۴] به ایجاد یک توازن در بین تأخیر در انتشار و بار ترافیکی پرداخته شده است. نتایج شبیه‌سازی نشان می‌دهد که بار موجود در سطح کنترلی را می‌توان با افزایش جزئی تأخیر متوازن‌سازی کرد. در [۱۵] به بررسی مسأله‌ی گمارش کنترلرها و آن هم از یک دید متفاوت پرداخته شده است. بر مبنای این دیدگاه، بجای کمینه‌سازی مستقیم تأخیر بین کنترلرها و سوئیچ ها، هدف این است که تعداد سوئیچ‌های کنترلی را در یک بازه‌ی زمانی که با تأخیر روبرو هستیم افزایش دهیم.

افزایش قابلیت اطمینان و تاب‌آوری: آقای ژانگ و همکارانش [۱۶] به ارائه‌ی یک الگوریتم مبتنی بر برش کمینه[۱۷] پرداخته‌اند تا بتوانند احتمال قطعی ارتباط در بین کنترلر و سوئیچ را به حداقل سطح ممکن برسانند. آقای هیو و همکارانش [۱۷] نیز به مطالعه‌ی مسأله‌ی گمارش کنترلر برای تضمین قابلیت اطمینان پرداخته‌اند و یک الگوریتم حریصانه را برای گمارش کارآمد و رسیدن به قابلیت اطمینان ارائه نموده‌اند. در [۱۸] هدفی که دنبال شده است این بوده که ارتباط بین دستگاه‌های هدایت‌کننده‌ی بسته‌های داده‌ای و کنترلرها به بالاترین سطح ممکن برسد. در این مطالعه خاطر نشان شده است که تنوع در مسیرها می‌تواند بقای شبکه را در شرایط بروز اشکال و خرابی افزایش دهد. تحقیقات مشابهی نیز صورت گرفته که می‌توان به [۱۹] اشاره کرد. شبیه‌سازی‌هایی نیز در [۲۰] و با هدف افزایش قابلیت اطمینان SDN ارائه شده است. در این مطالعه به گمارش مناسب کنترلرها پرداخته شده است. جدای از تحقیقاتی که به بهبود تاب‌آوری اتصالات شبکه

[16] Interfacing SDN Domain Controllers

[17] Min-cut algorithm

پرداخته‌اند، در [۲۱] روشی ارائه شده است که می‌تواند اشکالات کنترلر در شبکه‌های گسترده‌ی SDN را تحمل نماید.

کاهش هزینه‌ی توسعه و مصرف انرژی: پژوهشگران تحقیق [۲۲] به توسعه‌ی مدلی بهینه پرداخته‌اند و مصرف کلی انرژی را بر اساس شبیه‌سازی‌هایشان کاهش داده‌اند. چالش اصلی در این مدل، پیچیدگی بالای آن می‌باشد چرا که بعضی از محاسبات را نمی‌توان در کمتر از ۳۰ ساعت انجام داد. آن‌ها در این کار شرایط قرارگیری را برای بسط دادن یک شبکه مبتنی بر نرم‌افزار با کمترین هزینه مورد بررسی قرار دادند. بدین صورت که با داشتن طراحی یک شبکه موجود و تعدادی سوئیچ جدید که باید به شبکه قبلی اضافه شوند، مدل بررسی می‌کند که شبکه چگونه مجدداً سازماندهی شود تا هزینه بروزرسانی کمینه شود. از آنجایی که مسأله بسط دادن، کلی شده مسأله برنامه‌ریزی است، این مدل می‌تواند برای برنامه‌ریزی یک شبکه از ابتدا نیز مورد استفاده قرار بگیرد. آن‌ها برای حل این مسأله از یک راه‌حل برنامه‌ریزی خطی استفاده کردند. در [۲۳] راهکاری ارائه شده است که می‌تواند به صورت پویا به اضافه و یا حذف کنترلرها و آن هم بر اساس تغییر بارها بپردازد. یکی از روش‌ها برای به حداقل رساندن مصرف انرژی، روش GreCo می‌باشد که در [۲۴] ارائه شده است. در این روش، لینک‌های غیر ضروری خاموش شده و در عین حال حفظ ارتباط در بین سوئیچ‌ها و کنترلرها تضمین می‌شود. با توجه به شبیه‌سازی‌ها، با روش GreCo می‌توان تا ۵۵ درصد مصرف انرژی را در طی ساعات شلوغی و اوج کاهش داد. روش دیگری در [۲۵] ارائه شده که در آن در ابتدا تعداد کنترلرها مشخص شده و سپس کنترلرهای ضروری فعال می‌شوند تا به این شکل در مصرف انرژی صرفه‌جویی شود. یک مدل دیگر در [۲۶] برای به حداقل رساندن هزینه‌ی بروز رسانی توسعه یافته است و آن هم در زمانی که سوئیچ‌های جدید به شبکه اضافه می‌شوند. با توجه به اینکه این روش در شبکه‌های مبتنی بر نرم‌افزار محدود نمی‌باشد، می‌توان از آن برای ایجاد یک شبکه‌ی جدید و یا بروز رسانی شبکه‌های موجود استفاده نمود.

روش چند هدفی: در [۲۷]، مباحثی در خصوص مسأله‌ی گمارش کنترلرها و آن هم با تمرکز بر تاب‌آوری و اشکال در شبکه‌های هسته‌ای مبتنی بر نرم‌افزار ارائه شده است. یک مجموعه‌ی ابزاری با نام POCO ارائه شده که این امکان را به اپراتورهای شبکه داده تا گمارش را به صورت شبه بهینه انجام دهند. یافته‌های آن‌ها نشان می‌دهد که در اغلب توپولوژی‌ها، بیش از ۲۰ درصد از همه‌ی گره‌ها باید از نوع کنترلر بوده تا ارتباط پیوسته‌ی همه‌ی گره‌ها با یکی از کنترلرها و آن هم در سناریوهای خرابی تضمین شود. این مجموعه ابزار در [۲۸] به صورت توسعه‌یافته تر ارائه شده است به گونه‌ای که در آن از یک روش ابتکاری با دقت کمتر استفاده شده ولی زمان محاسباتی کمتری برای غلبه بر ماهیت پویای شبکه‌های بزرگ مقیاس دارد. ایجاد توازن در بین زمان و میزان صحت نیز به صورت کامل از طریق توپولوژی‌های مختلفی بررسی شده است. پژوهشگران در [۲۹] به ارائه‌ی یک چارچوب کنترل و مدیریت مبتنی بر نرم‌افزار و آن هم برای شبکه‌هایی با ستون فقرات ثابت پرداخته‌اند. الگوریتم‌هایی برای تخصیص مدیران و کنترلرها در یک لایه‌ی کنترلی ارائه شده که هدف آن رسیدن به یک توازن بار و مصرف انرژی می‌باشد. در [۳۰]، یک روش گمارش کنترلر و مبتنی بر چگالی ارائه شده است که مسأله‌ی گمارش کنترلر بر حسب تأخیر، تحمل‌پذیری در برابر اشکال و تعداد کنترلرها، ارائه شده است. ارزیابی‌های کارائی نشان می‌دهد که این روش می‌تواند به کارائی مطلوبی دست پیدا کرده و سرعت اجرا را نیز در عین حال کاهش می‌دهد. آن‌ها برای حل مسأله قرارگیری کنترلرها به چگالی گره‌ها در توپولوژی شبکه توجه کردند. بدین صورت که در ابتدا بر اساس چگالی گره‌ها، شبکه را به تعدادی خوشه‌ی مجزا تقسیم می‌کنند و به دلیل آنکه گره‌های درون هر خوشه اتصال قوی با دیگر گره‌های درون خوشه خود دارند و اتصال آن‌ها با گره‌های دیگر خوشه‌ها ضعیف است، در هر خوشه یک کنترل‌کننده قرار می‌گیرد. ژانگ و همکاران در [۳۱] به تدوین مسأله‌ی گمارش کنترلرها بر روی یک مدل چند هدفی پرداخته‌اند که هدف آن افزایش قابلیت اطمینان در بین کنترلرها و سوئیچ‌ها می‌باشد. نتایج شبیه‌سازی‌ها نشان می‌دهد که روش پیشنهادی می‌تواند کارائی شبکه را بهبود داده و به توان مطلوبی در بین این اهداف دست پیدا کند.

با توجه به بررسی‌ها، بسیاری از روش‌هایی که در بالا مطرح شد تنها تأخیر در انتشار بسته در بین کنترلرها و سوئیچ‌ها را در نظر می‌گیرد تا بتواند تأخیر را به حداقل سطح ممکن

برساند. البته تأخیر در انتشار تنها یکی از عوامل تأثیرگذار بر روی تأخیر سراسری می‌باشد. از جمله عوامل دیگر می‌توان به تأخیر در صف‌بندی کنترلرها اشاره کرد. بنابراین نیاز به تحقیقات گسترده‌ای برای لحاظ کردن همه‌ی مؤلفه‌ها می‌باشد. در مقاله‌ی [۳۲] که به عنوان مقاله‌ی پایه در این پژوهش مورد استفاده قرار گرفته است، همه‌ی تأخیرهای ممکن در بین کنترلرها و سوئیچ‌ها را در نظر گرفته و راهکاری را برای کاهش تأخیرها ارائه نموده است. علاوه بر مقالات و مطالعاتی که در بالا به آن‌ها اشاره شد، می‌توان به موارد زیر نیز در حوزه موضوع پژوهش اشاره کرد:

یک کار دیگر در خصوص قرارگیری کنترلرها توسط بَری و همکارانش انجام شده است [۳۳]. آنها دو الگوریتم مکاشفه-ای برای تأمین کنترلر به صورت پویا پیشنهاد دادند. اهداف آن‌ها شامل کمینه کردن زمان تنظیم جریان، ترافیک کنترلی و تخصیص مجدد سوئیچ به کنترلر می‌باشد.

هم‌چنین به تازگی تحقیقی توسط هیو و همکارانش انجام شده است که به مسأله قرارگیری کنترلرها از دیدگاه ذخیره انرژی نگاه می‌کند [۳۴]. آن‌ها برای کار خود، مسأله قرارگیری ذخیره‌کننده انرژی را توسط یک مسأله خطی عدد صحیح دودویی[18] مدل کردند. در این مدل، مصرف انرژی شبکه‌هایی که برای ترافیک کنترلی استفاده می‌شوند، تحت محدودیت‌های تأخیر مسیرهای کنترلی و بار کنترلرها کمینه می‌شوند. هم‌چنین آن‌ها با توجه به پیچیدگی مدل خطی عدد صحیح، یک الگوریتم مکاشفه‌ای ژنتیک نیز برای پیدا کردن راه‌حل‌های زیربهینه ارائه دادند.

یک کار متفاوت دیگر نیز توسط ایشیگاکی و همکارانش انجام شده است [۳۵]. آن‌ها در کار خود، راه‌حلی برای مسأله‌ی قرارگیری کنترلرها ارائه داده‌اند که قرارگیری را با توجه به بار روی گره‌های ارتباطی انجام می‌دهد. باید در نظر گرفت که پیام‌های کنترلی بین یک کنترلر و یک سوئیچ که در بخش داده ردوبدل می‌شوند، توسط دیگر سوئیچ‌ها به کنترلر می‌رسند. معمولاً الگوریتم‌های مسیریابی برای چنین انتقال‌هایی بر اساس کوتاه‌ترین مسیر عمل می‌کنند که باعث می‌شود دائماً برخی گره‌های خاص به عنوان نقاط بازپخش انتخاب شوند. به عبارت دیگر بسیاری از کوتاه‌ترین مسیرها بین سوئیچ‌ها و کنترلر ممکن است شامل گره‌های مشابه باشند. بدین صورت بار زیادی بر روی این گره‌ها به وجود می‌آید که ممکن است باعث خرابی در این نقاط شود. به همین دلیل در این مقاله قرارگیری با توجه به حداکثر بار بر روی این گره‌ها انجام شده است.

تمام موارد فوق تنها بخشی از مطالعات و پژوهش‌هایی بود که در خصوص گمارش و قرارگیری کنترلرها و توزیع آن‌ها در شبکه‌های مبتنی بر نرم‌افزار صورت گرفته است.

۴-روش پیشنهادی

در این بخش با توجه به بررسی‌های صورت گرفته در حوزه‌ی موضوع پژوهش و هم‌چنین با درنظر گرفتن معایب مقاله پایه [۳۲]، روش پیشنهادی خود را که ECNPA[19] نامگذاری کرده‌ایم، ارائه خواهیم کرد و به تشریح آن خواهیم پرداخت. عیب عمده الگوریتم به کار رفته در مقاله پایه در این است که در عمل خوشه‌بندی کنترلرها، تنها معیار تأخیر را در نظر می‌گیرد و توجهی به بار موجود روی لینک‌ها ندارد که این موضوع باعث به وجود آمدن ازدحام در لینک‌ها و ایجاد لینک‌های گلوگاه[20] می‌شود. ولی در روش پیشنهادی، علاوه بر تأخیر بین کنترلرها و سوئیچ‌ها، بار موجود روی لینک‌ها نیز در نظر گرفته شده و به تبع آن با دقت بالاتری عمل خوشه‌بندی را ادامه می‌دهیم. تأخیر و هم‌چنین بار موجود روی لینک‌ها در بین کنترلرها و سوئیچ‌های مربوط به آن‌ها را می‌توان یک ضرورت برای SDN در نظر گرفت چرا که همه‌ی توابع در یک شبکه‌ی مبتنی بر SDN می‌توانند به وسیله‌ی مبادله‌ی مکرر پیام ها در بین کنترلرها و سوئیچ‌ها صورت گیرد. در این بخش به بررسی مؤلفه‌های ممکن مربوط به تأخیر در بین کنترلرها و سوئیچ‌ها می‌پردازیم و در بخش بعد به تدوین مسأله‌ی گمارش کنترلرها خواهیم پرداخت.

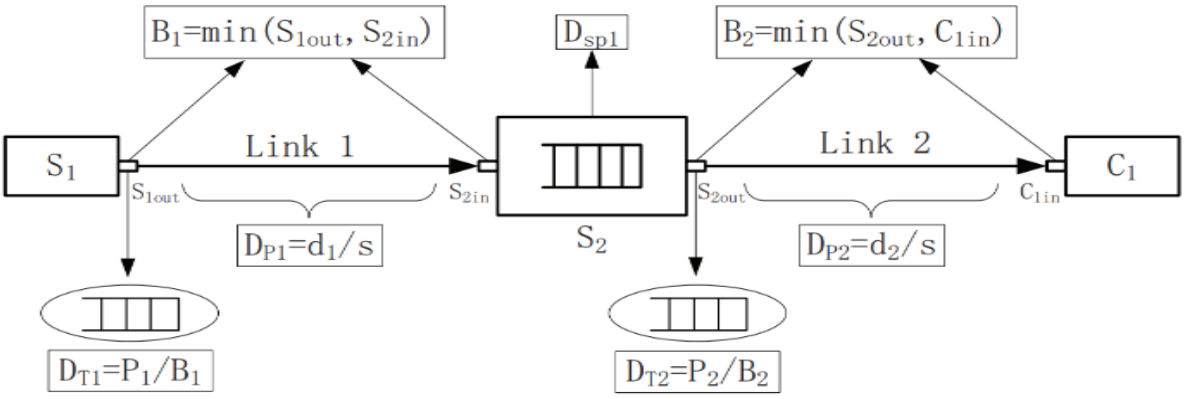
در یک شبکه‌ی مبتنی بر نرم‌افزار با تعداد مشخصی از گره‌ها و لینک‌ها، فرض کنید که $S = \{s_1, s_2, \dots, s_m\}$ بیانگر سوئیچ‌های موجود در شبکه باشد و $C = \{c_1, c_2, \dots, c_n\}$ نیز بیانگر کنترلرها باشد. فرض بر آن است که تعداد گام‌های[21] حرکتی مربوط به سوئیچ s_m تا کنترلر c_n به صورت $h(s_m, c_n)$ نمایش داده می‌شود. در شکل ۳، تأخیر انتها به انتها را در شرایطی مشاهده می‌کنید که انتقال یک بسته از سوئیچ s_1 به سمت کنترلر c_1 صورت می‌گیرد.

[18] Binary Integer Program (BIP)

[19] Enhanced Cluster-based Network Partitioning Algorithm

[20] Bottleneck

[21] Hop



شکل ۳- تأخیر انتها به انتها در انتقال بسته در SDN [۳۲]

تأخیر انتها به انتها مربوط به انتقال یک بسته‌ی داده‌ای از سوئیچ s_1 به کنترلر c_1، متشکل از سه مؤلفه می‌باشد: تأخیر در انتقال بسته (D_{Ti})، تأخیر در انتشار بسته (D_{Pi}) و تأخیر در پردازش سوئیچ (D_{SP_i}). تأخیر در انتقال به معنای زمان مورد نیاز برای انتقال بیت‌های یک بسته‌ی داده‌ای به سمت یک لینک می‌باشد که آن را به صورت $D_{Ti} = P_i/B_i$ نمایش می‌دهند و در آن، P_i بیانگر تعداد بیت‌های داده‌ای یک بسته در لینک i بوده و B_i نیز بیانگر پهنای باند لینک می‌باشد. در نظر داشته باشید که پهنای باند یک لینک خاص را می‌توان به وسیله‌ی حداقل نسبت انتقال واسطهای شبکه مشخص کرد، یعنی $B_1 = \min(S_{1out}, S_{2in})$ که در شکل ۳ نمایش داده شده است. تأخیر در انتشار بسته‌ی داده‌ای به معنای زمان مورد نیاز برای رسیدن یک بسته به مقصد می‌باشد. تأخیر در انتشار به صورت $D_{pi} = \frac{d_i}{s}$ مشخص شده که در آن، d_i بیانگر فاصله‌ی لینک i بوده و s بیانگر سرعت سیگنال رسانه‌ای است که برای انتقال داده‌ها بکار گرفته می‌شود. تأخیر در پردازش سوئیچ نیز به صورت D_{spi} نمایش داده می‌شود که تحت تأثیر بار سوئیچ i قرار می‌گیرد. بنابراین تأخیر انتها به انتها برای بسته‌ای که از سوئیچ S_m به سمت کنترلر c_n حرکت می‌کند در معادله ۱ نمایش داده شده است:

$$D_{e2e}(s_m, c_n) = \sum_{i=1}^{h(s_m,c_n)} \left(\frac{P_i}{B_i} + \frac{d_i}{S} + D_{SP}(i-1)\right) \quad (۱)$$

در نظر داشته باشید که سه مؤلفه از تأخیر انتها به انتها، وزن یکسانی در شبکه‌های مختلف ندارند. برای مثال در یک شبکه محلی[۲۲] (LAN) که پهنای باند کم سرعتی داشته و محدوده‌ی انتقال نیز محدود می‌باشد، تأخیر در انتقال بسته‌های داده‌ای (D_T)، غالب بوده در حالی که تأخیر در انتشار (D_P) را می‌توان نادیده گرفت. در مقابل، در یک شبکه‌ای که مجهز به سوئیچ‌های ۱۰ یا ۱۰۰ گیگابیت بر ثانیه می‌باشد [۳۶] [۳۷] و فاصله‌ی صدها مایلی دارد، تأخیر در انتقال بسته را می‌توان نادیده گرفت در حالی که تأخیر در انتشار بسته‌های داده‌ای یک تأخیر مهم می‌باشد. تأخیر در پردازش سوئیچ (D_{SP}) نیز به وسیله‌ی کارائی سوئیچ‌ها مشخص می‌شود. سوئیچ‌هایی با کارائی بالا و سوئیچ‌هایی با سرعت سیمی را همراه با تأخیر پایین و نوسان در شبکه‌های ستون فقرات مشاهده می‌کنیم. اخیراً دستاوردهای مهمی بر روی سوئیچ‌های SDN صورت گرفته است. برای مثال همان‌طور که در [۳۸] مطرح شده است، Corsa DP2100 [۳۷] می‌تواند به توان عملیاتی ۱۰۰ گیگابیت بر ثانیه دست پیدا کند و این بدین معنا است که تأخیر در پردازش سوئیچ (D_{sp}) را می‌توان در صورتی که از این نوع سوئیچ‌ها در یک شبکه‌ی ستون فقرات مبتنی بر نرم‌افزار استفاده شود، چشم‌پوشی کرد.

در ادامه به تدوین مسأله‌ی تقسیم‌بندی شبکه در SDN می‌پردازیم. برای یک شبکه‌ی مبتنی بر نرم‌افزار با تعداد مشخصی از گره‌ها و لینک ها، توپولوژی فیزیکی شبکه را به صورت یک گراف $G = (V, E)$ نمایش می‌دهیم که در آن،

[22] Local Area Network (LAN)

V بیانگر یک مجموعه از گره‌ها بوده و هر دوی سوئیچ‌ها و کنترلرها را در بر دارد. این بر مبنای این فرضیه بوده که کنترلرها در مکان‌هایی قرار دارند که سوئیچ‌ها نیز در آن مکان قرار دارند به گونه‌ای که کنترلرها و سوئیچ‌ها را می‌توان به هم متصل نمود [۱۱] [۱۲] [۲۸].

E بیانگر مجموعه‌ای از لینک‌های فیزیکی در بین این گره‌ها می‌باشد. فرض بر آن است که تعداد زیرگراف‌ها به صورت K مشخص می‌شود و بیانگر تعداد زیرشبکه‌های موجود در شبکه باشد. در زمانی که SDN در نظر گرفته می‌شود، پارتیشن شبکه را می‌توان به صورت $SDN_i(V_i, E_i)$ تعریف کرد:

$$\bigcup_{i=1}^{k} V_i = V; \bigcup_{i=1}^{K} E_i = E \quad (۲)$$

$$SDN_i \cap SDN_j = \emptyset, \quad \forall i \neq j, i, j \in k \quad (۳)$$

$$similarity(SDN_i) = TRUE, \quad \forall i \in K \quad (۴)$$

$$Similarity\left(SDN_i \cap SDN_j\right) = FALSE \quad \forall i \neq j, i, j \in k \quad (۵)$$

$$SDN_i \text{ is a connected region} \quad \forall i \in k \quad (۶)$$

معادله‌ی ۲ بکار گرفته شده تا نمایش داده شود که کل زیرشبکه‌ها باید همه‌ی اِلمان‌ها را پوشش دهند: گره‌ها و لینک‌ها. در معادله‌ی ۳، Ø به معنای یک مجموعه‌ی خالی بوده به این معنا که یک گره یا لینک تنها می‌تواند به یک زیرشبکه تخصیص داده شود و معادله‌ی ۴ نیز مبین این بوده که المان‌های موجود در یک زیرشبکه داری تشابه یکسانی می‌باشند. در معادله‌ی ۵ نشان داده شده است که اِلمان‌هایی که به زیرشبکه‌های مختلف تخصیص داده می‌شوند، دارای کمترین تشابه با یکدیگر هستند. منظور ما از تشابه، همان هزینه (تأخیر و بار) می‌باشد. به طور خاص، انتظار می‌رود که شبکه به گونه‌ای تقسیم شود که گره‌های موجود در یک خوشه دارای هزینه کمتری باشند. معادله‌ی ۶ نشان می‌دهد که همه‌ی رئوس موجود در زیرشبکه‌ها به وسیله‌ی لینک به هم متصل می‌باشند. معادلاتی که در بالا به آن‌ها اشاره شد، همگی در مقاله پایه مورد استفاده قرار گرفته‌اند. حال هدفی که در این مقاله دنبال می‌کنیم این است که کنترلرها را در یک شبکه‌ی گسترده‌ی مبتنی بر نرم‌افزار به گونه‌ای قرار دهیم تا بتوان حداکثر هزینه بین کنترلرها و سوئیچ‌ها را به حداقل سطح ممکن رساند. در روش پیشنهادی اگر $Load_i$ را میزان مشغول بودن لینک i ام در نظر بگیریم، میزان مشغول بودن مسیر بین سوئیچ s_m و کنترلر c_n از رابطه‌ی ۷ محاسبه می‌شود:

$$Load(s_m, c_n) = \max\{Load_i\} \quad (\forall i \in h(s_m, c_n)) \quad (۷)$$

هزینه را به صورت معادله‌ی ۸ نمایش می‌دهیم که شامل تأخیر انتها به انتها و درصد بار موجود یا همان میزان مشغول بودن لینک گلوگاه در مسیر بین کنترلر تا سوئیچ می‌باشد:

$$Cost = D_{e2e}(s_m, c_n) \times Load(s_m, c_n) \quad (۸)$$

بنابراین تابع هدف[23] را به صورت رابطه ۹ بیان می‌کنیم:

$$min\{max\{Cost\}\} \quad , \quad \forall s_m, c_n \in SDN_i \ , (i \in k) \quad (۹)$$

۴-۱- تشریح نحوه تقسیم‌بندی شبکه به k زیرشبکه با استفاده از الگوریتم پیشنهادی

در این بخش به ارائه‌ی یک الگوریتم بخش‌بندی شبکه بر مبنای خوشه‌بندی می‌پردازیم و چگونگی بخش‌بندی یک شبکه به زیرشبکه‌ها را با هدف کاهش ماکزیمم هزینه‌ی انتها به انتها ارائه می‌دهیم. مسأله‌ی بخش‌بندی شبکه مشابه با مسأله‌ی خوشه‌بندی می‌باشد و راهکارهای آن را می‌توان از الگوریتم‌های خوشه‌بندی الهام گرفت. البته الگوریتم‌های استاندارد خوشه‌بندی، مانند K-means و K-center را نمی‌توان به صورت مستقیم برای بخش‌بندی یک شبکه به زیرشبکه‌ها بکار گرفت که دلیل آن را می‌توان ناشی از موارد زیر دانست: اول اینکه انتخاب تصادفی مراکز اولیه نمی‌تواند این تضمین را بدهد که هر پارتیشن می‌تواند هزینه حداکثری در بین مرکز و گره‌های مربوط به آن را در زیرشبکه‌ها کاهش دهد. دوم اینکه فاصله‌ی اقلیدسی را نمی‌توان برای محاسبه‌ی فاصله‌ی بین دو گره جغرافیایی بکار گرفت. به منظور غلبه بر این معایب، از الگوریتم پیشنهادی برای حل مسأله‌ی بخش‌بندی شبکه استفاده می‌کنیم. در الگوریتم پیشنهادی، به محاسبه‌ی هزینه بین هر دو گره پرداخته می‌شود. مسیری با کوتاه‌ترین فاصله در الگوریتم پیشنهادی بکار گرفته شده تا تأخیر انتها به انتها

[23] Objective Function

محاسبه شود. به طور خاص، در یک توپولوژی شبکه با لینک‌ها و گره‌های مختلف، ماتریس مجاورت این توپولوژی در ابتدا محاسبه می‌شود. مختصات این گره‌ها را از روی نقشه‌ی گوگل به دست آوردیم و فاصله‌ی این لینک‌ها را نیز با استفاده از فرمول "haversine" محاسبه کردیم [۳۹] [۴۰]. هم‌چنین کوتاه‌ترین مسیر بین هر دو گره را با استفاده از الگوریتم دایجسترا محاسبه خواهیم کرد [۴۱]. مراحل الگوریتم پیشنهادی در زیر مشاهده می‌شود:

الگوریتم پیشنهادی برای خوشه‌بندی و توزیع کنترلرها

ورودی‌ها:

۱- توپولوژی شبکه که به صورت $G = (V, E)$ نمایش داده می‌شود و از داخل یک فایل XML وارد برنامه می‌شود که در آن V نشان‌دهنده‌ی گره‌ها و E نشان‌دهنده‌ی یال‌ها یا همان لینک‌های ارتباطی می‌باشد.

۲- تعداد زیر شبکه‌ها. (K)

شروع

گام اول: محاسبه‌ی تأخیر انتها به انتها بین هر دو گره دلخواه. $\forall a, b \in V \; l_{e2e}(a, b)$.

گام دوم: محاسبه‌ی میزان مشغول بودن لینک گلوگاه در مسیر بین هر دو گره دلخواه. $\forall a, b \in V \; load(a, b)$.

گام سوم: محاسبه‌ی اولین مرکز شبکه. گره‌ای که دارای کمترین هزینه (تأخیر و بار) نسبت به سایر گره‌ها باشد به عنوان اولین مرکز انتخاب می‌شود.

گام چهارم: پیدا کردن مرکز بعدی شبکه. گره‌ای که دارای بیشترین هزینه (تأخیر و بار) نسبت به مراکز قبلی می‌باشد، به عنوان مرکز بعدی انتخاب می‌شود.

گام پنجم: توزیع بردار $v(v \in V)$ بر روی یکی از خوشه‌ها و آن هم با استفاده از رابطه‌ی زیر:

$$v \in Cluster_i \; , \quad if \; Cost(v, c_i) < Cost(v, c_j) \; , \forall i, j \in \{1,2, \dots, k\}$$

گام ششم: بروز رسانی مراکز $C' = \{c'_1, c'_2, \dots, c'_k\}$ به گونه‌ای که جمع هزینه‌ی مربوط به همه‌ی گره‌ها در خوشه i نسبت به مرکز c'_i به حداقل سطح ممکن برسد.

$$c'_i = v_m \; , \quad if \; Cost(v_m, v) = minimum$$
$$\forall \; m \; \in size(cluster_i), v \; \in Cluster_i \; , \; i = \{1,2, \dots, k\}$$

گام هفتم: مراحل ۴، ۵ و ۶ را تکرار می‌کنیم تا اینکه شبکه به k زیر شبکه تقسیم شود.

پایان

خروجی‌ها:

۱- مراکز K خوشه.

۲- تخصیص سوئیچ‌ها به هر یک از خوشه‌ها.

شبه‌کد مربوط به محاسبه فاصله بین هر دو گره از دو سر یک لینک در شبکه WAN با استفاده از فرمول Haversine که در بالا به آن اشاره شد، به صورت زیر می‌باشد:

شبه‌کد محاسبه فاصله بین هر دو گره با فرمول Haversine

```
1)Begin
2)Function dist = CalcDistance(node1,node2)
3)         lat1 = node1(1);
4)         lng1 = node1(2);
5)         lat2 = node2(1);
6)         lng2 = node2(2);
7)         earthRadius = 3958.75;   % In
Miles(6370.99056 In Kilometers)
8)         dLat = toRadians(lat2 - lat1);
9)         dLng = toRadians(lng2 - lng1);
10)         a = sin(dLat / 2) * sin(dLat / 2) +
11)         cos(toRadians(lat1)) *
cos(toRadians(lat2)) *
12)         sin(dLng / 2) * sin(dLng / 2);
13)         c = 2 * atan2(sqrt(a), sqrt(1 - a));
14)         dist = earthRadius * c;
15)         meterConversion = 1/0.000621371192;
16)         dist=dist * meterConversion;
17)         Function r=toRadians(d)
18)             r=d/180*pi;
19)         End function
20)End function
21)End
```

شبه‌کد مربوط به محاسبه تأخیر انتها به انتها بین هر دو گره از دو سر یک لینک در شبکه WAN با توجه به معادله‌ی ۱ که در بالا به آن اشاره شد، به صورت زیر می‌باشد:

شبه‌کد محاسبه تأخیر انتها به انتها با معادله‌ی ۱

```
1)Begin
2)Function  [Delay]  =  CalcLatency(Distance)     %
Distance is in Meters
3)     SpeedOfLight=299792458;  % Meters Per Second
4)          WANFiberSpeed=(SpeedOfLight*65)/100;     %
Meters Per Second
5)     Delay=(Distance/WANFiberSpeed)*1000;  % Delay
in Millisecond
6)End function
7)End
```

شبه‌کد مربوط به محاسبه میزان مشغول بودن لینک گلوگاه در مسیر بین هر دو گره دلخواه در شبکه با توجه به معادله‌ی ۷ که در بالا به آن اشاره شد، به صورت زیر می‌باشد:

شبه‌کد محاسبه میزان مشغول بودن لینک گلوگاه با معادله‌ی ۷

```
1)Begin
% Find the bottleneck link load
2)For each node with indices i,j in Topology
3)tmp=[];
4)   For k=FindLinks(path,links)
5)     tmp=[tmp;loads(loads(:,1)==k,:)];
6)   End for
```

```
7)tmp=tmp(tmp(:,2)==max(tmp(:,2)),2);
8)AllNodesBottleNeckLinkLoad(i,j)=tmp(1);
9)AllNodesBottleNeckLinkLoad(j,i)=tmp(1);
10)End for
11)End
```

بطور خلاصه، اولین مرحله در الگوریتم پیشنهادی به این صورت است که می‌تواند مرکز واقعی شبکه (محور) را به راحتی پیدا کند. به طور خاص، الگوریتم پیشنهادی اقدام به محاسبه‌ی هزینه مربوط به هر گره نسبت به سایر گره‌ها نموده و گره ای را که دارای کمترین هزینه تا سایر گره‌ها می‌باشد را به عنوان محور انتخاب می‌کند. در گام بعد، الگوریتم پیشنهادی اقدام به پیدا کردن مرکز دوم شبکه می‌کند. این مرکز دوم، به عنوان گره ای انتخاب می‌شود که بیشترین هزینه را دارد. در گام بعد، الگوریتم پیشنهادی نقش محور (c_1) و مرکز دوم (c_2) را به عنوان دو مرکز اولیه انتخاب کرده و گره‌های مربوطه را به این مراکز تخصیص می‌دهد. به طور خاص به ازای هر گره‌ی n_i، هزینه آن نسبت به مراکز محاسبه شده تا دو مقدار $Cost_1 = Cost(n_i, c_1)$, $Cost_2 = Cost(n_i, c_2)$ به دست آید. این دو مقدار با هم مقایسه شده و گره به مرکزی تخصیص داده می‌شود که به آن هزینه‌ی کمتری دارد. برای مثال در صورتی که $Cost_1 < Cost_2$ باشد، گره ی n_i به c_1 تخصیص داده می‌شود. زمانی که گره به یک مرکز دیگر تخصیص داده شد، محور این خوشه بر مبنای حداقل هزینه‌ای که در گام سوم مشخص شده است محاسبه می‌شود. این فرآیند تا زمانی ادامه پیدا می‌کند که شبکه به k زیرشبکه تقسیم شود. قابل ذکر است که بر خلاف الگوریتم‌های متعارف خوشه‌بندی، که به صورت تصادفی اقدام به انتخاب یک باره‌ی همه k مرکز نموده و سپس همه‌ی آن‌ها را در طی تکرارهایی بهینه می‌سازند، الگوریتم پیشنهادی در ابتدا اقدام به تقسیم کل شبکه به تعدادی زیرشبکه نموده تا به k زیرشبکه دست پیدا کند. در طول هر بخش، می‌توان هزینه را در بین مراکز و گره‌های مربوط به آن‌ها کاهش داده و از این رو هزینه حداکثری در سطح قابل ملاحظه‌ای در مقایسه با الگوریتم‌های متعارف خوشه‌بندی مانندK-center و K-means کاهش پیدا می‌کند.

۵- ارزیابی روش پیشنهادی

در این بخش به ارزیابی روش پیشنهادی می‌پردازیم و آن را با راهکارهای موجود و آن هم تحت توپولوژی‌های واقعی که از پروژه‌ی Topology Zoo[۲۴] به دست آمده است [۴۲] مقایسه می‌نماییم. الگوریتم پیشنهادی در نرم‌افزار متلب به عنوان یک محیط رایانش عددی قوی پیاده سازی شده است. این نرم‌افزار به صورت گسترده در حوزه‌ی تحقیقات گمارش کنترلر SDN بکار گرفته می‌شود. شبیه‌سازی‌ها شامل گام‌های زیر می‌باشد:

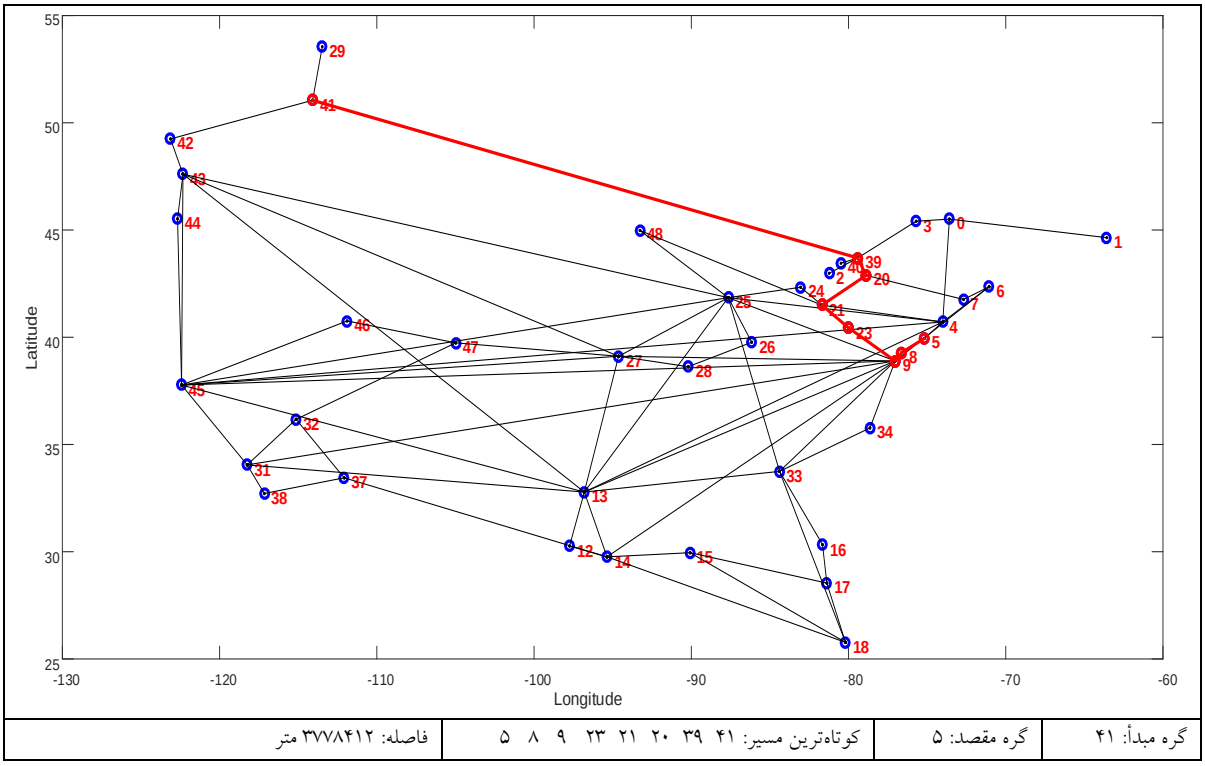
گام اول: در اولین مرحله با استفاده از تابعی با نام ImportGraphML دیتاست‌های مربوطه را که در قالب فایل‌های XML می‌باشند و شامل اطلاعاتی مانند موقعیت جغرافیایی[۲۵] گره‌های شبکه، مشخصات لینک‌ها، توپولوژی شبکه، و ... می‌باشند، وارد برنامه می‌کنیم. موقعیت جغرافیایی گره‌های شبکه شامل مشخصات طول جغرافیایی[۲۶] و عرض جغرافیایی[۲۷] می‌باشد که در جدولی به نام latlong ذخیره می‌شوند. اطلاعات مربوط به لینک‌ها در جدول links و اطلاعات مربوط به توپولوژی شبکه در جدول topology ذخیره می‌شود. تمامی عملیات مربوط به گام اول با اجرای تابع main انجام می‌شود.

گام دوم: فاصله‌ی بین هر دو گره از دو سر یک لینک با استفاده از فرمول Haversine که در تابع CalcDistance پیاده‌سازی شده است، محاسبه می‌شود.

گام سوم: کوتاه‌ترین مسیر بین هر دو گره دلخواه در شبکه با استفاده از الگوریتم Dijkstra که یکی از معروف‌ترین و محبوب‌ترین الگوریتم‌های مسیریابی می‌باشد، محاسبه می‌شود. این گام توسط تابع Dijkstra انجام شده و نتایج در جدول AllNodesDistances ذخیره می‌شود. در شکل ۴ نمونه‌ای از مسیریابی بین هر دو گره دلخواه از شبکه با استفاده از الگوریتم Dijkstra نمایش داده شده است. لازم به ذکر است که هم در روش پیشنهادی و هم در روش به کار رفته در مقاله پایه، از همین الگوریتم مسیریابی استفاده شده است.

[۲۴] پروژه Topology Zoo شامل صدها توپولوژی شبکه بوده که توسط سرویس‌دهندگان شبکه تدارک دیده شده است.

[25] Geographical Location
[26] Longitude
[27] Latitude



گره مبدأ: ۴۱	گره مقصد: ۵	کوتاه‌ترین مسیر: ۴۱ ۳۹ ۲۰ ۲۱ ۲۳ ۹ ۸ ۵	فاصله: ۳۷۷۸۴۱۲ متر

شکل ۴- نمونه‌ای از مسیریابی بین گره‌های شبکه‌ی Uunet با استفاده از الگوریتم Dijkstra

گام چهارم: در این مرحله با توجه به نتایج به دست‌آمده از مرحله قبل و با استفاده از تابع CalcLatency تاخیر انتها به انتها بین هر دو گره دلخواه محاسبه می‌شود. توجه به این نکته حائز اهمیت است که با توجه به این که گره‌های ما در سطح شبکه WAN توزیع شده‌اند و لینک‌های مورد استفاده فیبر نوری می‌باشد و با در نظر گرفتن این که سرعت انتقال سیگنال در فناوری فیبر نوری ۶۵ درصد سرعت نور است، محاسبات لازم در تابع CalcLatency انجام شده است. هم‌چنین نتایج به دست آمده از این مرحله در جدول AllNodesLatencies ذخیره شده است.

گام پنجم: اطلاعات مربوط به بار موجود روی هر لینک و درصد مشغول بودن آن توسط یک کنترلر مرکزی در جدولی به نام loads ذخیره می‌شود. در این مرحله با استفاده از تابعی به نام FindLinks ، لینک گلوگاه بین هر دو گره دلخواه به دست آمده و میزان مشغول بودن آن در جدولی به نام AllNodesBottleNeckLinkLoad ذخیره می‌شود. با شناسایی لینک‌های گلوگاه توانسته‌ایم ازدحام را در شبکه کنترل کنیم و با دقت بالاتری کنترلرها را توزیع کنیم و عمل خوشه‌بندی را انجام دهیم. لازم به توضیح است که برای پیاده‌سازی این مرحله، درصد مشغول بودن هر لینک را به صورت یک عدد تصادفی بین ۰ تا ۱۰۰ در نظر گرفته‌ایم که نشان‌دهنده وضعیت شبکه در آن لحظه مشخص می‌باشد.

گام ششم: با توجه به تابع هدفی که در بخش قبل معرفی کردیم، هزینه مربوط بین هر دو گره دلخواه محاسبه شده و در جدولی به نام AllNodesCosts ذخیره می‌شود.

گام هفتم: نتایج حاصل از اجرای الگوریتم پیشنهادی با الگوریتم K-means و CNPA مقایسه شده و نشان می‌دهیم که الگوریتم پیشنهادی بهتر عمل کرده است.

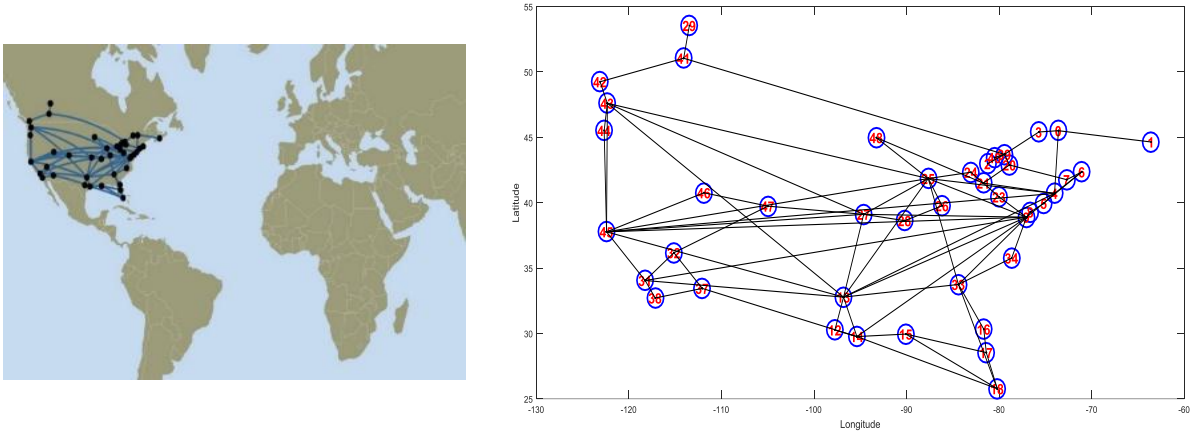
۵-۱- نتایج پیاده‌سازی

در این قسمت نتایج حاصل از پیاده‌سازی را با اجرای ۱۰۰ مرتبه از هر سه الگوریتم، در چند نمونه از توپولوژی‌های مربوط به پروژه Topology Zoo با هم مقایسه می‌کنیم. ضمناً با توجه به این که در مقاله پایه، توپولوژی Chinanet مربوط به کشور چین مورد بررسی قرار گرفته است و هم‌چنین جهت نشان دادن هر چه بهتر عملکرد الگوریتم پیشنهادی در توپولوژی‌های پیچیده‌تر و متفاوت‌تر در سایر کشورهای پیشرفته جهان مانند آمریکا، آلمان، و... به انتخاب

دیتاست‌های مربوطه پرداخته و نتایج به دست آمده را ارزیابی می‌کنیم.

۵-۱-۱- ارزیابی نتایج به دست آمده در توپولوژی Uunet مربوط به کشور آمریکا

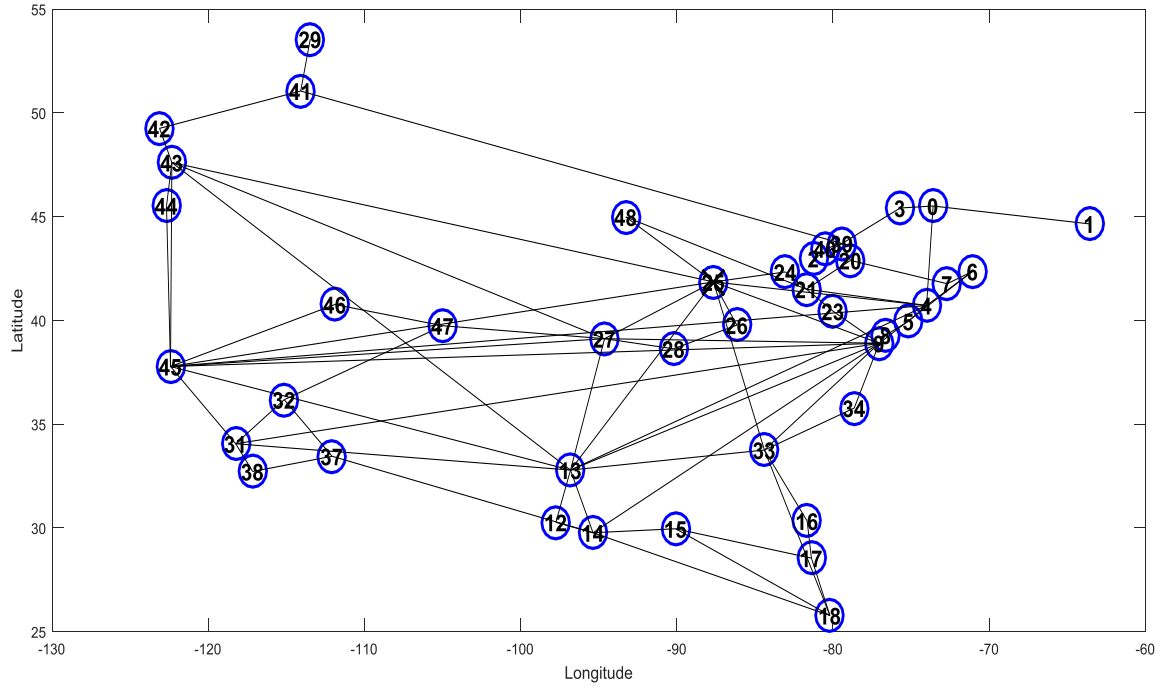
در شکل ۵ نمایی از توپولوژی واقعی شبکه‌ی Uunet را در مقابل توپولوژی پیاده‌سازی شده آن در نرم‌افزار متلب مشاهده می‌کنیم:



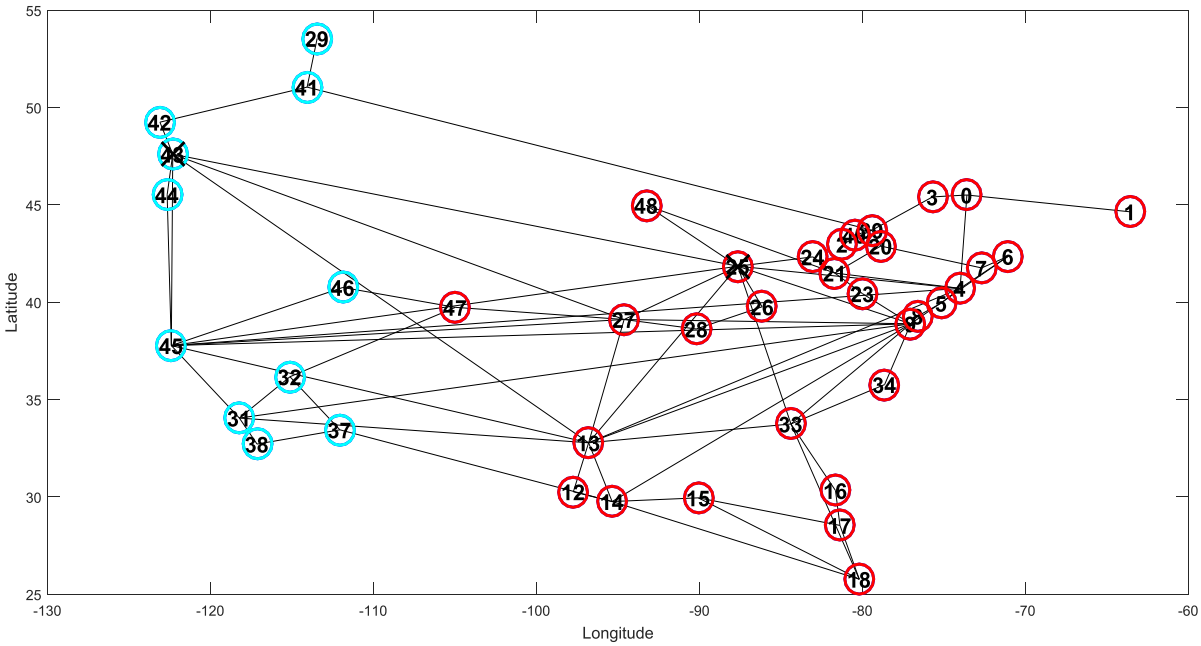
توپولوژی واقعی [۴۳] — توپولوژی پیاده‌سازی شده

شکل ۵- توپولوژی واقعی شبکه‌ی Uunet در مقابل توپولوژی پیاده‌سازی شده آن

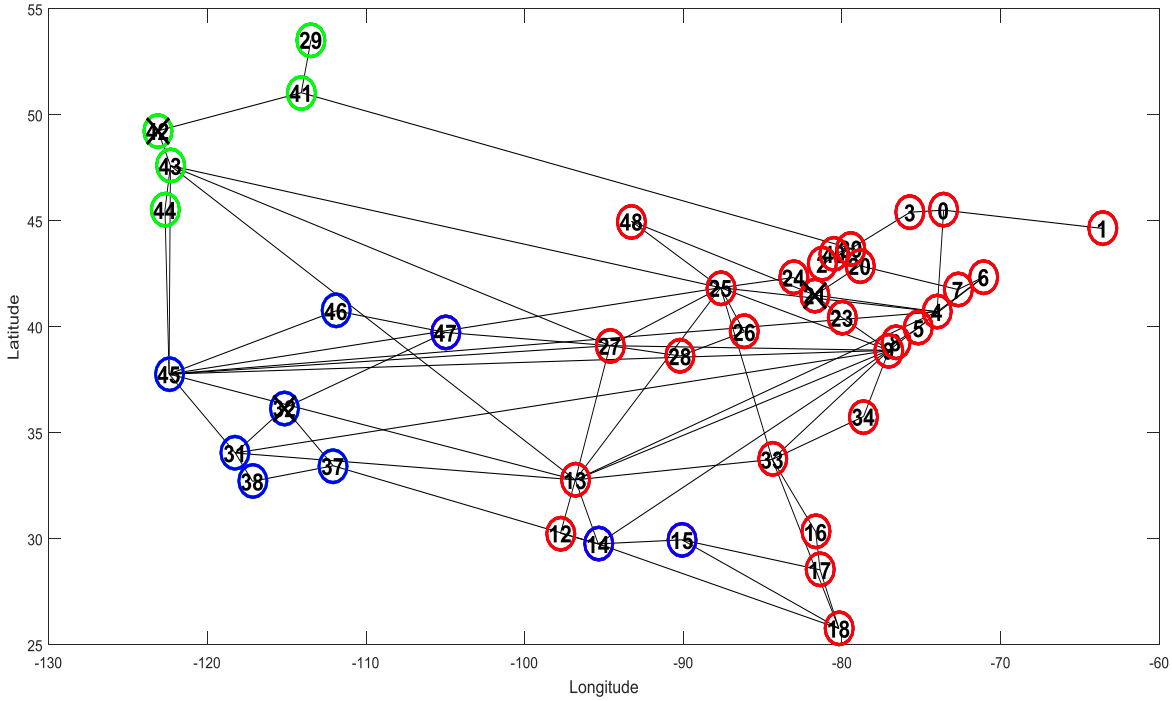
در ادامه، نتایج مربوط به اجرای هر یک از الگوریتم های CNPA و ECNPA را نمایش می‌دهیم. مراحل انتخاب سرخوشه‌ها در الگوریتم CNPA تا رسیدن به تعداد زیر-شبکه دلخواه در شکل ۶ نمایش داده شده است:



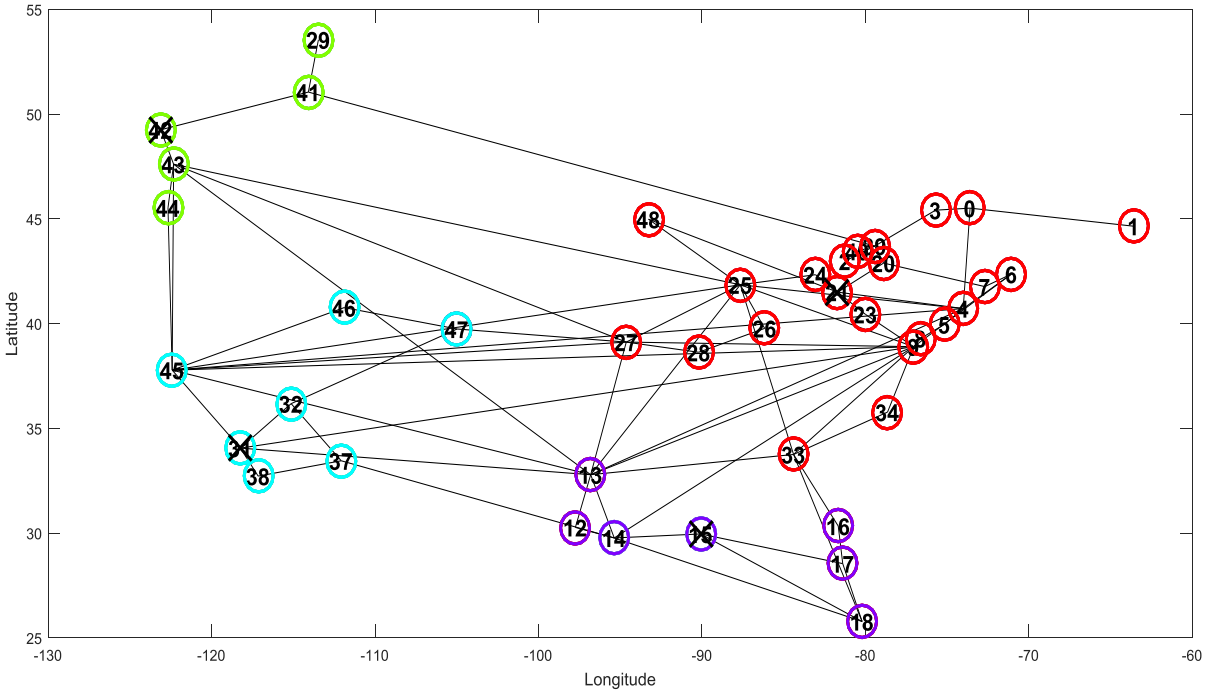
مرحله اول: ۲۵



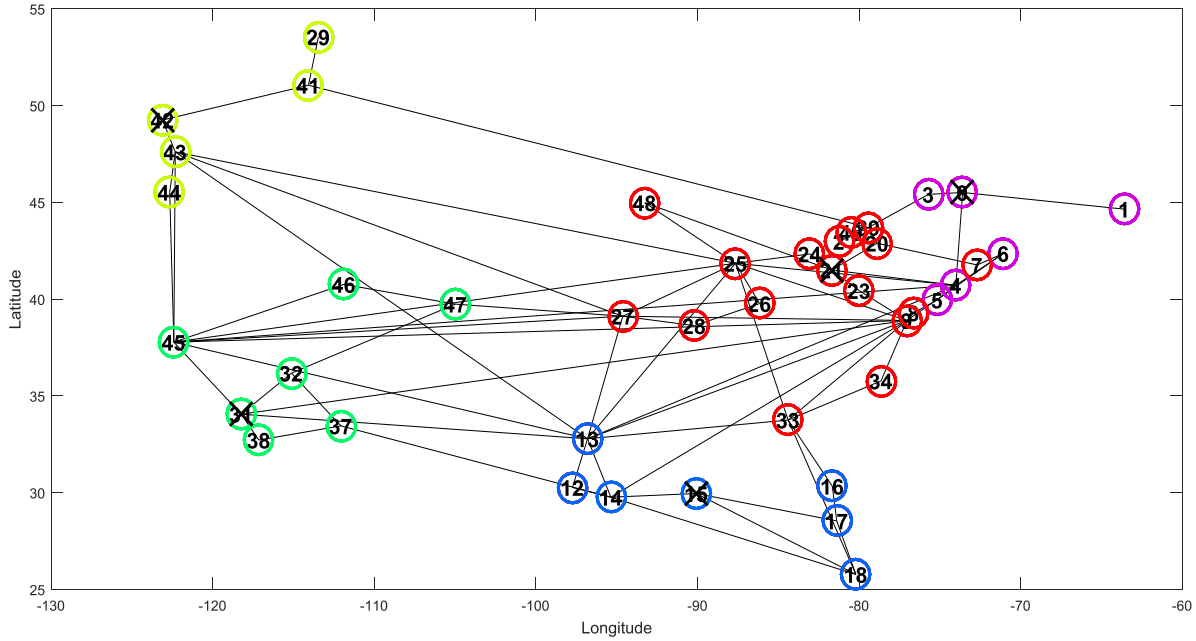
مرحله دوم: ۲۵ ۴۳



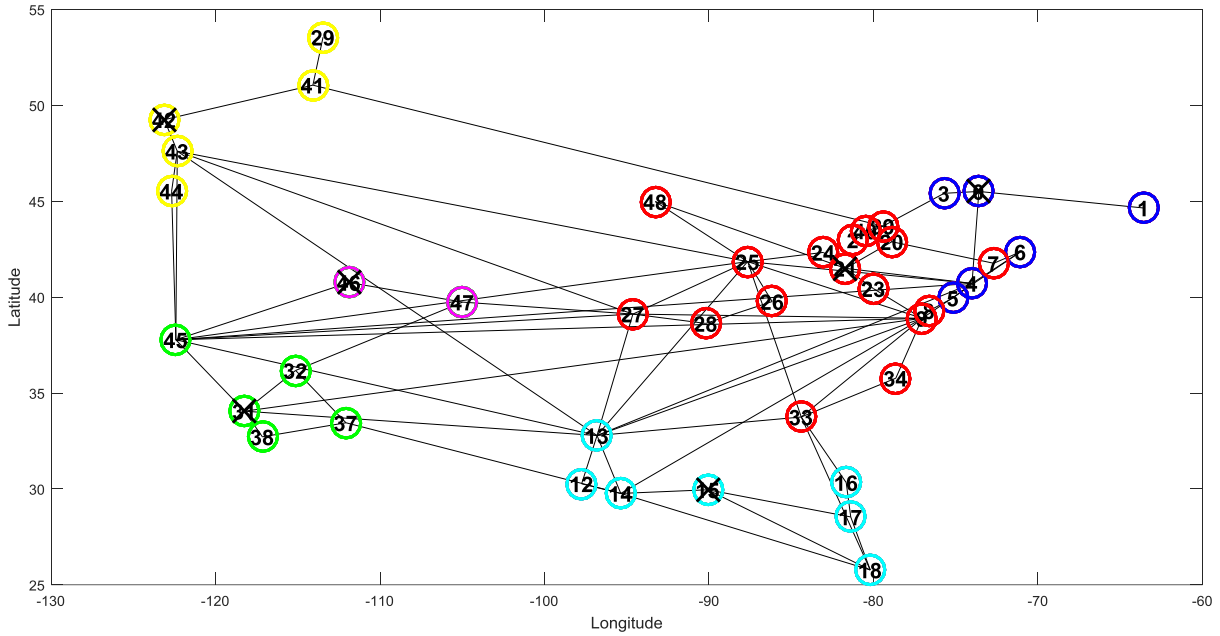
مرحله سوم: ۲۱ ۴۲ ۳۲



مرحله چهارم: ۲۱ ۴۲ ۳۱ ۱۵



مرحله پنجم: ۲۱ ۴۲ ۳۱ ۱۵ ۰

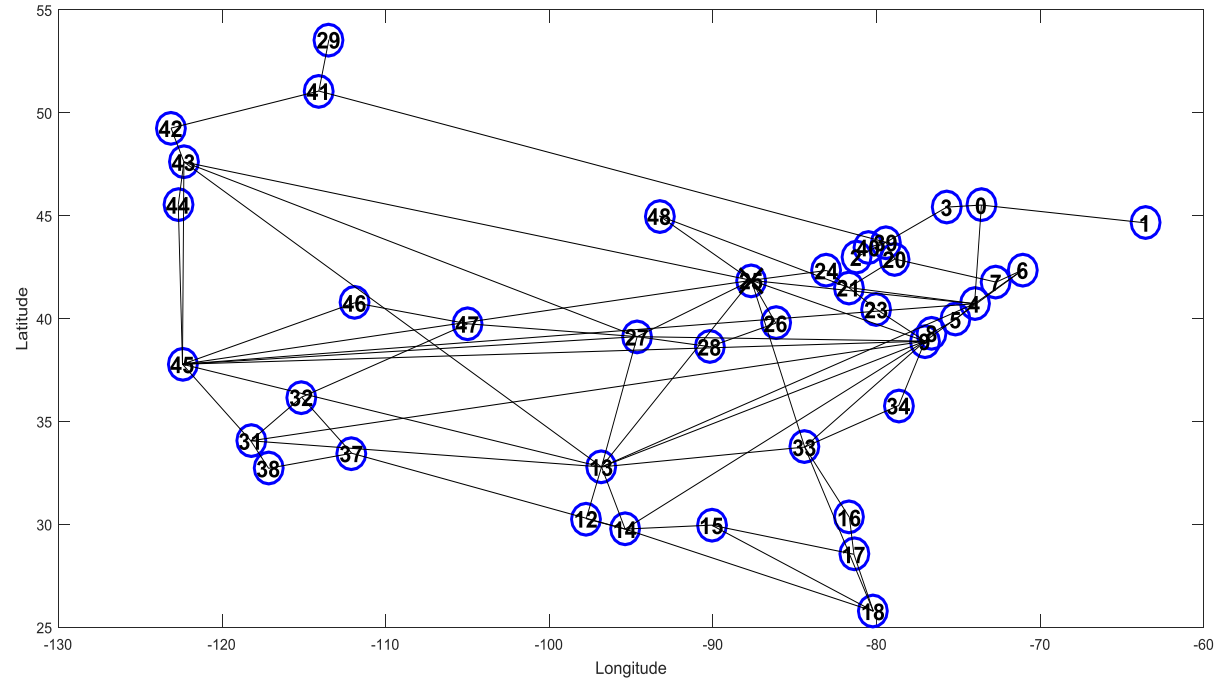


مرحله ششم: ۲۱ ۴۲ ۳۱ ۱۵ ۰ ۴۶

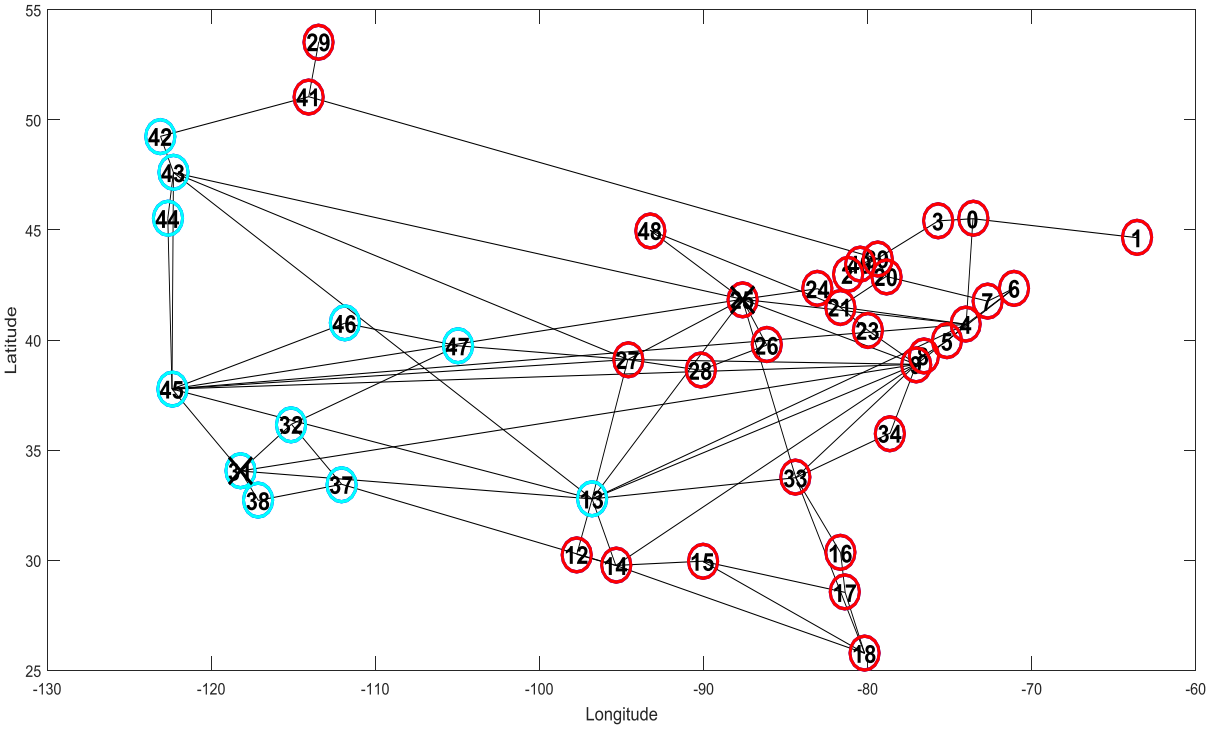
شکل ۶- اجرای الگوریتم CNPA روی شبکه Uunet به ازای k=6

حداکثر هزینه به دست آمده در تمام خوشه‌ها، بین سرخوشه ۲۱ (که معرف کنترلر می‌باشد) تا گره ۳۳ متعلق به آن (که معرف سوئیچ می‌باشد) با اجرای الگوریتم CNPA در توپولوژی Uunet برابر با ۶/۹۹۴۵ می‌باشد.

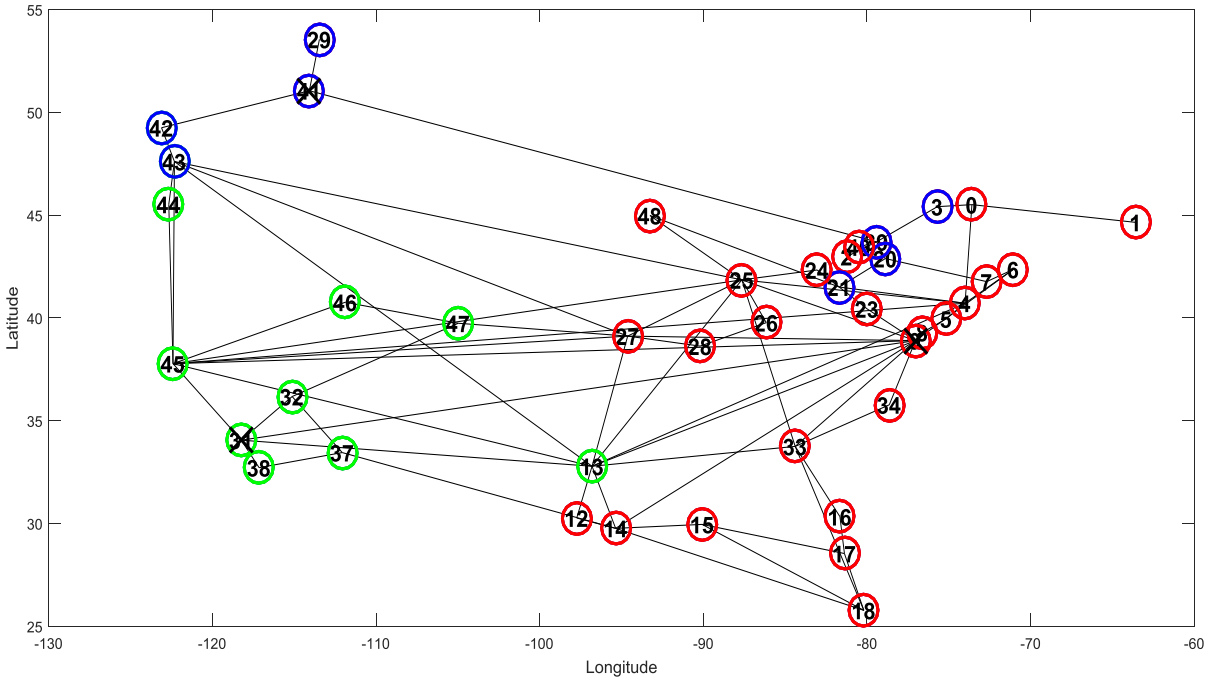
مراحل انتخاب سرخوشه‌ها در الگوریتم پیشنهادی (ECNPA) تا رسیدن به تعداد زیرشبکه دلخواه در شکل ۷ نمایش داده شده است:



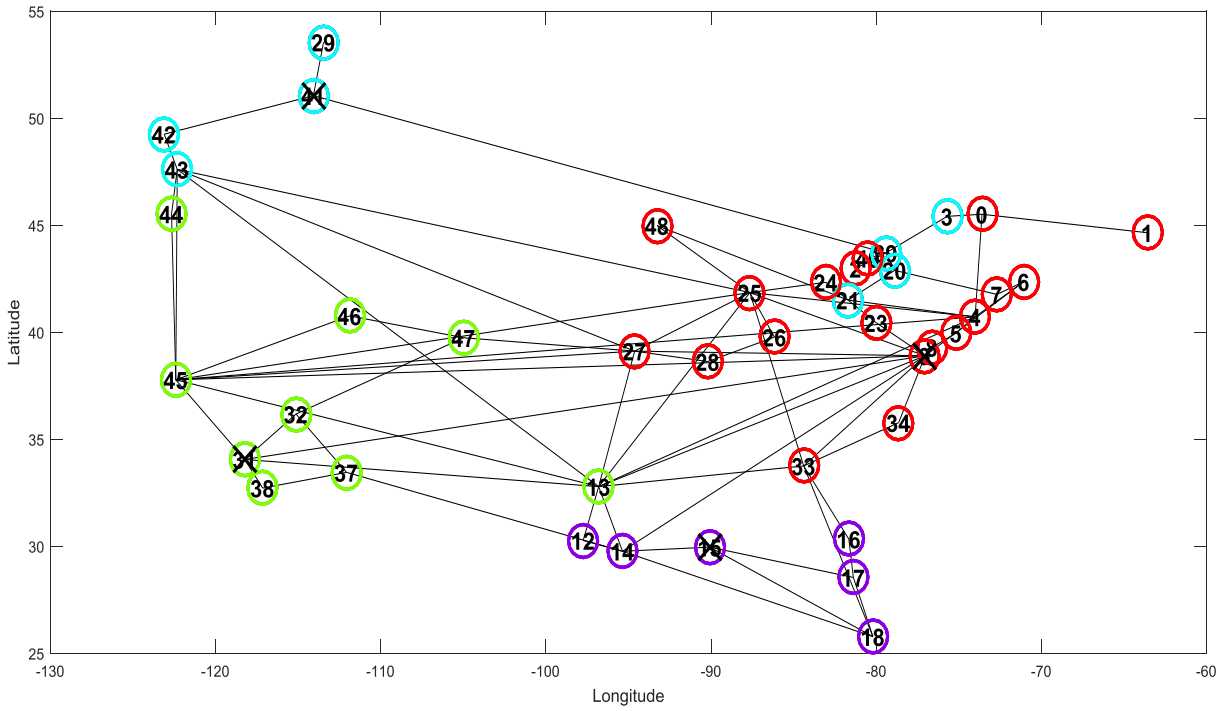
مرحله اول: ۲۵



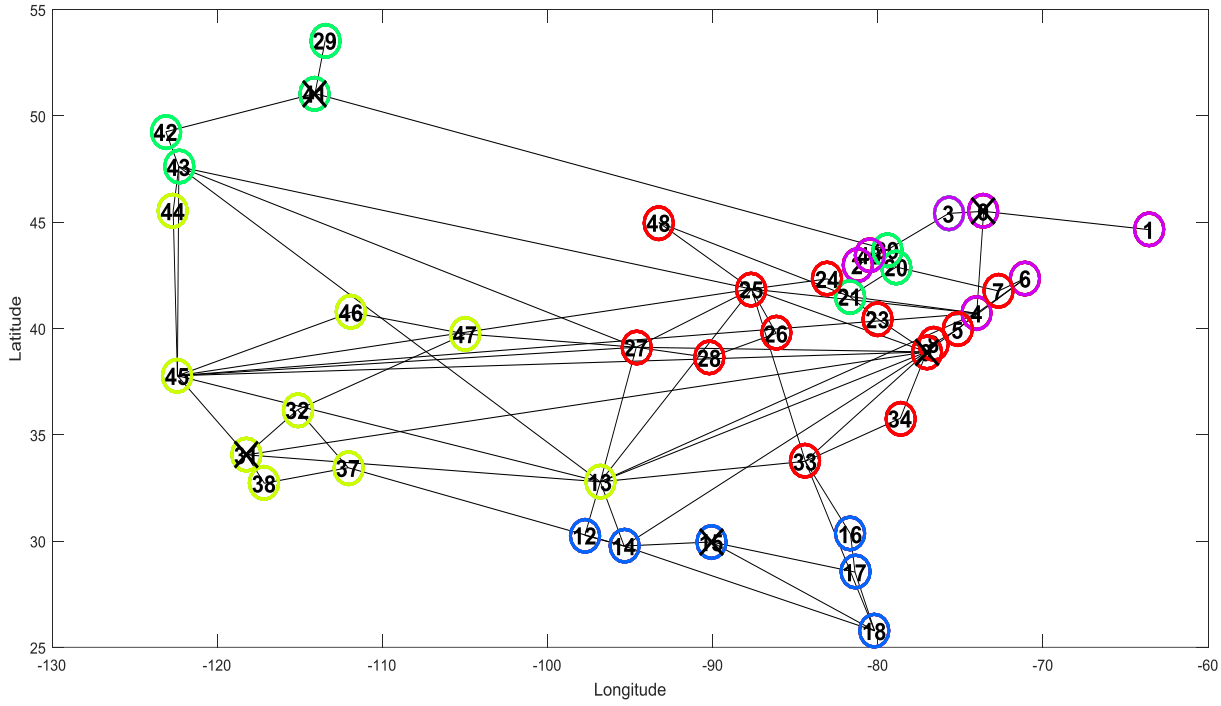
مرحله دوم: ۲۵ ۳۱



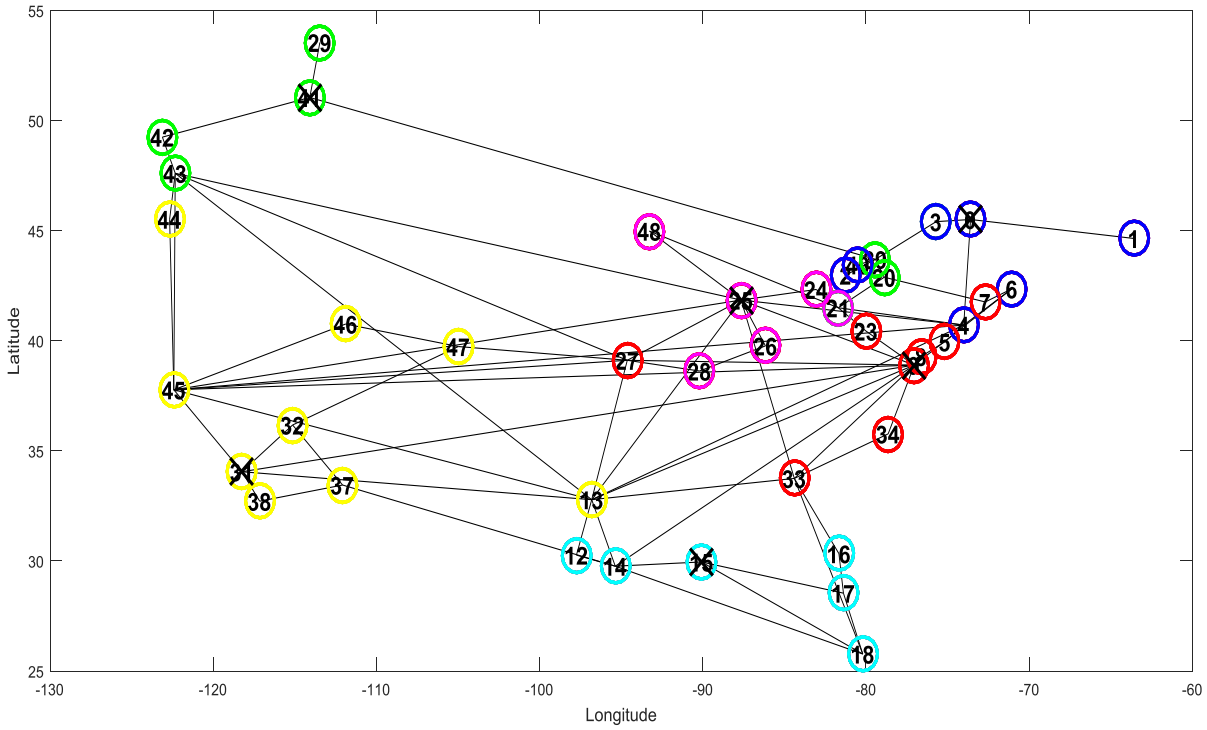
مرحله سوم: ۹ ۳۱ ۴۱



مرحله چهارم: ۹ ۳۱ ۴۱ ۱۵



مرحله پنجم: ۹ ۳۱ ۴۱ ۱۵ ۰

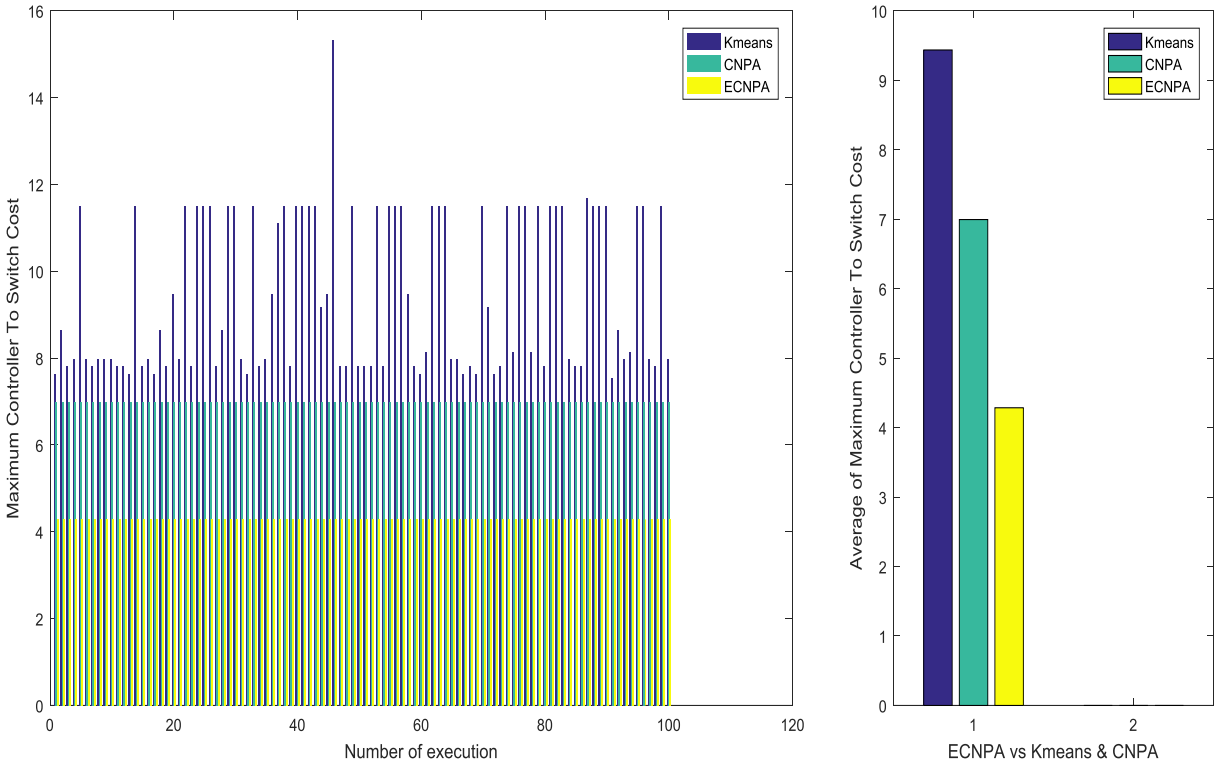


شکل ۷- اجرای الگوریتم پیشنهادی روی شبکه Uunet به ازای k=6

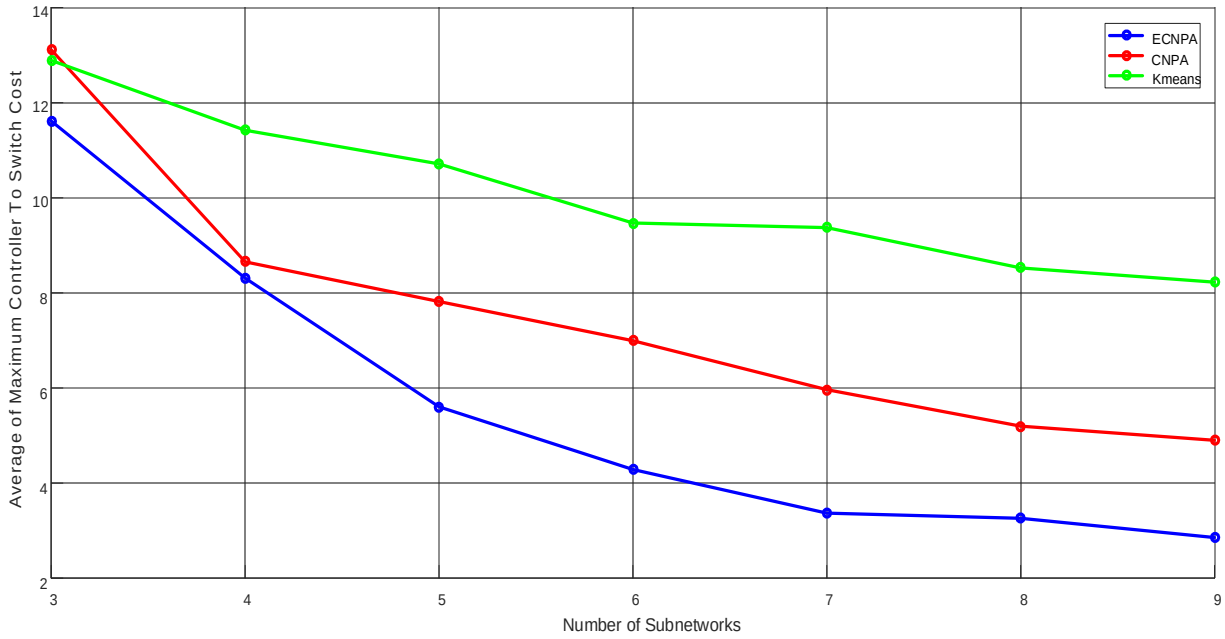
حداکثر هزینه به دست آمده در تمام خوشه‌ها، بین سرخوشه ۳۱ (که معرف کنترلر می‌باشد) تا گره ۴۶ متعلق به آن (که معرف سوئیچ می‌باشد) با اجرای الگوریتم پیشنهادی در توپولوژی Uunet برابر با ۴/۲۸۵۲ می‌باشد.

در نمودار ۱، نتایج به دست آمده در ۱۰۰ بار اجرای هر سه الگوریتم، نشان می‌دهد که در توپولوژی Uunet، میانگین حداکثر هزینه بین هر کنترلر و سوئیچ‌های مربوط به آن، در الگوریتم پیشنهادی مقدار کمتری می‌باشد. هم‌چنین نتایج شبیه‌سازی‌ها در توپولوژی Uunet نیز به وضوح نشان می‌دهد که اگرچه میزان تأخیر انتها به انتها در الگوریتم پیشنهادی ممکن است در بعضی از مواقع در مقایسه با الگوریتم CNPA کمی بیشتر باشد ولی در شرایطی که احتمال ازدحام در شبکه بالا می‌رود، الگوریتم پیشنهادی با شناسایی لینک‌های گلوگاه در مسیرهای ارتباطی هر گره با سایر گره‌ها، توانسته به خوبی ازدحام را در شبکه کنترل نماید و با در نظر گرفتن دو معیار تأخیر و میزان مشغول بودن لینک‌ها، فرایند قرارگیری و توزیع کنترل‌ها را در عمل خوشه‌بندی با دقت بالاتری انجام دهد و کارایی شبکه را افزایش دهد.

در نمودار ۲ نیز یک مقایسه‌ی کلی از هر سه الگوریتم زمانی که شبکه را به زیرشبکه‌های مختلفی تقسیم‌بندی می‌کنیم، ارائه شده است. نتایج به دست آمده نشان می‌دهد که با افزایش تعداد زیرشبکه‌ها در توپولوژی Uunet، میانگین هزینه‌ی انتها به انتها کاهش می‌یابد و الگوریتم پیشنهادی در کاهش میانگین هزینه‌ی انتها به انتها، نسبت به دو الگوریتم دیگر، بهتر عمل کرده است.



نمودار ۱- مقایسه نتایج الگوریتم پیشنهادی با الگوریتم K-means و CNPA در شبکه Uunet به ازای k=6



نمودار ۲- مقایسه‌ی کلی هر سه الگوریتم به ازای تعداد زیرشبکه‌های مختلف در توپولوژی Uunet

۶- نتیجه‌گیری

در این پژوهش به بررسی روش‌های جدیدی برای کاهش هزینه‌ی انتها به انتها (شامل تأخیر انتها به انتها و درصد مشغول بودن لینک‌ها) بین کنترلرها و سوئیچ‌های مرتبط با آن‌ها پرداختیم. در ابتدا تأخیر انتها به انتها در بین کنترلرها و سوئیچ‌ها را بررسی و به صورت کیفی تدوین نمودیم. در مرحله بعد به شناسایی لینک‌های گلوگاه که عامل ایجاد ازدحام هستند، پرداختیم و یک الگوریتم مبتنی بر خوشه-بندی را برای بخش‌بندی شبکه ارائه دادیم تا یک شبکه را به چندین زیرشبکه تقسیم کنیم تا بتوان هزینه انتها به انتها را

کاهش داد. به منظور ارزیابی کارائی الگوریتم پیشنهادی، شبیه‌سازی‌های زیادی را تحت توپولوژی‌های واقعی در پروژه‌ی Topology Zoo شامل توپولوژی Chinanet کشور چین، توپولوژی Uunet کشور آمریکا، توپولوژی DFN کشور آلمان، و توپولوژی Rediris کشور اسپانیا انجام دادیم. نتایج شبیه‌سازی‌ها نشان می‌دهد که الگوریتم پیشنهادی می‌تواند میانگین هزینه انتها به انتها در بین کنترلرها و سوئیچ‌های مربوطه را نسبت به الگوریتم‌هایی مانند K-means کاهش دهد. مراکز اولیه‌ی الگوریتم K-means به صورت تصادفی انتخاب می‌شوند و از این رو نتیجه‌ی پارتیشن شبکه به ازای هر بار اجرا متغیر است. بنابراین بدیهی است که ماکزیمم هزینه انتها به انتها در بین کنترلر و سوئیچ‌های مربوطه در زیرشبکه‌ها افزایش پیدا می‌کند. نتایج پیاده‌سازی به ازای صد بار اجرای هر یک از الگوریتم‌ها مورد بررسی قرار گرفته است و نتایج به دست آمده نشان می‌دهد که الگوریتم پیشنهادی می‌تواند همان نتایج را به ازای هر بار اجرا به دست آورد. دلیل آن این بوده که مراکز اولیه‌ی مناسبی محاسبه شده است و تعداد ثابتی از مراکز اولیه منجر به تقسیم بندی ثابت شبکه می‌شود. هم‌چنین الگوریتم پیشنهادی می‌تواند با شناسایی لینک‌های گلوگاه، از به وجود آمدن ازدحام در شبکه جلوگیری به عمل آورد.

در انتها لازم به ذکر است که در راستای استفاده از کنترلر-های توزیع‌شده و قرار دادن آن‌ها در شبکه‌های مبتنی بر نرم‌افزار می‌توان از الگوریتم‌های ابتکاری و فرا ابتکاری مانند الگوریتم ازدحام ذرات، الگوریتم ژنتیک، الگوریتم کلونی زنبور عسل، الگوریتم تکامل تفاضلی، و ... نیز جهت انجام دقیق‌تر و بهتر عمل خوشه‌بندی و توزیع کنترلرها استفاده کرد.

منابع

1.N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, *et al.*, "OpenFlow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, pp. 69-74, 2008.

2.W. Miao, G. Min, Y. Wu, H. Wang, and J. Hu, "Performance modelling and analysis of software-defined networking under bursty multimedia traffic," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 12, p. 77, 2016.

3.H. Huang, H. Yin, G. Min, H. Jiang, J. Zhang, and Y. Wu, "Data-driven information plane in software-defined networking," *IEEE Communications Magazine*, vol. 55, pp. 218-224, 2017.

4.D. Levin, M. Canini, S. Schmid, F. Schaffert, and A. Feldmann, "Panopticon: Reaping the Benefits of Incremental {SDN} Deployment in Enterprise Networks," in *2014 {USENIX} Annual Technical Conference ({USENIX}{ATC} 14)*, 2014, pp. 333-345.

5.N. Feamster, J. Rexford, and E. Zegura, "The road to SDN: an intellectual history of programmable networks," *ACM SIGCOMM Computer Communication Review*, vol. 44, pp. 87-98, 2014.

6.*OpenFlow Switch Specification Version 1.5.1*. Available: https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-switch-v1.5.1.pdf

7.B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, and future of programmable networks," *IEEE Communications Surveys & Tutorials*, vol. 16, pp. 1617-1634, 2014.

8.A. Ominike Akpovi, A. Adebayo, and F. Osisanwo, "Introduction to Software Defined Networks (SDN)," 2016.

9.W. Stallings, "Software-defined networks and openflow," *The internet protocol Journal*, vol. 16, pp. 2-14, 2013.

10.G. Wang, Y. Zhao, J. Huang, and W. Wang, "The controller placement problem in software defined networking: A survey," *IEEE Network*, vol. 31, pp. 21-27, 2017.
11.B. Heller, R. Sherwood, and N. McKeown, "The controller placement problem," in *Proceedings of the first workshop on Hot topics in software defined networks*, 2012, pp. 7-12.
12.G. Yao, J. Bi, Y. Li, and L. Guo, "On the capacitated controller placement problem in software defined networks," *IEEE Communications Letters*, vol. 18, pp. 1339-1342, 2014.
13.S. Khuller and Y. J. Sussmann, "The capacitated k-center problem," *SIAM Journal on Discrete Mathematics*, vol. 13, pp. 403-418, 2000.
14.Y. Hu, T. Luo, W. Wang, and C. Deng, "On the load balanced controller placement problem in Software defined networks," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, 2016, pp. 2430-2434.
15.L. Han, Z. Li, W. Liu, K. Dai, and W. Qu, "Minimum control latency of SDN controller placement," in *2016 IEEE Trustcom/BigDataSE/ISPA*, 2016, pp. 2175-2180.
16.Y. Zhang, N. Beheshti, and M. Tatipamula, "On resilience of split-architecture networks," in *2011 IEEE Global Telecommunications Conference-GLOBECOM 2011*, 2011, pp. 1-6.
17.Y.-N. Hu, W.-D. Wang, X.-Y. Gong, X.-R. Que, and S.-D. Cheng, "On the placement of controllers in software-defined networks," *The Journal of China Universities of Posts and Telecommunications*, vol. 19, pp. 92-171, 2012.
18.L. F. Müller, R. R. Oliveira, M. C. Luizelli, L. P. Gaspary, and M. P. Barcellos, "Survivor: An enhanced controller placement strategy for improving SDN survivability," in *2014 IEEE Global Communications Conference*, 2014, pp. 1909-1915.
19.Y. Jimenez, C. Cervelló-Pastor, and A. J. García, "On the controller placement for designing a distributed SDN control layer," in *2014 IFIP Networking Conference*, 2014, pp. 1-9.
20.Y. Hu, W. Wang, X. Gong, X. Que, and S. Cheng, "On reliability-optimized controller placement for software-defined networks," *China Communications*, vol. 11, pp. 38-54, 2014.
21.M. Tanha, D. Sajjadi, and J. Pan, "Enduring node failures through resilient controller placement for software defined networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1-7.
22.A. Sallahi and M. St-Hilaire, "Optimal model for the controller placement problem in software defined networks," *IEEE communications letters*, vol. 19, pp. 30-33, 2015.
23.H. K. Rath, V. Revoori, S. M. Nadaf, and A. Simha, "Optimal controller placement in Software Defined Networks (SDN) using a non-zero-sum game," in *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, 2014, pp. 1-6.
24.A. Ruiz-Rivera, K.-W. Chin, and S. Soh, "GreCo: An energy aware controller association algorithm for software defined networks," *IEEE communications letters*, vol. 19, pp. 541-544, 2015.
25.M. T. I. ul Huque, G. Jourjon, and V. Gramoli, "Revisiting the controller placement problem," in *2015 IEEE 40th Conference on Local Computer Networks (LCN)*, 2015, pp. 450-453.
26.A. Sallahi and M. St-Hilaire, "Expansion model for the controller placement problem in software defined networks," *IEEE Communications Letters*, vol. 21, pp. 274-277, 2017.
27.D. Hock, M. Hartmann, S. Gebert, M. Jarschel, T. Zinner, and P. Tran-Gia, "Pareto-optimal resilient controller placement in SDN-based core networks," in *Proceedings of the 2013 25th International Teletraffic Congress (ITC)*, 2013, pp. 1-9.
28.S. Lange, S. Gebert, T. Zinner, P. Tran-Gia, D. Hock, M. Jarschel, *et al.*, "Heuristic approaches to the controller placement problem in large scale SDN networks,"

IEEE Transactions on Network and Service Management, vol. 12, pp. 4-17, 2015.
29.D. Tuncer, M. Charalambides, S. Clayman, and G. Pavlou, "Adaptive resource management and control in software defined networks," *IEEE Transactions on Network and Service Management*, vol. 12, pp. 18-33, 2015.
30.J. Liao, H. Sun, J. Wang, Q. Qi, K. Li, and T. Li, "Density cluster based approach for controller placement problem in large-scale software defined networkings," *Computer Networks*, vol. 112, pp. 24-35, 2017.
31.B. Zhang, X. Wang, L. Ma, and M. Huang, "Optimal controller placement problem in Internet-oriented software defined network," in *2016 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 2016, pp. 481-488.
32.G. Wang, Y. Zhao, J. Huang, and Y. Wu, "An effective approach to controller placement in software defined wide area networks," *IEEE Transactions on Network and Service Management*, vol. 15, pp. 344-355, 2017.
33.M. F. Bari, A. R. Roy, S. R. Chowdhury, Q. Zhang, M. F. Zhani, R. Ahmed, *et al.*, "Dynamic Controller Provisioning in Software Defined Networks," in *CNSM*, 2013, pp. 18-25.
34.Y. Hu, T. Luo, N. C. Beaulieu, and C. Deng, "The energy-aware controller placement problem in software defined networks," *IEEE Communications Letters*, vol. 21, pp. 741-744, 2017.
35.G. Ishigaki and N. Shinomiya, "Controller placement algorithm to alleviate burdens on communication nodes," in *2016 International Conference on Computing, Networking and Communications (ICNC)*, 2016, pp. 1-5.
36.*SDN-Enabled Programmatic Control of the Network*. Available: http://www.brocade.com/en/backend-content/pdf-page.html?/content/dam/common/documents/content-types/solution-brief/brocade-mlx-service-provider-sb.pdf
37.*Corsa's DP2100 SDN switching and routing platform*. Available: http://www.corsa.com/products/dp2100/
38.R. Daniels and D. Whittaker. (2015). *Benchmarking the SDN Switch*. Available: https://www.opennetworking.org/images/stories/sdn-solution-showcase/germany2015/Spirent%20-%20Benchmarking%20the%20SDN%20Switch.pdf
39.C. Veness, "Calculate distance and bearing between two Latitude/Longitude points using Haversine formula in JavaScript," *Movable Type Scripts*, 2011.
40.G. Wang, Y. Zhao, J. Huang, and R. M. Winter, "On the data aggregation point placement in smart meter networks," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, 2017, pp. 1-6.
41.S. Skiena, "Dijkstra's algorithm," in *Implementing discrete mathematics: combinatorics and graph theory with mathematica*, ed: Addison-Wesley Reading, MA, 1990, pp. 225-227.
42.S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *IEEE Journal on Selected Areas in Communications*, vol. 29, pp. 1765-1775, 2011.
43.(2019, 02/05/2019). *The Internet Topology Zoo*. Available: http://www.topology-zoo.org/

سال یازدهم، شماره‌های ۳۹ و ۴۰، بهار و تابستان ۱۳۹۸

صص: ۷۳-۹۰

فصلنامه علمی- پژوهشی

فناوری اطلاعات و ارتباطات ایران

ارائه الگوی معماری دانش منابع انسانی در سازمان‌های دانش‌بنیان با استفاده از رویکرد آمیخته

*عبداله ساعدی **رضا سپهوند ***سید نجم الدین موسوی ****محمد حکاک

*دانشجوی دکتری مدیریت منابع انسانی، دانشگاه لرستان، خرم آباد، ایران

** استاد گروه مدیریت بازرگانی، دانشگاه لرستان، خرم آباد، ایران

***دانشیار گروه مدیریت بازرگانی، دانشگاه لرستان، خرم آباد، ایران

**** دانشیار گروه مدیریت بازرگانی، دانشگاه لرستان، خرم آباد، ایران

تاریخ دریافت: ۱۳۹۸/۰۴/۰۷ تاریخ پذیرش: ۱۳۹۹/۰۴/۲۵

چکیده

معماری دانش منابع انسانی بیش از هر ابزار یا عنصر دیگری در شیوهٔ ایجاد، سازماندهی کردن، اندوختن، توزیع و کاربرد دانش برای دستیابی به اهداف سازمانی ضروری است. پژوهش حاضر باهدف ارائه الگوی معماری دانش منابع انسانی در سازمان‌های دانش‌بنیان با استفاده از رویکرد آمیخته انجام پذیرفت. این پژوهش بر پایه پژوهش‌های آمیخته و به‌صورت کمی و کیفی است که ازنظر هدف، کاربردی و حیث ماهیت و روش، توصیفی پیمایشی است. جامعه آماری پژوهش حاضر را سازمان‌های دانش‌بنیان استان لرستان تشکیل می‌دهد که ۳۰ نفر از خبرگان آن‌ها بر اساس اصل کفایت نظری و با استفاده از روش نمونه‌گیری هدفمند انتخاب‌شده‌اند. ابزار گردآوری اطلاعات در بخش کیفی پژوهش، مصاحبه نیمه ساختاریافته و در بخش کمی نیز پرسشنامه است. در بخش کیفی، داده‌ها و اطلاعات با استفاده از نرم‌افزار Atlas.ti و روش کدگذاری تحلیل و شاخص‌های معماری دانش منابع انسانی شناسایی شدند. در بخش کمی پژوهش، با استفاده از نرم‌افزار Matlab و روش مدل‌سازی ساختاری تفسیری مدل نهایی پژوهش تدوین و ارائه‌شده است. نتایج پژوهش دربرگیرنده شاخص‌ها و مؤلفه‌های معماری دانش منابع انسانی و ارائه مدل معماری دانش منابع انسانی در سازمان‌های دانش-بنیان است. بدین ترتیب، یافته‌ها علاوه بر تدوین مدل معماری دانش منابع انسانی، حاکی از شناسایی مؤلفه‌های اصلی معماری دانش منابع انسانی، زیرساخت‌های مدیریت دانش، ویژگی‌های حرفه‌ای، ویژگی‌های موقعیتی و دستاوردهای معماری دانش منابع انسانی است.

واژه‌های کلیدی: معماری دانش، معماری دانش منابع انسانی، رویکرد آمیخته، سازمان‌های دانش‌بنیان

۱- مقدمه

تحولات هزاره سوم دیگر به سازمان‌ها اجازه نخواهد داد تا در مقابل فشارهای ناشی از مهارت‌ها، توانایی‌ها و تکنولوژی‌های رقیب به رویکردها و روش‌های سنتی امیدوار باشند. راهکارهایی لازم است تا از طریق آنها فرصت مناسبی برای بهسازی عملکرد منابع انسانی، ارتقای تصمیم‌گیری، بهبود اثربخشی و کارایی سازمانی، افزایش روند خلاقیت و نوآوری و تسریع خدمت رسانی به ارباب رجوع را در اختیار سازمان قرار دهد [۱]. مسلماً آینده پذیرای سازمان‌هایی خواهد بود که سهم بیشتری از دانش را به خود اختصاص دهند، نه سهم بیشتری ازمنابع طبیعی. چرا که دانش کشف کننده منابع دیگر است و توان و شایستگی سازمان را ابدی و لایزال خواهد ساخت. در عصر حاضر مزیت اصلی برای رقابت

نویسنده عهده‌دار مکاتبات : رضا سپهوند sepahvand.re@lu.ac.ir

میان سازمان‌ها در سرمایه دانشی[1] آنها نهفته است و همه آنها به این امر واقف و آگاهند که اساسی‌ترین عامل برتری سازمان، نه در دارایی‌های انباشته ناشی از کارخانجات و یا حتی بازارهای بزرگ پیش روی آنهاست، بلکه در دانش و جریان-های آن قرار دارد. در این راستا، به خوبی یقین دارند که با وجود نرخ بالای تغییرات در محیط کسب‌وکار کنونی، دارایی-های ملموس (همچون سرمایه، زمین، مواد اولیه و ...) پیشتازی آنها در صحنه رقابت را رقم نخواهد زد [۲]. دانش سرمایه انسانی یکی از اصلی‌ترین و حیاتی‌ترین منابع راهبردی جهت پیش برد اهداف سازمان و درواقع تنها منبع قابل‌اتکا برای فتح بازارهای جدید به شمار می‌رود [۳]. بدیهی است که درک این واقعیت‌ها ما را به سمت مفاهیم و رویکردهای جدیدی به‌منظور مشخص ساختن مکان و چگونگی کسب و تبادل دانش برای واکنش مفید و مؤثر به تغییرات و فرصت‌های حاصل از آن رهنمون می‌سازد. این رویکرد در ادبیات مدیریت تحت عنوان معماری دانش منابع انسانی[2] از آن یاد می‌شود. پارادایمی که با راه‌حلی ارزش افزا و آینده‌ساز در حوزه دانش به‌منظور بقای سازمان به یک ضرورت مبدل گشته است. معماری دانش منابع انسانی با مجموعه‌ای منطقی از اصول و استانداردها مهندسی زیرساخت مدیریت دانش کارکنان را هدایت می‌کند. درواقع نحوه کسب، خلق، سازمان‌دهی، ذخیره، اشتراک، و به‌کارگیری دانش را برای سرمایه انسانی مشخص می‌کند [۴].

بدون شک پیچیدگی‌های عصر اطلاعات، هزاره جدید را جولانگاه تاخت‌وتاز سازمان‌هایی خواهد کرد که قابلیت‌های متحیرکننده‌ی دانش را به خدمت درآورده و با استفاده از آن در ماراتن نفس‌گیر رقابت جهانی، موفق‌تر عمل نمایند [۵]. باوجوداین، بسیاری از سازمان‌ها به نیک دریافته‌اند که بهره‌برداری صحیح از دانش می‌تواند بزرگ‌ترین و مهم‌ترین سلاح راهبردی در راستای بازآفرینی و نوسازی استراتژی‌ها باشد تا راهی نو در فرایند تولید ثروت را برای آن‌ها فراهم آورد. همین امر اهمیت روزافزون مشخص کردن ساختار پایگاه دانش، نحوه به هم پیوستن اجزای دانش، چگونگی تبدیل و تحول اطلاعات به دانش را بیش‌ازپیش برای سازمان‌ها نمایان ساخته است [۶]. در محیط کسب‌وکار امروزی، دانش و خردمندی تنها راه ایجاد ثروت در سازمان‌ها و جوامع تلقی می‌گردد. سرمایه‌های انسانی و طبیعی نیز زمانی به ثروت تبدیل می‌شوند که با دانش آمیخته و عجین شده باشند. تلاش جهت دستیابی به دانش برای سازمان‌های با فعالیت دانشی، هر روز پر رنگ‌تر می‌شود. زیرا دانش دارایی اصلی آن-هاست و مدیران و رهبران سازمان برای استفاده مؤثر و کارآمد از این دارایی باید بتوانند آن را مدیریت نموده و به منظور ارتقاء بهره‌وری آن را بکار گیرند. به بیانی دیگر می باید دانش موجود به نحوی سامان‌دهی شود و در اختیار افراد قرارگیرد که بتوان از همه پتانسیل‌های موجود در راستای پیش برد اهداف و مقاصد سازمان استفاده نمود. بنابراین، مسئله مهم و قابل‌بحث در سازمان‌های کنونی به‌ویژه سازمان‌های دانش‌بنیان که نقش بسزایی در هم افزایی علم و ثروت، توسعه اقتصاد دانش محور، تحقق اهداف علمی، اقتصادی، رونق اقتصادی جوامع و جلوگیری از مهاجرت نخبگان دارند، مسئله معماری دانش منابع انسانی و چگونگی کسب، خلق، سازمان‌دهی، ذخیره، اشتراک، و به‌کارگیری دانش است. ازاین‌رو، لازم و ضروری است که مدیران سازمان‌های دانش‌بنیان با شناخت و آگاهی کافی از پدیده معماری دانش منابع انسانی بتوانند با به دست آوردن، اندوختن و توزیع مناسب منابع دانش برای نیل به بصیرت و رسالت سازمان خویش تلاش نماید. بنابراین، با توجه به اینکه رویکرد معماری دانش منابع انسانی به‌طورکلی و به‌طور اخص در سازمان‌های دانش‌بنیان کمتر موردتوجه قرارگرفته است. لذا پژوهش حاضر بر آن است تا با استفاده از رویکرد ساختاری ـ تفسیری مدلی را در جهت معماری دانش منابع انسانی تدوین نماید.

۱- مروری بر مبانی نظری

۱-۲- دانش منابع انسانی

از دهه۱۹۶۰ توجه به دانش، مهارت‌ها و تخصص‌های سرمایه انسانی به‌طور گسترده در ادبیات اقتصاد، حسابداری، استراتژی، مالی و ... موردتوجه قرار گرفت. دانش و مهارت منابع انسانی جزء ارزشمندترین دارایی‌ها و مولد ثروت برای سازمان و جامعه محسوب می‌شود. زیرا [۷] دانش منابع انسانی منبع قدرت و کلید موفقیت بیان شده که برای اطمینان از عملکرد عالی سازمان لازم و ضروری است [۷]. در حقیقت دانش منابع انسانی این قابلیت را در فرد ایجاد می‌کند که نه تنها با شایستگی لازم در فرایندهای دانشی درگیر می-شود؛ بلکه توانایی تولید، به کارگیری دانش، خلاقیت و نوآوری را در فرایندهای دانش دارند. آنها به خوبی سرمایه دانشی را به درون سازمان انتقال می دهند و از قدرت زیادی در سازمان برخوردارند که تصمیمات آنها تأثیر مهمی بر کل روند فعالیت

1 . Knowledge Capital
2 . Human Resources Knowledge Architecture

سازمان دارد [۸]. همچنین می‌توان اذعان داشت که نیروی کار صاحب دانش این مهارت و توانایی را دارند که بدون وجود سیستم‌ها و نیروهای کنترلی (ناظر) کار و فعالیت خود را انجام دهند. زیرا اعمال محدودیت و ممنوعیت، موجب یأس و سرخوردگی و در مواردی سرکشی آنان در سازمان می‌گردد [۹]. در تعریفی دیگر، منابع انسانی دانش مدار کسانی هستند که برای سازمان بسیار با ارزشند و نقش تعیین کننده‌ای در شکست یا موفقیت سازمان را رقم می‌زند. در واقع می‌تواند بطور مدام سازمان را در حال تولید و مشتریان آن را راضی نگه دارند [۱۰]. دانشی که در اختیار منابع انسانی قرار دارد یا به عبارتی نیروی انسانی که دارای سطح بالایی از مهارت، تجربه و قدرت شناختی بالا؛ ضمن افزایش ارزش سرمایه‌های دیگر می‌توانند تصمیمات، فرایندها و راه‌حل‌های مؤثرتری در هنگام مواجه با مسائل و بحران‌ها برای سازمان ایجاد نماید [۱۱].

۲-۲- معماری دانش منابع انسانی

در عصری که ما زندگی می‌کنیم دانش همواره یکی از مهترین عوامل کسب موفقیت برای فرد و سازمان تلقی می‌گردد. در واقع دانش است که در آینده مسیر حرکت فرد و سازمان را مشخص و جایگاه آن را درمحیط رقابتی تعیین خواهد کرد [۱۲]. بدیهی است با افزایش ارزش دانش و اهمیت روزافزون کارکنان دانشی، منطقی به نظر می رسد تا با استفاده از معماری دانش منابع انسانی سازمان بتوان فرصتی برای ایجاد مزیت رقابتی خلق کرد [۱۳]. معماری دانش منابع انسانی در وهله اول قابلیت پیدا شدن[3] و قابل استفاده بودن دانش[4] را نمایان می‌سازد. به این ترتیب، در معماری دانش منابع انسانی فهرست کردن، طبقه بندی، کدگذاری و تعیین محل کردن به نحوی انجام می‌شود که دانش به راحتی در دسترس و مورد استفاده قرار می‌گیرد [۱۴]. [۱۵] معماری دانش اصطلاحاً از مدل‌سازی دانش نشأت می‌گیرد و معتقد است که معماری دانش منابع انسانی با ارائه چارچوبی اطلاعات خام نیروی کار را به دانش عملی تبدیل می‌کند. درواقع با طراحی سیستمی تمام داده‌ها و اطلاعات پراکنده‌شده را به هم مرتبط و آن‌ها را به یک ابزار قدرتمند تبدیل می‌کند [۱۶]. معماری یا مهندسی دانش را می‌توان ابزاری مهم برای ارزیابی دانش ضمنی نیروی انسانی، چارچوبی یکپارچه برای تبیین، استنتاج و حفظ دانش کارکنان در راستای اهداف استراتژیک سازمان و یا به‌عبارتی‌دیگر، مفهومی که ارتباط بین مأموریت، اهداف و مقاصدی که در طی برنامه به‌صورت مجزا درآمده را برقرار می‌کند؛ تعریف کرد [۱۷]. معماری کردن دانش کارکنان باعث می شود که مؤلفه‌ها، روابط درونی، روابط بیرونی و سیاست های طراحی دانش به وضوح بیان شود. رابطه بین مدیریت دانش و معماری دانش منابع انسانی را اینگونه بیان می‌کنند که مدیریت دانش به شکلی نظامند و سیستماتیک امکان مدیریت دانش آشکار مورد نیاز سازمان و تمامی فرایندهای مورد نیاز جهت خلق، سامان دهی، تسهیم، کاربرد و جاری سازی دانش در راستای اهداف سازمانی را فراهم می‌آورد. اما در مقابل، معماری دانش منابع انسانی برای رخ دادن فرایندهای مدیریت دانش مورد استفاده قرار می‌گیرد. در واقع در معماری دانش منابع انسانی هدف چگونگی و چرایی رخ دادن فرایندهای مدیریت دانش است [۱۸]. در جدول شماره یک برخی از تعاریف معماری دانش منابع انسانی ارائه‌شده است.

جدول۱- تعاریف معماری دانش منابع انسانی

ردیف	تعاریف	منبع
۱	چگونگی و نحوه سامان دهی، کدگذاری و شیوه های دستیابی به دانش را فراهم می آورد	[۱۹]
۲	ارزیابی‌های منظم و دقیق دانش برای تعیین اینکه چه دانشی در کجای سازمان ذخیره‌شده است.	[۲۰]
۳	اقدامی عملی در خصوص چیدمان دانش به گونه ای که قابل درک باشد	[۲۱]
۴	ایفای نقش اصلی در حمایت از فرایندهای عملیاتی و مدیریتی در سازمان	[۲۲]
۵	طراحی چارچوبی خاص برای دانش کارکنان به‌منظور کسب منافع رقابتی	[۲۳]
۶	تغییر در شکل و ماهیت دانش، نحوه ایجاد و بکارگیری آن برای سود آوری سازمان	[۲۴]

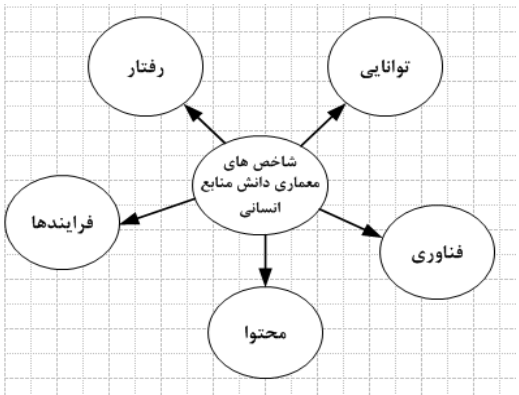
شاید بتوان گفت پایه و اساس معماری دانش منابع انسانی، شناخت طبعیت و ماهیت دوگانه دانش یعنی دانش پنهان و آشکار است. کانون توجه این رویکرد آن است که دانشی که عمدتاً حاصل تجربه‌های واقعی فرد است را به طور شایسته‌ای سازماندهی و معماری نماید [۲۵]. با توجه به مطالعات و پژوهش‌های صورت گرفته در خصوص ابعاد معماری دانش منابع انسانی، در این پژوهش تشریح ابعاد معماری دانش منابع انسانی از مدل چوران[5] (۲۰۰۱) و اسنایمن و کروگر[6] (۲۰۰۴)

3 . Findable
4 . Usable
5 . Chevron
6 . Snyman & Kruger

استفاده شده است. دلیل این امر آن است که مدل محققین مذکور راهنمای عمل پژوهش حاضر قرار گرفته و این ابعاد بر مبنای دیدگاه خبرگان پایه و اساس مدل پژوهش حاضر نیز قرار گرفت. بنابراین، در شکل شماره یک ابعاد معماری دانش منابع انسانی ارائه و هر یک به‌طور مختصر تشریح شده است.

توانایی منابع انسانی[۷]: توانایی منابع انسانی به سطح بالایی از تحصیلات، تجربه و مهارت‌های نظری و تحلیلی اطلاق می‌شود [۲۶]. درواقع توانایی و قابلیت این‌گونه کارکنان در سازمان بادانش فنی و عملی بالا، کسب دانش و به‌کارگیری آن، مهارت‌های ارتباطی، انگیزش، توانایی کشف فرصت‌ها و رویارویی با چالش‌ها، قدرت تجزیه‌وتحلیل و ترکیب و تفسیر داده‌ها و اطلاعات شناخته می‌شوند. به عبارتی این شاخص بیان می‌کند که افراد با این مهارت‌ها و قابلیت هایی که دارند بهتر می‌توانند دانش و اطلاعات را مورد تجزیه و تحلیل، تفسیر قرار دهند و سازمان را در شناسایی و جذب دانش مهم و ضروری یاری رسانند [۲۷].

رفتار منابع انسانی[۸]: اشاره به رفتارهایی دارد که نیروی انسانی در محیطی که فرایندهای مدیریت دانش در آن رخ می‌دهد؛ از خود بروز می‌دهند [۲۸]. این بدان معناست که کارکنان به چه طریقی در فرآیندهای مدیریت دانش رفتار می‌کنند؟ فرد برای کمک به موفقیت سازمانی می تواند در قالب سخنرانی، مقاله و ... دانش خود را در اختیار دیگران قرار دهد تا آنها نیز از دانش منتفع شوند. از طرفی فردی که به دنبال دانش است با دریافت دانش آن را با آموخته های خود مقایسه می‌کند و به کاستی‌های خود پی برده و آن را جبران می‌نماید این عمل انتقال و دریافت باعث بازسازی دانش می‌شود.



شکل ۱- ابعاد معماری دانش منابع انسانی [۲۴] و [۲۶].

فرایند (شیوه عمل) منابع انسانی[۹]: فرایندها دربرگیرنده چگونگی انجام کار می‌باشند. به عبارتی فرایندها چرایی و چگونگی انجام فعالیت‌های منابع انسانی را بیان می‌کند. چرا کارکنان اقدام به کسب، شناسایی و یا تسهیم دانش می‌کنند؟ بدین ترتیب، کارکنان دانشی با چگونگی یا شیوه عمل خود (فرایندها) سازمان را در دستیابی به اهداف و مأموریت‌ها یاری می‌رسانند [۲۶]. درواقع منابع انسانی دانش‌محور با کمک این فرایندها (شناسایی، خلق، ذخیره‌سازی، تسهیم و ...) رشد و تعالی سازمان را رقم می‌زنند. به‌عنوان‌مثال: کلانتن و همکاران[۱۰] (۲۰۰۲) نتیجه اشتراک یا تسهیم دانش (مبادله دانش افراد با یکدیگر) را پاسخگویی سریع به نیازهای مشتری، شناخت بازار و نوآوری و همچنین تسهیل در تغییرات سازمانی عنوان می‌کنند [۲۷].

فناوری (ابزار مورداستفاده) منابع انسانی[۱۱]: فناوری به‌عنوان یکی از مهم‌ترین اجزای معماری دانش کارکنان محسوب می‌شود. چراکه نقش تسهیل‌کنندگی مهمی را در ذخیره‌سازی، ارتباط، تولید دانش و ... ایفا می‌کند. [۲۸]. به عبارتی فناوری بستر مناسبی را برای بهبود همه فعالیت ها و فرایندهای مدیریت دانش فراهم ساخته و همه انواع روش های تولید دانش (جامعه پذیری، درونی سازی، بیرونی سازی و ترکیب) را تسریع و توسعه می دهد [۲۹].

محتوا (پایگاه دانش) منابع انسانی[۱۲]: یعنی پایگاه دانش مشارکتی که به‌صورت الکترونیکی استخراج‌شده است. به عبارتی محتوا یا مضمون در معماری دانش منابع انسانی به تجمیع و ذخیره‌سازی (پایگاه دانش) داده‌های تولید شده با استفاده از تکنولوژی اشاره دارد. به‌این‌ترتیب، جمع‌آوری و تجمیع دانش کارکنان به سازمان کمک خواهد کرد تا با بهره‌مندی از داده‌های ساخت‌یافته بتوانند از تجربیات و دانش به‌دست‌آمده مجدداً استفاده نمایند [۲۴]. بزرگ‌ترین و مهم‌ترین پایگاه دانش (محتوا) توزیع‌شده را، دانش انباشته‌شده در ذهن کارکنان عنوان می‌کند. و معتقد است اگر این دانش ضمنی به مدت طولانی در اختیار افراد و گروه‌ها باقی بماند؛ سازمان نمی‌تواند بر ارزش خود بیفزاید [۳۰].

۲-۳- پیشینه پژوهش

پژوهش عبدالهی و حسین زاده (۱۳۹۷) نشان داد که معماری دانش کارکنان تأثیر مثبت و معناداری بر

7 . Human Resource Ability
8 . Human Resource Behavior
9 . Human Resource Process
10 . Calantone et al
11 . Human Resource Technology
12 . Human Resource Content

هویت‌سازمانی، توانمندسازی روان‌شناختی آنان دارد. ورعی و همکاران (۱۳۹۵) در انجام پژوهشی ضمن بررسی ابعاد گوناگون مقوله معماری دانش، چارچوبی را جهت معماری دانش درسازمان‌های کلان‌مقیاس پیشنهاد نموده که نتایج آن حکایت از سودمندی این چارچوب جهت مدیریت دانش در سازمان‌های کلان‌مقیاس را دارد. آن‌ها بیان داشتند که یافته‌های این پژوهش می‌تواند به طراحان جهت معماری دانش سازمان‌های متبوع یاری رساند. عابدینی بلترک و نیلی (۱۳۹۳) در پژوهش خود از معماری دانش به‌عنوان نوعی فلسفه یادگیری یاد می‌کنند که برساختن دانش توسط یادگیرندگان به‌صورت فردی و اجتماعی اشاره دارد که یادگیرندگان دانش خود را بر اساس طرح‌واره‌ها یا عقاید موجود می‌سازند. محمدی فاتح و جوکار (۱۳۹۰) در پژوهش خود دریافتند که معماری دانش کارکنان رویکردی است که سازمان را در جایی قرار می‌دهد که به اهداف سازمانی خود دست یابد. همچنین نشان دادند که معماری دانش کارکنان درجه اطمینان سازمان را در انجام فرایندهای مدیریت دانش افزایش می‌دهد. بات و همکاران (۲۰۱۸) در پژوهشی معماری را معادل سامان‌دهی دانش کارکنان قلمداد می‌کنند. آنان نشان دادند که معماری دانش باعث می‌شود دانش کارکنان به روز رسانی شوند و آنها همواره در انجام وظایف و مسئولیت‌های خود بیشترین توان و قابلیت را داشته و نیرویی خلاق و نوآور برای سازمان خود باشند. روزیکیک و میکیک[13] (۲۰۱۷) نیز در پژوهش خود معماری دانش منابع انسانی را چارچوبی منسجم و یکپارچه برای استخراج و ذخیره‌سازی دانش کارکنان قلمداد می‌کنند و معتقدند این رویکرد ابزاری مناسب جهت تحقق اهداف استراتژیک سازمان، کیفیت و سرعت پاسخگویی و همچنین افزایش انطباق‌پذیری در برابر تغییرات محیطی می‌باشد. یافته های گنت و همکاران (۲۰۱۷) نیز نشان می‌دهد که معماری دانش کارکنان موجب می‌شود مسائل سازمانی از طریق به ایجاد دانش جدید، حداقل رساندن افزونگی و دوباره کاری‌ها حل شوند. نتایج پژوهش دراگنی و همکاران[14] (۲۰۱۷) حاکی از آن بود که به‌کارگیری معماری دانش به‌طور مستقیم و غیرمستقیم ساختارهای پیچیده و پویا اطلاعاتی ایجاد کرده و این قابلیت را به سازمان می‌دهد تا در راستای بهبود فعالیت‌ها و حمایت از تصمیمات به نحوه مطلوبی استفاده نماید. همچنین ناواک (۲۰۱۷) با توجه به نتایج پژوهش خود بیان می‌کند که معماری دانش منابع انسانی رویکردی است که از طریق آن سازمان دانش کارکنان خود را مدام رصد کرده، نقد و مورد بررسی قرار می‌دهد. تا در صورت لزوم و با استفاده از سازکارهای مناسب، زمینه را برای تغییر، کسب، سامان‌دهی، ذخیره‌سازی و کاربرد دانش موجود و جدید فراهم آورد. فرناندز (۲۰۱۳) با انجام مطالعه‌ای اظهار داشت که شناخت طبعیت و ماهیت دوگانه دانش یعنی دانش پنهان و آشکار اساس و مبنای معماری دانش منابع انسانی است. پژوهش سندکوه[15] (۲۰۱۵) نشان داد که معماری دانش کارکنان ضمن افزایش قابلیت و انعطاف‌پذیری سازمان در برابر تغییرات محیطی، بهبود بهره-وری سازمان را نیز در پی دارد. یافته‌های گو و ژانگ (۲۰۱۲) حکایت از آن دارد که معماری دانش کارکنان میزان همراستا سازی فعالیت‌ها با مأموریت و اهداف سازمانی را افزایش داده و قابلیت سازمان را در واکنش به تغییرات محیطی افزایش می-دهد. اورز[16] (۲۰۰۸) در پژوهش خود نشان داد که معماری دانش با ارائه چارچوبی ساختاریافته جمع‌آوری دانش را در راستای موفقیت در بازارهای رقابتی تسهیل می‌نماید. ال کادی[17] (۲۰۰۳) در پژوهش خود بیان داشت که در صورت ایجاد و پیاده‌سازی معماری دانش کارکنان، دانش موجود به‌طور مدام در دسترس سازمان قرارگرفته و سازمان این قابلیت را برای توسعه و بازیابی دانش خود کسب خواهد کرد. کریستسکو وکریستسکو (۲۰۰۸) در تحقیقی بیان داشتند زمانی که سازمان رویکرد معماری دانش را به کار می‌گیرد بهتر می‌تواند منابع مالی خود را برای انجام فعالیت‌ها تخصیص دهد. همچنین تغییر، ذخیره و بکارگیری دانش کسب شده، بهتر عمل می‌کند. لانه و همکاران (۲۰۰۶) نشان دادند که معماری دانش کارکنان، سازمان را در امر فهرست کردن، طبقه بندی، کدگذاری و تعیین محل دانش کمک می‌کند و همین امر میزان دسترسی به دانش را افزایش می‌دهد. مورگانوالپ و ساج (۲۰۰۳) برای توصیف معماری دانش کارکنان در پژوهشی بیان کردند که این رویکرد یکی از مهمترین ابزارها برای فهرست‌بندی کردن دانش درسازمان است.

13 . Ruzicic & Micic
14 . Dragoni et al
15 . Sandkuh
16 . Evers
17 . Al-Kadi

۲-روش‌شناسی پژوهش

پژوهش حاضر رویکردی آمیخته و به‌صورت کمی و کیفی است. که ازنظر هدف، کاربردی و از حیث ماهیت و روش، توصیفی پیمایشی است. جامعه آماری پژوهش را سازمان‌های دانش‌بنیان در استان لرستان تشکیل می‌دهد که با توجه به اطلاعات به‌دست‌آمده تعداد آن‌ها برابر ۱۶شرکت است که در حوزه‌های صنعتی، خدماتی، تولیدی و داروسازی مشغول به فعالیت می‌باشند. بدین ترتیب، جامعه آماری پژوهش خبرگان متشکل از مدیران ارشد و میانی شرکت‌های دانش‌بنیان استان لرستان می‌باشند که با استفاده از روش نمونه‌گیری هدفمند و بر اساس حوزه فعالیت تعداد۳۰ نفر به‌عنوان اعضای نمونه انتخاب‌شده است. بدین شکل که از هر حوزه (صنعتی، خدماتی، تولیدی و داروسازی) بر اساس اصل کفایت نظری، تا سرحد اشباع داده‌ها و اطلاعات موردنیاز جمع‌آوری گردید. ابزاری گردآوری اطلاعات در بخش کیفی پژوهش مصاحبه نیمه ساختاریافته است که روایی و پایایی آن به ترتیب با استفاده از ضریب CVR و آزمون کاپای ـ کوهن تأیید گردید. همچنین ابزار گردآوری اطلاعات در بخش کمی پرسشنامه است که روایی و پایایی آن به ترتیب با استفاده از روایی محتوا و آزمون مجدد تأیید شد. در بخش کیفی پژوهش، داده‌های به‌دست‌آمده از مصاحبه با استفاده از نرم‌افزار Atlas.ti و روش کدگذاری تحلیل و مؤلفه‌های معماری دانش منابع انسانی شناسایی شدند. لازم به ذکر است که داده‌ها بر اساس شیوه کدگذاری باز، محوری و انتخابی و درنهایت ایجاد مقوله‌ها و مؤلفه‌ها تحلیل شد. همچنین در بخش کمی پژوهش، با استفاده از رویکرد مدل‌سازی ساختاری تفسیری (ISM) مدل نهایی پژوهش تدوین و ارائه‌شده است. رویکرد مدل‌سازی ساختاری تفسیری به ایجاد و جهت دادن به روابط پیچیده کمک می کند. این روش نه تنها بینشی را در خصوص روابط در بین عناصرمختلف یک سیستم فراهم می سازد بلکه ساختاری را مبتنی بر اهمیت و یا تأثیرگذاری عناصر بر هم ایجاد ساخته و نمایشی تصویری ارئه می‌دهد و همچنین بر مبنای قضاوت گروهی از افراد تعیین می کند که آیا روابطی میان این عناصر وجود دارد یا خیر [۳۱]. از این رو در پژوهش حاضر برای مشخص کردن روابط میان مؤلفه‌های معماری دانش منابع انسانی و ارائه مدلی جامع و منسجم بر مبنای قضاوت افراد از از رویکرد مدل‌سازی ساختاری تفسیری (ISM) استفاده شده است. همچنین مقدار روایی و پایایی پژوهش در بخش کیفی در جدول شماره دو نشان داده شده است.

جدول۲. روایی و پایایی پژوهش در بخش کیفی

روایی		پایایی	
ابزار مورد استفاده	مقدار	ابزار مورد استفاده	مقدار
ضریب CVR	۰/۴۹	کاپای ـ کوهن	۰/۷۸

۳-یافته‌های پژوهش

۳-۱- الف) یافته‌های جمعیت شناختی

در جدول شماره سه ویژگی‌های جمعیت شناختی اعضای نمونه موردبررسی ارائه‌شده است.

جدول۳- ویژگی‌های جمعیت شناختی نمونه

حوزه فعالیت	سن	تعداد	حوزه فعالیت	سن	تعداد
خدماتی	۴۰-۳۰	۳	تولیدی	۴۰-۳۰	۳
	۵۰-۴۱	۵		۵۰-۴۱	۴
	۵۱ به بالا	۱		۵۱ به بالا	۱
	تحصیلات			تحصیلات	
	لیسانس	۱		لیسانس	۲
	فوق‌لیسانس	۵		فوق‌لیسانس	۳
	دکتری	۳		دکتری	۳
	سابقه کار			سابقه کار	
	کمتر از ۱۰	۱		کمتر از ۱۰	۱
	۲۰-۱۰	۶		۲۰-۱۰	۴
	بیشتر از ۲۰	۲		بیشتر از ۲۰	۳
	معیارانتخاب			معیار انتخاب	
	تحصیلات مرتبط	۹			

	سابقه کار بالا			تحصیلات مرتبط	۸
	اشراف اطلاعاتی			سابقه کار بالا	
				اشراف اطلاعاتی	
صنعتی	سن		داروسازی	سن	
	۴۰-۳۰	۳		۴۰-۳۰	۳
	۵۰-۴۱	۲		۵۰-۴۱	۴
	۵۱ به بالا	۱		۵۱ به بالا	۰
	تحصیلات			تحصیلات	
	لیسانس	۲		لیسانس	۱
	فوق‌لیسانس	۳		فوق‌لیسانس	۳
	دکتری	۱		دکتری	۳
	سابقه کار			سابقه کار	
	کمتر از ۱۰	۰		کمتر از ۱۰	۱
	۲۰-۱۰	۴		۲۰-۱۰	۴
	بیشتر از ۲۰	۲		بیشتر از ۲۰	۲
	معیار انتخاب			معیارانتخاب	
	تحصیلات مرتبط	۶		تحصیلات مرتبط	۷
	سابقه کار بالا			سابقه کار بالا	
	اشراف اطلاعاتی			اشراف اطلاعاتی	

۳-۲- ب) یافته‌های بخش کیفی

در این بخش با استفاده ازنظر خبرگان مجموعه‌ای از شاخص‌ها و مؤلفه‌های معماری دانش منابع انسانی شناسایی شد. لازم به ذکر است که شیوه استخراج شاخص‌ها و مؤلفه‌های معماری دانش منابع انسانی با بررسی متون مصاحبه‌ها با کمک نرم‌افزار Atlas.ti و روش کدگذاری شاخص‌ها و مؤلفه‌های معماری دانش منابع انسانی در جدول شماره چهار نشان داده‌شده است.

جدول۴- شاخص‌ها و مؤلفه‌های معماری دانش منابع انسانی

کدهای انتخابی	کدهای محوری	کدهای باز
دستاوردهای معماری دانش منابع انسانی	سازمانی چابک، سازمانی یادگیرنده، سازمانی با مزیت رقابتی پایدار، سازمانی پیشرو	کیفیت و سرعت پاسخگویی، انطباق‌پذیری، افزایش مستمر توانایی افراد، ترویج تفکر سیستمی، توانایی انتقال دانش در سازمان، منحصربه‌فرد بودن، ایجاد ارزش برای مشتری، اشتراک‌گذاری اهداف یا کارکنان، روحیه بالا کارکنان
مشخصه‌های موقعیتی معماری دانش منابع انسانی	شناخت بهتر بازار و مشتری، افزایش خلاقیت و نوآوری، تسهیل و حفظ ارتباطات، همسویی دانش با اهداف سازمانی	شناسایی دقیق خواسته‌ها و نیازها، درک بهتر نیات رفتاری، کشف چیزهای جدید، به‌کارگیری ایده‌ها، تشخیص فرصت‌ها و تهدیدها، آسان ساختن اتصالات، جریان‌های ارتباطی مناسب و به‌موقع، وجود تعامل و تعاون بین کارکنان، ارائه دانش صحیح در راستای مأموریت‌ها، تناسب دانش با استراتژی‌ها
ویژگی‌های حرفه‌ای معماری دانش منابع انسانی	بازیابی و بهبود دانش، ممیزی دانش، هدایت دانش ضمنی و صریح، پیاده‌سازی موفقیت‌آمیز دانش	توسعه دانش، استخراج دانش، بررسی و سنجش دانش، ارزش‌گذاری دانش، بیرونی سازی دانش، توان به‌کارگیری دانش تجربی کارکنان، رهبری دانش کارکنان در راستای اهداف، اجرای اثربخش دانش، استفاده بهینه از دانش
زیرساخت‌های مدیریت دانش	ساماندهی دانش، کشف دانش، چگونگی کسب، چگونگی انتقال، چگونگی تحول دانش	طبقه‌بندی و دسته‌بندی دانش، یکپارچه‌سازی دانش، شناسایی و ردیابی دانش، آشکارسازی حوزه دانش، ثبت و ضبط دانش کارکنان، گردآوری دانش، اشتراک و تسهیم دانش، تبادل و جابه‌جایی دانش، تبدیل داده به اطلاعات، پالایش دانش
شاخص اصلی معماری دانش	رفتارهای کارکنان، فرایندها (شیوه عمل)،	اقدامات کارکنان در فرایندهای مدیریت دانش، (ذخیره‌سازی،

منابع انسانی	توانایی کارکنان، محتوا، فناوری	تسهیم)، چگونگی فعالیت کارکنان، چرایی اقدامات افراد، تجربه و مهارت‌های نظری، انگیزش بالا پایگاه دانش مشارکتی، دانش ذخیره‌شده، تسهیلگر فرایندهای مدیریت دانش، برقراری ارتباطات شبکه‌ای، اتوماتیک سازی وظایف خاص

۴-۲- ج) یافته‌های کمی پژوهش

۴-۲-۱- تشکیل ماتریس خود تعاملی (SSIM)

در این مرحله پس از شناسایی متغیرها نوبت به واردکردن آن‌ها در ماتریس ساختاری روابط درونی متغیرها می‌شود. این ماتریس یک ماتریس به ابعاد متغیرها می‌باشد که در سطر و ستون آن متغیرها به ترتیب ذکر می‌شود. سپس روابط میان متغیرها در یک مقایسه زوجی به‌وسیله نمادهایی تعیین می‌گردد [۳۲]. درواقع جهت تشکیل ماتریس روابط درونی متغیرها از خبرگان شرکت‌کننده در فرایند ISM خواسته می‌شود، تا بر اساس اصول زیر ماتریس را تکمیل نمایند.

V: عامل سطر (i) می‌تواند بر عامل ستون (j) تأثیرگذار باشد.

A: عامل ستون (j) می‌تواند بر عامل سطر (i) تأثیرگذار باشد.

X: هر دو سطر (i) و ستون(j) بر یکدیگر تأثیر می‌گذارند.

O: بین عنصر سطر و عنصر ستون (i,j) ارتباطی وجود ندارد.

جدول۵- تشکیل ماتریس

متغیرها	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
توانایی کارکنان	V	A	V		O	V	O	V	V	O	V	O	V	V	V	X	V	O		V	A	V
چگونگی کشف دانش	X	A			O	V	O	V	V	O	V	V	V	O	O	O	A	O		X	A		
رفتارهای کارکنان	V				O	V	O	O	O	O	V	O	O	V	V	X	V	O		V			
چگونگی انتقال دانش	A				O	V	O	V	V	O	X	V	V	O	V	A	A	O					
تسهیل و حفظ ارتباطات	O	O			O	V	A	O	X	O	O	A	O	O	O	O	O						
ساماندهی دانش	X	V	O		V	A	V	V	O	V	X	V	V	A	O	O							
فرایندها (شیوه عمل)	O	O	O		V	X	O	O	V	O	V	O	V	X	O								
سازمانی پیشرو	A	X	X		O	A	O	A	O	A	A	O	O	A									
محتوا	O	V	V		A	X	O	V	O	V	V	O	V										
ممیزی دانش	A	O	O		O	A	O	V	V	X	O	X											
بازیابی و بهبود دانش	O	O	V		V	O	V	O	O	V	A												
چگونگی تحول دانش	A	V	V		O	A	O	V	O	V													
هدایت دانش ضمنی و صریح	O	V	V		X	O	V	O	V														
همسویی دانش با اهداف سازمانی	A	O	O		A	O	X	O															
سازمانی چابک	A	V	X		O	A	O																
افزایش خلاقیت و نوآوری	A	O	O		X	O	O																
فناوری	V	O	O		O	V																	
پیاده‌سازی موفقیت‌آمیز دانش	O	V	V		V																		
شناخت بهتر بازار و مشتری	O	O	O																				
سازمانی یادگیرنده	A	V																					
مزیت رقابتی پایدار	A																						
چگونگی کسب دانش																							

۲-۴-۲- ماتریس دریافتی

از تبدیل ماتریس خود تعاملی به یک ماتریس دو ارزشی (صفر و یک)، ماتریس دریافتی یا دستیابی شکل می‌گیرد. به این ترتیب، برای استخراج ماتریس دریافتی باید در هر سطر عدد یک را جایگزین نمادهای X و V و همچنین عدد صفر را جایگزین نمادهای A و O در ماتریس خود تعاملی نمود. بنابراین، ماتریس دستیابی اولیه با توجه با قاعده فوق شکل می‌گیرد. به طور کلی قوانین برای تبدیل ماتریس خود تعاملی به ماتریس دریافتی اولیه به شرح زیر است:

- اگر نماد خانه (i,j) معادل V باشد؛ آنگاه ارزش آن خانه معادل۱ و ارزش خانه قرینه صفر می‌باشد.
- اگر نماد خانه (i,j) معادل A باشد؛ آنگاه ارزش آن خانه معادل صفر و ارزش خانه قرینه ۱ می‌باشد.
- اگر نماد خانه (i,j) معادل X باشد؛ آنگاه ارزش آن خانه و ارزش خانه قرینه ۱ خواهد بود.
- اگر نماد خانه (i,j) معادل O باشد؛ آنگاه ارزش آن خانه و ارزش خانه قرینه صفر خواهد بود.

جدول ۶- ماتریس دریافتی

متغیرها	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
توانایی کارکنان	**۱**	۱	۰	۱	۰	۱	۱	۱	۱	۱	۰	۱	۰	۱	۱	۰	۱	۰	۱	۱	۱	۰
چگونگی کشف دانش	۰	**۱**	۰	۱	۰	۰	۰	۰	۰	۰	۱	۱	۰	۱	۱	۰	۱	۰	۰	۱	۱	۱
رفتارهای کارکنان	۱	۱	**۱**	۱	۰	۱	۱	۱	۱	۰	۰	۱	۰	۰	۰	۰	۰	۱	۱	۰	۰	۰
چگونگی انتقال دانش	۰	۱	۰	**۱**	۰	۰	۰	۱	۰	۱	۱	۱	۰	۱	۱	۰	۰	۰	۰	۱	۱	۰
تسهیل و حفظ ارتباطات	۰	۰	۰	۰	**۱**	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰
ساماندهی دانش	۰	۱	۰	۱	۰	**۱**	۰	۰	۰	۱	۱	۱	۰	۱	۱	۰	۱	۰	۰	۱	۱	۱
فرایندها (شیوه عمل)	۱	۰	۱	۱	۰	۰	**۱**	۰	۰	۱	۰	۱	۰	۱	۰	۰	۱	۱	۰	۰	۰	۰
سازمانی پیشرو	۰	۰	۰	۰	۰	۰	۰	**۱**	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۱	۰
محتوا	۰	۰	۰	۰	۰	۰	۱	۱	**۱**	۱	۰	۱	۱	۰	۰	۱	۰	۰	۰	۱	۱	۰
ممیزی دانش	۰	۰	۰	۰	۰	۰	۰	۰	۰	**۱**	۱	۰	۱	۱	۰	۱	۰	۰	۰	۰	۰	۰
بازیابی و بهبود دانش	۰	۰	۰	۰	۱	۰	۰	۰	۰	۱	**۱**	۰	۱	۰	۰	۱	۰	۱	۱	۱	۰	۰
چگونگی تحول دانش	۰	۰	۰	۱	۰	۱	۰	۰	۰	۱	۰	**۱**	۱	۰	۰	۰	۰	۱	۰	۱	۱	۰
هدایت دانش ضمنی و صریح	۰	۰	۰	۰	۰	۰	۱	۰	۰	۱	۰	۰	**۱**	۰	۱	۰	۱	۰	۱	۱	۱	۰
همسویی دانش با اهداف سازمانی	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	**۱**	۰	۱	۰	۰	۰	۰	۰	۰
سازمانی چابک	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	**۱**	۰	۰	۰	۰	۱	۱	۰
افزایش خلاقیت و نوآوری	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	**۱**	۰	۱	۰	۰	۰	۰
فناوری	۱	۱	۱	۰	۰	۱	۱	۱	۱	۰	۰	۱	۰	۱	۰	۰	**۱**	۰	۰	۰	۰	۱
پیاده‌سازی موفقیت‌آمیز دانش	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۱	۰	۰	۰	۰	**۱**	۱	۱	۰	۰
شناخت بهتر بازار و مشتری	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	**۱**	۰	۰	۰
سازمانی یادگیرنده	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	**۱**	۱	۰
مزیت رقابتی پایدار	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	**۱**	۰
چگونگی کسب دانش	۰	۱	۰	۱	۰	۱	۰	۱	۰	۱	۰	۱	۰	۱	۱	۰	۱	۰	۰	۱	۱	**۱**

۲-۴-۳- ماتریس دریافتی نهایی

پس‌ازآنکه ماتریس دستیابی اولیه به دست آمد؛ باید سازگاری درونی آن نیز برقرار شود. یعنی اگر عامل A منجر به عامل B شود و عامل B هم منجر به عامل C شود، درنتیجه عامل A باید منجر به عامل C شود. و اگر در ماتریس دستیابی این حالت برقرار نباشد باید ماتریس اصلاح‌شده و روابطی که ازقلم‌افتاده جایگزین شوند. ازاین‌رو، به‌منظور سازگار کردن ماتریس می‌توان به جمع‌آوری مجدد نظرات خبرگان و تکرار این فرآیند تا به دست آوردن سازگاری؛ و یا به‌کارگیری قوانین ریاضی؛ به این صورت که ماتریس دستیابی اولیه را به توان (K+1) می‌رسانند و $K\geq 1$ می‌باشد. لازم به ذکر است که عملیات به توان رساندن ماتریس طبق قاعده بولن[18] (1=1*1 و 1=1+1) انجام می‌گیرد. و یا همچنین از نرم‌افزار

18 . Boolean

Matlabدر این خصوص استفاده نمود. در پژوهش حاضر جهت سازگار کردن ماتریس دستیابی اولیه از نرم‌افزار مطلب استفاده که نتایج در جدول شماره ۶ قابل‌مشاهده است. لازم به ذکر است که در این جدول اعدادی که علامت * گرفته‌اند، نشان می‌دهد که در ماتریس دستیابی اولیه صفر بوده و پس از سازگاری عدد یک گرفته‌اند.

جدول ۷- ماتریس دریافتی نهایی

متغیرها	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
توانایی کارکنان	۱	۱	۱*	۱	۰	۱	۱	۱	۱	۱	۱*	۱	۱*	۱*	۱	۱	۱	۱*	۱	۱	۱	۱*
چگونگی کشف دانش	۰	۱	۰	۱	۱*	۱*	۰	۱*	۰	۱*	۰	۱	۱	۱	۱	۱	۰	۱*	۱*	۱	۱	۱
رفتارهای کارکنان	۱	۱	۱	۱	۰	۱	۱	۱	۱*	۱*	۱*	۱	۱*	۱*	۱*	۱*	۱*	۱	۱	۱*	۱*	۱*
چگونگی انتقال دانش	۰	۱	۰	۱	۱*	۱*	۰	۱	۰	۱	۱	۱	۱*	۱	۱	۱*	۰	۱*	۱*	۱	۱	۱*
تسهیل و حفظ ارتباطات	۰	۰	۰	۰	۱	۰	۰	۱*	۰	۰	۰	۰	۱	۱*	۱*	۱*	۰	۰	۰	۱*	۰	۰
ساماندهی دانش	۰	۱	۰	۱	۱*	۰	۱	۱*	۱*	۱	۱	۱	۱*	۱	۱	۱	۰	۱	۱*	۱*	۱	۱
فرایندها (شیوه عمل)	۱	۱*	۱	۱	۱*	۱*	۱	۱*	۱*	۱	۱*	۱	۱*	۱	۱*	۱*	۱	۱	۱*	۱*	۱*	۰
سازمانی پیشرو	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۱*	۰	۰	۰	۰	۱	۱	۰
محتوا	۱*	۱*	۱*	۰	۱	۱	۱	۱	۰	۱	۱*	۱	۱*	۱	۱*	۱*	۱	۱*	۰	۱	۱	۱*
ممیزی دانش	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۱	۱	۰	۱	۱*	۱	۰	۱*	۱*	۱*	۱*	۰
بازیابی و بهبود دانش	۰	۰	۰	۱*	۱	۱*	۰	۱*	۱*	۱	۱*	۱	۱*	۱*	۱	۱*	۰	۱	۱	۱	۱*	۰
چگونگی تحول دانش	۰	۱*	۰	۱	۱*	۱	۰	۱	۰	۱*	۱	۱	۱*	۱	۱	۱*	۰	۱*	۱*	۱	۱	۱*
هدایت دانش ضمنی و صریح	۰	۰	۰	۰	۱*	۰	۰	۱	۱*	۱	۱*	۰	۱	۱	۱*	۱	۰	۱	۱*	۱	۱	۰
همسویی دانش با اهداف سازمانی	۰	۰	۰	۰	۱	۰	۰	۱*	۰	۰	۰	۰	۱	۰	۱*	۱	۰	۰	۰	۱*	۱*	۰
سازمانی چابک	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۱	۱	۰
افزایش خلاقیت و نوآوری	۰	۰	۰	۱	۰	۰	۰	۱*	۰	۱*	۰	۱*	۱	۱*	۱	۰	۱*	۱*	۱	۰	۱*	۰
فناوری	۱	۱	۱	۰	۱*	۱	۱	۱	۱	۰	۱	۰	۰	۱	۰	۱	۱	۰	۰	۰	۰	۱
پیاده‌سازی موفقیت‌آمیز دانش	۰	۰	۰	۰	۱*	۱*	۱*	۱*	۱*	۱*	۰	۱	۱	۱*	۰	۰	۱	۰	۱	۱	۱	۰
شناخت بهتر بازار و مشتری	۰	۰	۰	۰	۱*	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۱	۰	۰	۰	۰
سازمانی یادگیرنده	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۱	۱	۱	۰
مزیت رقابتی پایدار	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۱*	۰	۰	۰	۰	۱*	۱	۰
چگونگی کسب دانش	۰	۱	۰	۱	۱*	۰	۱	۰	۱	۰	۰	۱*	۱	۱	۱	۰	۰	۰	۱*	۱	۱	۱

۴-۴-۲- تعیین روابط و سطح بندی شاخص‌ها

برای تعیین روابط و سطح بندی شاخص‌ها باید مجموعه ورودی و مجموعه خروجی برای هر کدام از شاخص‌ها مطابق ماتریس دستیابی نهایی مشخص شود. مجموعه خروجی شامل خود شاخص (بعد) و شاخص‌هایی که از آن تأثیر می پذیرد. یعنی برای تعیین مجموعه خروجی هر بعد، باید سطر مربوط به آن را بررسی کرد و تعداد یک-های این سطر را مورد توجه قرار داد. همچنین مجموعه ورودی شامل خود شاخص (بعد) و شاخص‌هایی که بر آن تأثیر می‌گذارند. یعنی برای تعیین مجموعه ورودی هر بعد، باید ستون مربوط به آن را بررسی کرد و تعداد یک های این ستون را مورد توجه قرار داد. پس از مشخص نمودن مجموعه ورودی و مجموعه خروجی برای هر شاخص، مجموعه مشترک (خانه های مشترک بین مجموعه ورودی و خروجی) تعیین می‌گردد و سپس نوبت به سطح‌بندی متغیرها می‌رسد. بدین شکل که متغیری که مجموعه خروجی و مجموعه مشترک آن کاملاً یکسان باشد در بالاترین سطح مدل سلسه مراتب ساختاری تفسیری قرار می‌گیرد و بعد از تعیین این متغیر یا متغیرها، آنها را از جدول حذف می‌کنیم و آنها منشأ هیچ متغیر یا متغیرهایی نخواهند بود. به عبارتی دیگر، هنگامی که سطح بالایی تعیین گردید، از دیگر شاخص‌ها (متغیرها) تفکیک می شود و با شاخص‌های (متغیرها) باقی مانده جدول بعدی را تشکیل و سطح دوم را مشخص می‌نمایم. این عمل برای تمامی شاخص‌ها (متغیرها) تکرار می‌گردد تا همگی آنها سطح‌بندی شوند. بنابراین، جدول شماره هشت سطح‌بندی متغیرها را نشان می‌دهد.

جدول ۸- تعیین سطوح (اولویت‌بندی) متغیرها

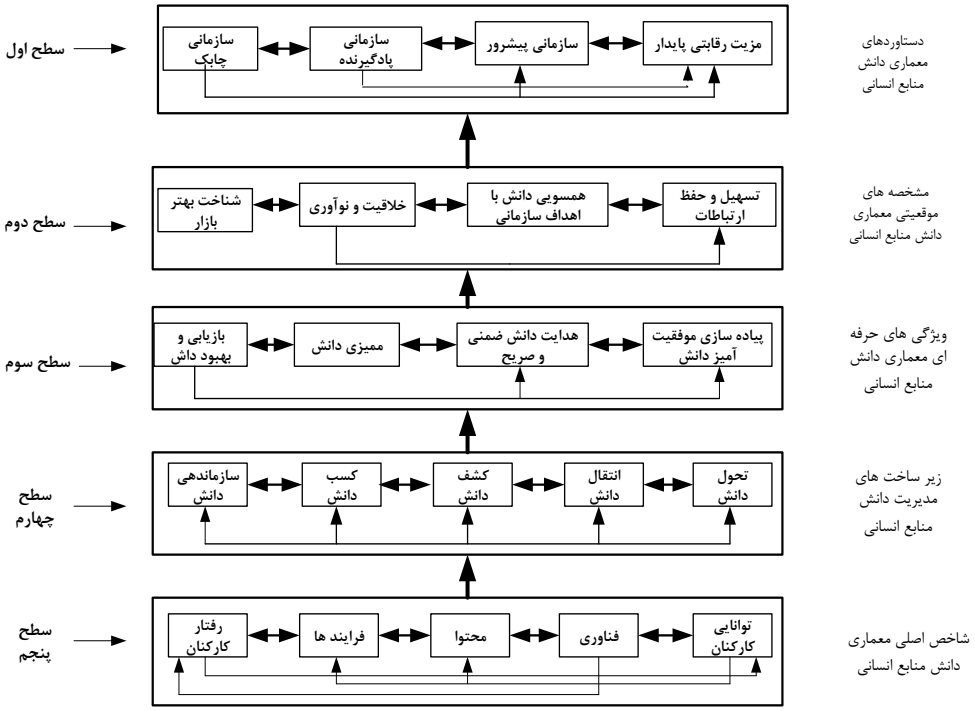
متغیرها	مجموعه ورودی	مجموعه خروجی	مجموعه مشترک	سطح
توانایی کارکنان	۱،۳،۷،۹،۱۷	۱،۲،۳،۴،۶،۷،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۷،۱۸،۱۹،۲۰،۲۱،۲۲،	۱،۳،۷،۹،۱۷	پنجم
چگونگی کشف دانش	۱،۲،۳،۴،۶،۷،۹،۱۲،۱۷،۲۲	۲،۴،۵،۶،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۸،۱۹،۲۰،۲۱،۲۲،	۲،۴،۶،۹،۱۲،۲۲	چهارم
رفتارهای کارکنان	۱،۳،۷،۱۷	۲،۳،۴،۵،۶،۷،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۷،۱۸،۱۹،۲۰،۲۱،۲۲،	۳،۷،۱۷	پنجم
چگونگی انتقال دانش	۱،۲،۳،۴،۶،۷،۹،۱۱،۱۲،۱۷،۲۲	،۲،۴،۵،۶،۸،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۸،۱۹،۲۰،۲۱،۲۲	۲،۴،۶،۱۱،۱۲،۲۲	چهارم
تسهیل و حفظ ارتباطات	۲،۳،۴،۵،۶،۷،۱۱،۱ ۲،۱۳،۱۴،۱۶،۱۸،۲۲	۵،۸،۱۴،۱۵،۱۶،۲۰	۵،۱۴،۱۶	دوم
ساماندهی دانش	،۱،۲،۳،۴،۶،۷،۹،۱۱ ۱۲،۱۶،۱۷،۱۸،۲۲	۲،۴،۵،۶،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۸،۱۹،۲۰،۲۱،۲۲،	۲،۴،۶،۹،۱۱،۱۲،۱۶،۱۸،۲۲	چهارم
فرایندها (شیوه عمل)	۱،۳،۷،۹،۱۷،۲۲	،۱،۲،۳،۴،۵،۶،۷،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۷،۱۸،۱۹،۲۰،۲۱	۱،۳،۷،۹،۱۷	پنجم
سازمانی پیشرو	۱،۲،۳،۴،۵۶،۷،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۷،۱۸،۱۹،۲۰،۲۱،۲۲،	۸،۱۵،۲۰،۲۱	۸،۱۵،۲۰،۲۱	اول
محتوا	۱،۳،۶،۷،۹،۱۱،۱۳،۱۶،۱۷،۱۸	،۱،۲،۳،۴،۶،۷،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۷،۱۸،۱۹،۲۰،۲۱،۲۲	۱،۳،۶،۷،۹،۱۱،۱۳،۱۶،۱۷،۱۸	پنجم
ممیزی دانش	،۱،،۲،۳،۴،۶،۷،۹،۱۰،۱۱ ۱۲،۱۳،۱۶،۱۷،۱۸،۲۲	۸،۱۰،۱۱،۱۳،۱۴،۱۵،۱۶،۱۸،۱۹،۲۰،۲۱	۱۰،۱۱،۱۳،۱۶،۱۸	سوم
بازیابی و بهبود دانش	،۱،۲،۴،۶،۷،۹،۱۰ ۱۱،۱۲،۱۳،۱۷،۱۸،۲۲	،۴،۵،۶،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۸،۱۹،۲۰،۲۱	۴،۶،۹،۱۰،۱۱،۱۲،۱۳	سوم
چگونگی تحول دانش	۱،۲،۳،۴،۶،۷،۹،۱۰ ۱۱،۱۲،۱۶،۱۷،۱۸،۲۲	،۲،۴،۵،۶،۸،۱۰،۱۱،۱۲،۱۳،۱۴،۱۵ ۱۶،۱۸،۱۹،۲۰،۲۱،۲۲	۲،۴،۶،۱۰،۱۱،۱۲،۱۶،۱۸،۲۲	چهارم
هدایت دانش ضمنی و صریح	۱،۲،۳،۴،۶،۷،۹،۱۰ ۱۱،۱۲،۱۳،۱۶،۱۷،۱۸،۲۲	۵،۸،۹،۱۰،۱۱،۱۳،۱۴،۱۵،۱۶ ۱۸،۱۹،۲۰،۲۱	۹،۱۰،۱۱،۱۳،۱۸	سوم
همسویی دانش با اهداف سازمانی	۱،۲،۳،۴،۵،۶،۷،۹،۱۰،۱۱،۱۲ ،۱۳،۱۴،۱۵،۱۶،۱۷،۱۸،۱۹،۲۲	۵،۸،۱۴،۱۵،۱۶،۱۹،۲۰،۲۱	۵،۱۴،،۱۶،۱۹	دوم
سازمانی چابک	،۱،۲،۳،۴،۵،۶،۷،۸،۹،۱۰،۱۱،۱۲ ۱۳،۱۴،۱۵،۱۶،۱۷،۱۸،۲۰،۲۱،۲۲	۸،۱۵،۲۰،۲۱	۸،۱۵،۲۰،۲۱	اول
افزایش خلاقیت و نوآوری	،۱،۲،۳،۴،۵،۶،۷،۸،۹،۱۰،۱۱،۱۲ ۱۳،۱۴،۱۶،۱۷،۱۸،۱۹،۲۰،۲۲	،۶،۷،۹،۱۰،۱۲،۱۳،۱۴،۱۵ ۱۶،۱۷،۱۹،۲۰،۲۱	،۶،۷،۹،۱۰،۱۲،۱۳،۱۴ ۱۶،۱۷،۱۹،۲۰،۲۱	دوم
فناوری	۱،۳،۷،۹،۱۶،۱۷،۱۸	۱،۲،۳،۴،۶،۷،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۷،۱۸،۱۹،۲۰،۲۱،۲۲،	۱،۳،۷،۹،۱۶،۱۷،۱۸	پنجم
پیاده‌سازی موفقیت‌آمیز دانش	۱،۲،۳،۴،۶،۷،۹،۱۰،۱۱،۱۲،۱۳،۱۷،۱۸	۵،۶،۷،۸،۹،۱۰،۱۲،۱۳،۱۴،۱۵ ۱۶،۱۷،۱۸،۱۹،۲۰،۲۱	۶،۷،۹،۱۰،۱۲،۱۳،۱۷،۱۸	سوم
شناخت بهتر بازار و مشتری	۱،۲،۳،۴،۵،۶،۷،۸،۱۰،۱۱،۱۲ ۱۳،۱۴،۱۶،۱۷،۱۸،۱۹،۲۲	۵،۸،۱۴،۱۵،۱۶،۱۹	۵،۸،۱۴،۱۶،۱۹	دوم
سازمانی یادگیرنده	،۱،۲،۳،۴،۵،۶،۷،۸،۹،۱۰،۱۱،۱۲ ۱۳،۱۴،۱۵،۱۶،۱۷،۱۸،۲۰،۲۱،۲۲	۸،۱۵،۱۶،۲۰،۲۱	۸،۱۵،۱۶،۲۰،۲۱	اول
مزیت رقابتی پایدار	،۱،۲،۳،۴،۶،۷،۸،۹،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۷،۱۸،۲۰،۲۱،۲۲	۸،۱۵،۲۰،۲۱	۸،۱۵،۲۰،۲۱	اول
چگونگی کسب دانش	۱،۲،۳،۴،۶،۹،۱۲،۱۷،۲۲	،۲،۴،۵،۶،۷،۸،۱۰،۱۱،۱۲،۱۳ ۱۴،۱۵،۱۶،۱۹،۲۰،۲۱،۲۲	۲،۴،۶،۱۲،۲۲	چهارم

۵-۴-۲- ترسیم مدل

پس از تعیین روابط و مشخص شدن سطح متغیرها می‌توان مدل نهایی آن را با توجه به ماتریس دستیابی سازگار شده ترسیم کرد. به همین منظور متغیرهای پژوهش برحسب سطح (با توجه به جدول شماره هفت) از بالا به پایین مرتب می‌شوند. در این پژوهش شاخص‌ها و

مؤلفه‌های معماری دانش منابع انسانی در پنج سطح دسته‌بندی شدند. شکل شماره دو مدل نهایی پژوهش را نشان می‌دهد.

.



شکل ۲- مدل مفهومی پژوهش

۶-۴-۲- تجزیه‌وتحلیل MICMAC

در این مرحله هدف آن است که قدرت نفوذ و وابستگی متغیرها مشخص گردد. بدین صورت که با جمع کردن ورودی‌های ۱ در هر سطر، میزان قدرت نفوذ، و همچنین با جمع کردن ورودی‌های ۱ در هر ستون میزان قدرت وابستگی شاخص‌ها تعیین می‌شود. لازم به ذکر است که در تجزیه و تحلیل متغیرها براساس قدرت نفوذ و قدرت وابستگی، شاخص‌ها در چهار گروه قرار می‌گیرند. طبقه اول را متغیرهای مستقل تشکیل می‌دهند، متغیرهایی که در این طبقه قرار می‌گیرند دارای قدرت نفوذ بالا و وابستگی اندکی هستند که در اصطلاح به آنها متغیرهای کلیدی گفته می‌شود. متغیرهای خود مختار (خودگردان) طبقه دوم را شامل می شوند که قدرت نفوذ و وابستگی پایینی دارند و بر روابط و پویایی‌های سیستم تأثیر اندکی دارند. در طبقه سوم نیز متغیرهای وابسته قرار می گیرند که دارای قدرت وابستگی زیاد و قدرت نفوذ کمی هستند. و در نهایت در طبقه چهارم متغیرهای پیوندی (متصل) قرار می گیرند که از قدرت نفوذ و وابستگی بالایی برخوردار هستند و هرگونه عملی بر روی آنها منجر به تغییر سایرین خواهد. جدول شماره نه میزان قدرت نفوذ و وابستگی متغیرها را نشان می‌دهد.

جدول ۹- میزان قدرت نفوذ و وابستگی

قدرت نفوذ	مستقل / متصل																							
۲۲																								
۲۱													۹			۷ ۱۷	۱	۳						
۲۰																								
۱۹													۲											
۱۸										۶		۴												
۱۷									۱۲					۲۲										
۱۶									۱۸ ۱۱															
۱۵									۱۳ ۱۴															
۱۴																								
۱۳			۱۶																					
۱۲																								
۱۱									۱۰															
۱۰																								
۹																								
۸				۱۵																				
۷																								
۶						۱۹		۵																
۵		۲۰																						
۴		۸	۲۱																					
۳																								
۲																								
۱	وابسته / خودمختار																							
	۲۳	۲۲	۲۱	۲۰	۱۹	۱۸	۱۷	۱۶	۱۵	۱۴	۱۳	۱۲	۱۱	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱	
میزان وابستگی																								

۵- بحث و نتیجه‌گیری

وجود عدم اطمینان‌ها در محیط و تغییرات شدید در بازده زمانی کوتاه، چالش‌های زیادی را برای سازمان‌ها به همراه آورده است. در چنین شرایطی سازمان‌ها برای مقابله با این تهدیدات و همچنین حفظ جایگاه خود؛ باید رویکرد مناسبی در چگونگی کسب دانش، انتقال دانش، ذخیره‌سازی آن، ... داشته باشند. معماری دانش منابع انسانی ابزار مهمی است که از طریق آن سازمان‌ها بهتر می‌توانند دانش خود را در راستای ارائه خدمات بهتر، بهبود کیفیت کالا و خدمات، کاهش دوباره‌کاری‌ها، قابلیت ارائه خدمات بهنگام و ... به‌کارگیرند. پژوهش حاضر باهدف ارائه الگوی مدل معماری دانش منابع انسانی در سازمان-های دانش‌بنیان با استفاده از رویکرد آمیخته انجام پذیرفت. با توجه به یافته‌های پژوهش مؤلفه‌ها و شاخص-های معماری دانش منابع انسانی در قالب پنج سطح قرار می‌گیرند. به‌این‌ترتیب، یافته‌ها حکایت از آن دارد که رفتار کارکنان، فناوری، توانایی کارکنان، محتوا و فرآیندها به‌عنوان ابعاد اصلی و پنجگانه معماری دانش منابع انسانی پایه و مبنایی مهم برای اجرای موفقیت‌آمیز آن در سازمان است. همچنین یافته‌ها مبین آن است که ساماندهی دانش، کشف دانش، چگونگی کسب، چگونگی انتقال و چگونگی تحول دانش زیرساخت‌های مدیریت دانش منابع انسانی هستند که در صورت توجه به آن‌ها معماری دانش منابع انسانی به‌خوبی به ثمر می‌نشیند. به‌عبارت‌دیگر، یکپارچه‌سازی دانش، شناسایی و ردیابی دانش، گردآوری دانش، اشتراک و تسهیم دانش و ... می‌تواند سازمان را در

به‌کارگیری ویژگی‌های حرفه‌ای معماری دانش منابع انسانی یاری رساند. علاوه براین، نتایج مبین آن است که بازیابی و بهبود دانش، ممیزی دانش، هدایت دانش ضمنی و صریح و پیاده‌سازی موفقیت‌آمیز دانش ویژگی‌های حرفه‌ای معماری دانش منابع انسانی را تشکیل می‌دهند. بر این اساس، ویژگی‌های حرفه‌ای در معماری دانش منابع انسانی مبتنی بر توسعه، بررسی و سنجش دانش و ارزش-گذاری آن است. همچنین در ارتباط با ویژگی‌های حرفه‌ای بیرونی سازی دانش، توان به‌کارگیری دانش تجربی کارکنان و رهبری دانش کارکنان در راستای اهداف سازمان به‌وضوح دیده می‌شود. از طرف دیگر یافته‌های پژوهش چهار ویژگی موقعیتی را برای معماری دانش منابع انسانی شناسایی کرده است. ازاین‌رو مهم‌ترین ویژگی‌های موقعیتی معماری دانش منابع انسانی شناخت بهتر بازار و مشتری، افزایش خلاقیت و نوآوری، تسهیل و حفظ ارتباطات و همسویی دانش با اهداف سازمانی هستند که در سایه‌سار معماری دانش منابع انسانی با افزایش توان خلاقیت و نوآوری، تشخیص فرصت‌های پیش رو و با شناختی مناسب و آگاهانه از بازار و نیازهای مصرف‌کنندگان؛ خود را به‌عنوان سردمداران موفقیت در حوزه فعالیت خود نشان دهند. درنهایت یافته‌های پژوهش حاکی از آن است که رقابت مزیتی پایدار، سازمانی چابک، سازمانی یادگیرنده و سازمانی پیشرو چهار پیامد اصلی معماری دانش منابع انسانی است. بدین منظور با توجه به یافته‌ها می‌توان اذعان داشت که کیفیت و سرعت پاسخگویی، انطباق‌پذیری، افزایش مستمر توانایی افراد، ترویج تفکر سیستمی، توانایی انتقال دانش در سازمان، منحصربه‌فرد بودن، ایجاد ارزش برای مشتری، اشتراک‌گذاری اهداف یا کارکنان و روحیه بالای کارکنان از مهم‌ترین پیامدهای معماری دانش منابع انسانی است که سازمان‌های دانش‌بنیان می‌توانند با اجرای معماری دانش منابع انسانی آن‌ها را دارا باشند. در خصوص وجوه اشتراک و افتراق پژوهش حاضر با دیگر پژوهش‌های صورت گرفته باید عنوان نمود که پژوهش حاضر با روزیکیک و میکیک (۲۰۱۷) و سندکوه (۲۰۱۵) همخوانی دارد. بدین شکل که در پژوهش آن‌ها معماری دانش منابع انسانی چارچوبی منسجم و یکپارچه برای استخراج و ذخیره‌سازی دانش کارکنان، کیفیت و سرعت پاسخگویی، افزایش بهره‌وری و انطباق‌پذیری سازمان قلمداد می‌شود؛ کما اینکه در پژوهش حاضر انطباق‌پذیری و چابکی سازمان ازجمله پیامدهای معماری دانش منابع انسانی شناسایی‌شده است. همچنین نتایج پژوهش حاضر با پژوهش ال‌کادی (۲۰۰۳) دارای اشتراک است. در پژوهش ال‌کادی (۲۰۰۳) معماری دانش کارکنان ابزاری برای توسعه و بازیابی دانش شناخته‌شده است. با توجه به اینکه در پژوهش حاضر نیز بازیابی و بهبود دانش به‌عنوان یکی از ویژگی‌های حرفه‌ای معماری دانش منابع انسانی شناسایی‌شده؛ می‌توان عنوان نمود که این دو پژوهش دارای همخوانی هستند. در ارتباط با وجه نوآوری پژوهش حاضر لازم به ذکر است که تا زمان انجام پژوهش، هیچ پژوهشی که مبادرت به طراحی مدل معماری دانش منابع انسانی در سازمان‌های دانش‌بنیان نموده باشد، یافت نشد. درنهایت در خصوص محدودیت‌های پژوهش حاضر باید گفت که کم بودن ادبیات و مبانی نظری یا به عبارتی کم بودن پژوهش‌های مشابه مرتبط با معماری دانش منابع انسانی به‌عنوان مهم‌ترین محدودیت پژوهش حاضر است.

۵-۱- پیشنهادات کاربردی

- پژوهش حاضر به مدیران سازمان‌های دانش‌بنیان در جهت نیل به موفقیت و برای ورود به عرصه رقابت جهانی و توفیق در بازارهای داخلی پیشنهاد می‌کند که به معماری دانش منابع انسانی توجه و عنایت ویژه‌ای داشته باشند و اجرای معماری دانش مبادرت ورزند.
- بدین منظور مدیران سازمان‌های دانش‌بنیان باید در راستای اهداف و نیازهای اطلاعاتی حال و آینده خود توانایی‌ها، رفتارها و فرایندهای کارکنان دانشی را مدنظر قرار دهند. چراکه قدرت تجزیه‌وتحلیل بالای کارکنان، اقدامات و فعالیت‌های آن‌ها در کسب و شناسایی و یا تسهیم دانش و همچنین چرایی و چگونگی رفتارهای کارکنان در فرایندهای مدیریت ضمن قرار دادن سازمان در مسیر تعالی و رشد؛ توانایی و قابلیت آن را در تأثیرگذاری بر محیط پیرامون خود افزایش دهند.
- بر مبنای یافته‌ها پژوهش حاضر به سازمان‌های دانش بنیان پیشنهاد می‌کند که این رویکرد می‌تواند دانش کارکنان را که به هر علتی منسوخ و کهنه شده؛ و به

احتمال زیاد دارای اثرات منفی است و می‌تواند وضعیت سازمان را به مخاطره اندازد، شناسایی نموده و با انجام اقدامات و فعالیت‌هایی همچون بازیابی، تجزیه و تحلیل و یا آمایش دانش بتواند خطرات احتمالی و یا میزان وابستگی به آن را کاهش دهد.

- همچنین پژوهش حاضر به سازمان‌های دانش‌بنیان پیشنهاد می‌کند تمامی تلاش و توان خود را در کسب و تسهیم دانش ضمنی (دانشی نهفته در فرد) به کار گیرند. چرا که حاصل تزریق تخصص و تجربه فرد به دیگر کارکنان تحول و تعالی در سازمان خواهد بود و دیری نخواهد پایید که چنین سازمانی گوی سبقت را از دیگر سازمان‌ها خواهد ربود.
- پژوهش حاضر با توجه به قابلیت‌های معماری دانش و براساس یافته‌ها به سازمان‌های دانش بنیان توصیه می کند که طراحی و نحوه چگونگی سازماندهی دانش می-توانند ضمن بهبود ارتباط‌ها؛ یادگیری و جذب اطلاعات جدید را پشتیبانی نمایند.
- بدیهی است دانش ابزاری استراتژیکی برای رویارویی با عدم اطمینان‌های محیطی است. این امر مستلزم آن است که سازمان‌های دانش‌بنیان با ارج نهادن به معماری دانش منابع انسانی، ضمن بازیابی و بهبود دانش بتوانند دانش‌های منسوخ و کهنه خود را کنار گذاشته و متناسب با نیازهای خود دانش جدید را کسب نمایند تا با افزایش قدرت خلاقیت و نوآوری جایگاه خود را در عرصه رقابتی بهبود بخشند.

۵-۲- پیشنهاد برای تحقیقات آتی

- درنهایت به پژوهشگران آتی توصیه می‌شوند که عوامل اثرگذار بر شکل‌گیری معماری دانش منابع انسانی را در سازمان‌های دانش‌بنیان بسنجند و یا اثرگذاری معماری دانش منابع انسانی را بر مفاهیم با نزدیک بودن معانی مناسب مورد تحلیل و ارزیابی قرار دهند. پژوهش حاضر به پژوهشگران آتی پیشنهاد می‌کند؛ مدل معماری دانش منابع انسانی را در قلمرو مکانی دیگر تدوین و ارائه نمایند تا امکان مقایسات تطبیقی را فراهم آورده و به این ترتیب، فهم

گسترده‌تری از مفهوم معماری دانش منابع انسانی را کسب کرد. همچنین با استفاده از فرایند رتبه‌بندی تفسیر (IRP) به تحلیل روابط متغیر مذکور با مفاهیم مرتبط بپردازند.

منابع

1. Lyu, Ch., Yang, J., Zhang, F., Thompson S.H & Tian Mu, T (2020). How do knowledge characteristics affect firm’s knowledge sharing intention in interfirm cooperation? An empirical study, Journal of Business Research, 115, 48-66.

2. Attar, M., Kang, K & Sohaib, O (2019). Knowledge Sharing Practices, Intellectual Capital and Organizational Performance, International Conference on System Sciences, 5578-5587.

3. Zaim, H., Keceli, Y., Jaradat, A., Kastrati, S (2018). The effects of knowledge management processes on human resource management: Mediating role of knowledge utilization, Journal of Science and Technology Policy Management, 9(3), 310-328.

4.Wickramasinghe, N. (2003). Do We Practise What We preach? Are Knowledge Management Systems in Practice Truly Reflective of Knowledge Management Systems in theory?. Business Process Management Journal, 9(3), 295-316.

5. Govender, L., Peruma, R & Perumal, S (2018). Knowledge management as a strategic tool for human resource management at higher education institutions, South African Journal of Information Management, 20(1), 1-10.

۶. عبدالهی، مریم؛ حسین زاده، علی (۱۳۹۷). بررسی تأثیر معماری دانش بر هویت‌سازمانی با تأکید بر نقش میانجی توانمندسازی روان‌شناختی و سرمایه اجتماعی (موردمطالعه: کارکنان شهرداری مشهد)، مجلهٔ علوم اجتماعی دانشکدهٔ ادبیات و علوم انسانی، سال پانزدهم، ۱۴۲-۱۱۳.

7. Heizmann, H & Olsson, M.R (2015). Power matters: the importance of Foucault’s power/knowledge as a conceptual lens in

KM research and practice, Journal of Knowledge
Management, 19(4), 756-769.
8. Wipawayangkool, K &Teng, J (2018). Profiling knowledge workers' knowledge sharing behavior via knowledge internalization, Knowledge Management Research & Practice,9 (2), 1-14.
9. Nick Bontis, A (2016). Negotiate, reciprocate, or cooperate? The impact ofexchange modes on inter-employee knowledge sharing, Journal of Knowledge Management, 20(4), 1-60.
10. Gaizauskiene, L & Tuncikiene, Z (2016**).** The concept and role of knowledge worker and workplace fit in learning organisation, International Journal of Learning and Change, 5(3), 1-10.
10. Rozewski, P., Jankowski, J., Brodka, P & Michalski, R. (2015). Knowledge worker‘s collaborative learning behavior modeling in an organozational social network, Computers in Human Behavior, 51, 1248-1260.
11. Tsai, S (2018). Innovative behaviour of knowledgeworkers and social exchange attributesoffinancial incentive: implicationsfor knowledge management, Journal of Knowledge Management, 22(8), 1757-1781.
12. Kianto, A., Vanhala, M., & Heilmann, P. (2016). The impact of knowledge management on job satisfaction. Journal of Knowledge Management, 20(4), 621–6 Garcia, M. (2015). How to measure knowledge management: dimensions and model. VINE, 45(1), 107–125. 36.
13. Calabrese, (2006). Knowledge-based organizations in context", VINE, 36(1), 12 -
16.
14. Pech, R. (2003). Developing a leadership knowledge architecture: a cognitive approach" Leadership and Organization Development Journal, 24(1), 32-42.
15. Sowa, J.F. (2000). Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co.: Pacific Grove, CA.
16. Sandkuhl, K (2015). Patten-Based Knowledge Architecture For Information Logistics, Revista Investigacion Operacional, 36(1), 36-44.
17. Ruzicic, V & Micic, Z (2017). Creating a strategic national knowledge architecture: A Comparative analysis of knowledge source innovation in the ICS subfields of multimedia and IT security, journal computers & security, 18(5), 455-466.
18. Wickramasinghe, N. (2003). Do we practise what we preach? Are knowledge management systems in practice truly reflective of knowledge management systems in theory?. *Business Process Management Journal*, *9*(3), 295-316.
19. Lee VH, Leong LY, Hew TS, Ooi KB (2013) Knowledge Management: a Key Determinant in Advancing Technological innovation? J Knowl Manag 17(6):848–872.

20. Kamhawi, E (2010). The three tiers architecture of knowledgeflow andmanagement Activities, Information and Organization, 7(1), 169- 186.
21. Morganwalp, J & Sage, A. P. (2003). A system of systems focused enterprise architecture framework and an associated architecture development process. Information, Knowledge, Systems Management 3: 87-105.

22. Holm, J,. Olla, P,. Moura, M & Warhaut, M (2006). Creating architectural approaches Toknowledge management: an example fromthe space industry, Journal Of Knowledge Management, 10(2), 36-51.
23. Evers, H. D. (2008). Knowledge hubs and knowledge clusters: Designing a knowledge Architecture for development. Paper presented at the conference of Knowledge Architecture for Development: Challenges ahead for Asian Business and Governance, Singapore, SMU.
24. Butt, A., Nawaz, F., Hussain, S & Sousa, M (2018). Individual knowledge management engagement, knowledge-worker productivity, and

innovation performance in knowledge-based organizations: the implications for knowledge processes and knowledge-based systems, Comput Math Organ Theory, 1-21.
25. Fernandez R (2013). The factors determining knowledge worker productivity within the Irish IT Industry (Doctoral dissertation, Dublin Business School).
http://esource.dbs.ie/handle/10788/1753

26. Tyugu, E (2005). Understanding knowledge architectures, Knowledge-Based Systems, 19(3), 50- 56.

22. Bardolet, C., Sellens, J & Royo, M (2018). Knowledge Workers and Job Satisfaction: Evidence from Europe, Springer Science +Business Media, LLC, J Knowl Econ, 1-25.
23. Atapattu, M (2018). High performance work practices and knowledge workers' propensity for knowledge management processes, Knowledge Management Research & Practice,1-11.
24. Snyman, R., and C. J. Kruger. 2004. The interdependency between strategic management and strategic knowledge management. Journal of knowledge management 8: 5-19.
25. Kianto,A., Shujahat, M,. Hussain, S., Nawaz, F & Ali, M (2018). The impact of knowledge management on knowledgeworker productivity, Baltic Journal of Management, 1-21.
26. Chevron, J. S. (2001). Developing an integrated enterprise-wide knowledge architecture. Paper presented at the APQC conference of Next Generation KM, APQC, Houston, TX: pp. 1-20.
۲۷. علامه، سید محسن؛ عسگری، نوربخش؛ خزایی پول، جواد (۱۳۹۵). بررسی فرهنگ‌سازمانی بر عملکرد سازمانی: تأکید بر نقش تسهیم دانش و چابکی سازمانی با رویکرد کارت امتیازی متوازن، مدیریت فرهنگ‌سازمانی، ۱۴(۲)، ۴۵۳-۴۷۴.
28. Mishra, P., Kishore, S & Shivani,Sh (2018). The Role of Information Technology for Knowledge Management: An Empirical Study of the IndianCoal Mining Industry, Journal of Global Information Technology Management,21(3), 208-225.
29. Choi, B. & Lee, B (2003). Knowledge Management as a catalyst for innovation within organiza-tions, Organisation Study,18(7), 403- 417
30. Darroch, J (2005).Knowledge management, innovation, and firm performance, Journal of Knowledge Management,. 3(9), 101–115.
31. Faisal, M., Banwet, D.K. and Shankar, R. (2006). Supply chain risk mitigation: modelling the enablers, Business Process Management, 12(4), 535-552
32. Ravi V. & Shanker R. (2005). Analysis of interactions among the barriers of reverse logistics; Technological Forecasting and Social Changes,72.
33. Lasnik, V. E. (2000). Architects of knowledge: an emerging hybrid profession for educational communications, in: STC's 50th Annual Conference, Dallas TX. pp.: 132-136.

سال یازدهم، شماره‌های ۳۹و۴۰، بهار و تابستان ۱۳۹۸ فصلنامهٔ علمی- پژوهشی

صص:۹۱ -۱۰۸ فناوری اطلاعات و ارتباطات ایران

بررسی تاثیر استفاده از انواع استراتژی‌های بازاریابی در شبکه‌های اجتماعی بر جلب اعتماد مشتریان

* فرزانه میلانی ** سیدجعفر زنوزی

* کارشناس ارشد مدیریت بازرگانی- بازاریابی دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ایران

** استادیار گروه مدیریت بازرگانی دانشکده اقتصاد و مدیریت، دانشگاه ارومیه، ایران

تاریخ دریافت: ۱۳۹۸ /۰۶/۱۹ تاریخ پذیرش: ۱۳۹۹/۰۴/۱۸

چکیده

هدف از پژوهش حاضر، بررسی تاثیر استفاده از انواع استراتژی‌های بازاریابی در شبکه‌های اجتماعی بر جلب اعتماد مشتریان می‌باشد. بازاریابی از طریق شبکه‌های اجتماعی فرصت‌های مناسبی را برای شرکت‌ها بوجود آورده تا مشتریان زیادی را به طرف خود جذب کنند. از چالش‌های پیش روی بازاریابی در این شبکه‌ها می‌توان به جلب اعتماد و جذب مشتریان اشاره کرد. لذا برای ایجاد مزیت رقابتی لازم است شرکت‌ها استراتژی‌های مناسبی را برای جلب اعتماد مشتریان در این شبکه‌ها بکار گیرند. جامعه آماری این پژوهش شامل تمامی کاربران ایرانی عضو سایت‌های شبکه‌های اجتماعی است که تحت تاثیر تبلیغات شرکت‌ها قرار می‌گیرند. حجم نمونه آماری نیز با روش نمونه‌گیری گلوله برفی ۴۴۶ نفر است. روش پژوهش، توصیفی پیمایشی بوده و ابزار گردآوری داده‌ها پرسشنامه می‌باشد. برای آزمون فرضیه‌های پژوهش از روش حداقل مربعات جزئی (PLS) و نرم‌افزار SmartPLS 3 استفاده شده است. نتایج بدست آمده نشان می‌دهد که هر چهار متغیر، استراتژی‌های بازاریابی مبادله‌ای، رابطه‌ای، پایگاه داده و مبتنی بر دانش در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارند. البته این تأثیر در استراتژی مبادله‌ای منفی است که به معنی رابطه معکوس بین این نوع استراتژی و اعتماد مشتریان می‌باشد. نتایج پژوهش حاکی از آن است که بازاریابی مبتنی بر دانش دارای بیشترین تأثیر مثبت بر جلب اعتماد مشتریان می‌باشد.

واژه‌های کلیدی: استراتژی‌های بازاریابی، شبکه‌های اجتماعی، اعتماد مشتریان، بازاریابی الکترونیکی.

۱- مقدمه

شبکه‌های اجتماعی یکی از بزرگترین تحولات اینترنت در قرن ۲۱ می‌باشد که به عنوان رسانه جدید از بازاریابی مستقیم هر ساله اهمیت خود را هم برای کاربران و هم برای بازاریابانی که آنها را در طرح رسانه‌ای خود جا می‌دهند بیشتر می‌کنند [۱]. در سال ۲۰۲۰ تعداد کاربران شبکه‌های اجتماعی در سراسر جهان همچنان رو به رشد است که از ۳ میلیارد نفر عبور کرده و پیش‌بینی می‌شود تا سال ۲۰۲۳، تقریباً ۸۰ درصد از جمعیت آنلاین، حداقل ماهانه از شبکه‌های اجتماعی بازدید کنند [۲]. با توجه به مخاطبان گسترده بالقوه موجود که ساعت‌های زیادی را صرف استفاده از این شبکه‌ها می‌کنند، بازاریابان به عنوان کانال بازاریابی از این بستر استقبال می‌کنند [۳]. در واقع، به نظر می‌رسد که شبکه‌های اجتماعی راه ارتباطی جدید و کم هزینه‌تری برای تعامل بیشتر و شخصی‌سازی با مشتریان هدف داشته باشند. این امر به نوبه خود، به

نویسنده عهده‌دار مکاتبات : سیدجعفر زنوزی j.zonoozi@gmail.com

شرکت‌ها کمک می‌کند تا در مقایسه با استفاده از ابزارهای سنتی بازاریابی، فعالیت‌های بازاریابی خود را به صورت کارآمدتر و موفقیت‌آمیز انجام دهند [۴، ۵].

در تعاملات انسانی، اعتماد همیشه یک فاکتور مهم در مبادلات توام با ریسک بوده است. همچنین در ادبیات شبکه‌های اجتماعی آنلاین اشاره شده است که اعتماد یک متغیر مهم در روابط است [۶، ۷] و مهم‌ترین عامل در به دست آوردن وفاداری مشتری در کسب‌وکارهای آنلاین در نظر گرفته می‌شود، از آنجایی که مسائل حریم خصوصی و امنیت موانع عمده کانال‌های اینترنتی هستند. بدون اعتماد، مشتریان اطلاعات شخصی خود را به فروشندگان نخواهند داد. در واقع، اعتماد حتی در خریدهایی با ریسک کمتر، از فروشندگان آنلاینی که شناخته شده هستند نیز بسیار مهم است [۸]. اعتماد می‌تواند تعامل بین فروشنده و خریدار را تسهیل کند، فرصت‌هایی را برای شرکت‌های آنلاین فراهم می‌کند تا به اهداف خود برسند و تاثیر مثبتی نیز بر قصد خرید مشتریان دارد [۹، ۱۰]. مطالعات بسیاری بر اهمیت اعتماد در شبکه‌های اجتماعی، تجارت اجتماعی[1] و تجارت الکترونیکی[2] تاکید کرده‌اند [۵، ۷، ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶]. با توجه به اینکه فرصت بزرگی برای تبلیغات در شبکه‌های اجتماعی به روی بازاریاب‌ها و شرکت‌های تولیدی باز شده است، لذا برای بهره‌گیری بهینه از بستر شبکه‌های اجتماعی، لازم است استراتژی‌های مناسبی را برای جلب اعتماد و جذب مشتریان در این شبکه‌ها بکار گیرند. در این پژوهش به دنبال بررسی این موضوع خواهیم بود که آیا استفاده از انواع استراتژی‌های بازاریابی در شبکه‌های اجتماعی بر جلب اعتماد مشتریان می‌تواند موثر باشد.

۱- مبانی نظری و پیشینه پژوهش

۱-۲- اعتماد

بسیاری از پژوهشگران، بسته به موضوع تخصصی خود، اعتماد را به گونه‌ای خاص تعریف کرده‌اند [۱۷]. تعریفی از اعتماد که مورد توافق اکثریت است توسط مایر و همکارانش ارائه شده است که بیان می‌کند اعتماد "تمایل یک طرف به قرار گرفتن در معرض آسیب پذیری اقدامات طرف دیگر بر اساس این انتظار که طرف مورد اعتماد اقدامی خاص را که برای اعتمادکننده اهمیت دارد برای او انجام خواهد داد صرف نظر از توانایی نظارت و کنترل طرف اعتمادکننده بر اعتمادشونده" است [۱۸]. آسیب‌پذیری در وضعیت آنلاین با توجه به ماهیت اینترنت برجسته می‌شود. از آنجایی که اینترنت اجازه نمی‌دهد تا مصرف‌کننده دامنه مشابهی از نشانه‌های ذاتی که در خرید فیزیکی سنتی انجام می‌گیرد گردآوری کند، ریسک درک شده بیشتر است و در نتیجه ارزیابی اعتماد توسط مصرف‌کننده باید صورت گیرد [۱۹]. اعتماد به معنای آن است که "فرد می‌تواند به قول یا وعده دیگری اتکا کند و طرف مقابل در شرایط غیر قابل پیش‌بینی، با حسن نیت عمل خواهد کرد" [۹].

ویژگی‌های اعتماد آنلاین

ونگ و ایموریان در سال ۲۰۰۵، چهار ویژگی برای اعتماد بیان کرده‌اند. این مفاهیم ویژه عبارتند از:

اعتمادکننده و اعتمادشونده[3]: در هر ارتباط مبتنی بر اعتماد، دو طرف خاص وجود دارند، اعتمادکننده و اعتمادشونده. این دو نقش اعتماد، می‌تواند بر عهدهٔ یک فرد، یک شرکت و یا حتی یک محصول قرار گیرد. در دنیای آنلاین، دو طرف اعتمادکننده و اعتمادشونده برای برقراری روابط مبتنی بر اعتماد ضروری هستند، اما بر موجودیت‌های خاصی اشاره دارند. در فضای مجازی بر خلاف محیط واقعی، اعتمادکننده، مشتری و یا کسی است که از طریق وبسایت به انجام تجارت الکترونیکی می‌پردازد و در مقابل، اعتمادشونده، وبسایت موردنظر یا فروشندگان وبسایت می‌باشد.

آسیب‌پذیری[4]: اعتماد، شامل قبول آسیب‌پذیری توسط اعتمادکننده است. به علت پیچیدگی و ریسک فراوان موجود در تجارت الکترونیک، فروشندگان می‌توانند به شیوه‌ای غیر قابل پیش‌بینی در اینترنت رفتار کنند. خریداران و مصرف‌کنندگان نیز اغلب در مورد ریسک موجود در هنگام استفاده از وبسایت و نیز ریسک موجود در هنگام تراکنش آنلاین نامطمئن می‌باشند. حتی زمانی که مصرف‌کنندگان اینترنتی تنها به بررسی وبسایت پرداخته و از آن خریدی نمی‌کنند باز هم داده و اطلاعات به صورت اتوماتیک از فعالیت‌های آنها جمع‌آوری شده که

[1] social commerce
[2] electronic commerce
[3] Trustor and trustee
[4] Vulnerability

بعدها می‌تواند مورد سوء استفاده قرارگیرد و یا بدون آگاهی یا رضایت آنان توزیع شود.

رفتار ناشی از اعتماد[5]: اعتماد منجر به اقدامات و رفتارهایی می‌شود که اغلب همراه با ریسک‌پذیری است. شکل و نوع اقدام به موقعیت مورد نظر بستگی دارد و این رفتارها می‌توانند محسوس و یا نامحسوس باشند. اعتماد مشتری در تجارت الکترونیکی می‌تواند منجر به دو نوع اقدام از سوی مشتری شود. ۱) خرید مجازی که اغلب همراه با ارائهٔ کارت اعتباری و اطلاعات شخصی است. ۲) تماشای فروشگاه‌های الکترونیکی. این اقدامات می‌تواند منجر به نتایج مثبت برای فروشندگان الکترونیکی شود که می‌توان آن را فروش بالقوه دانست.

موضوعیت ذهنی[6]: اعتماد موضوعی ذهنی و بطور مستقیم مرتبط و تأثیرپذیر از تفاوت‌های شخصی و عوامل محیطی است. افراد مختلف نقش اعتماد را در سناریوهای مختلف، متفاوت ارزیابی کرده و تلقی‌های گوناگونی از آن دارند. اعتماد در محیط مجازی نیز مانند اعتماد در محیط واقعی، مفهومی ذهنی است. سطح اعتماد مورد نیاز برای ایجاد تراکنش‌های/ معاملات آنلاین با توجه به ویژگی‌های شخصی هر نفر متفاوت است همچنین انسان‌ها طرز تلقی‌های گوناگونی نسبت به فناوری دارند [۱۷].

ابعاد و مولفه‌های اعتماد آنلاین

برخی از محققان اعتماد را به عنوان یک ساختار تک بعدی می‌بینند [۲۰]؛ با این حال، در ادبیات ابعاد متعددی از اعتماد شناسایی شده است. جدول ۱ مطالعات پیشین ابعاد اعتماد را نشان می‌دهد.

جدول ۱- ابعاد اعتماد[7]

ابعاد شناسایی شده اعتماد	نویسندگان
صداقت، خیرخواهی، توانایی	لیانگ و همکاران[8]، ۲۰۲۰ [۱۱]
صداقت، خیرخواهی، توانایی	یان و همکاران[9]، ۲۰۱۹ [۱۲]
توانایی، خیرخواهی، صداقت، قابلیت پیش‌بینی	حاجلی و همکاران[10]، ۲۰۱۷ [۹]
شایستگی، خیرخواهی، صداقت	الیویرا[11]، ۲۰۱۷ [۱۰]
شایستگی، توانایی، خیرخواهی، صداقت و درستکاری	گرابنر-کراتر و بیتر[12]، ۲۰۱۵ [۱۳]
توانایی، خیرخواهی، صداقت، قابلیت پیش‌بینی	پالمر و هیو[13]، ۲۰۱۳ [۱۴]
توانایی، خیرخواهی، صداقت، قابلیت پیش‌بینی	مک نایت و چروانی[14]، ۲۰۰۲ [۲۱]
خیرخواهی	لارزلر و هاستون[15]، ۱۹۸۰ [۲۰]

در شبکه‌های اجتماعی به علت نبود ارتباطات چهره به چهره به دست آوردن درک بیشتری از اعتماد ضروری است [۱۱]. در واقع، اعتماد کاتالیزور تسهیل کننده مهم تعامل اجتماعی و روابط بلندمدت است [۲۲]. در این شبکه‌ها، مشتریان با تبلیغات، تصاویر/ فیلم‌ها/ اخبار، پیشنهادات و لایک‌های مرتبط با فروشندگان الکترونیکی روبرو می‌شوند. با این وجود، اعتماد در شبکه‌های اجتماعی می‌تواند میزان اتکا مشتری را به اعتبار محتویات و فعالیتهای فروشندگان الکترونیکی تعیین کند. در راستای طبقه‌بندی ابعاد اعتماد که توسط مک نایت و چروانی در سال ۲۰۰۲ [۲۱] پیشنهاد شده است ابعاد اعتماد شامل توانایی[16]، خیرخواهی[17]، صداقت[18] و قابلیت پیش‌بینی[19] است.

توانایی- اشاره به سطح شایستگی درک شده افراد و یا شرکت‌ها برای انجام برخی رفتارهای مورد نظر است. در این زمینه، مشتری ارزیابی می‌کند که آیا فروشنده شایستگی‌ها، مهارت‌ها و تخصص لازم برای برآوردن نیازهای آنها و انجام مبادله موفقیت آمیزی را دارد.

خیرخواهی- اشاره به برداشت اعتمادکننده به تلاش‌های اعتمادشونده و همچنین تمایل به دستیابی به برخی ارزش‌هایی که در یک رابطه بدون پاداش مطلوب هستند دارد. به عبارتی، نشان دادن حسن نیت جدا از انگیزه سود محوری است. در یک محیط مجازی، جامعه خیرخواه باید حداقل به طور فعال به سوالات مصرف‌کننده پاسخ دهند و فعالانه به دنبال یافتن نیازها و خواسته‌های اعضا باشند.

صداقت- به رفتار صالح اشاره دارد. در یک محیط مجازی، صداقت حاکی از انطباق ارزش‌ها، اصول و قوانین رایج پذیرفته شده است. صداقت به اعتقاد درک شده مشتریان از این که فروشنده مجازی صادقانه عمل خواهد کرد و

[5] Produced actions
[6] Subjective matter
[7] Trust dimensions
[8] Leong et al.
[9] Yeon et al.
[10] Hajli et al.
[11] Oliveira et al.
[12] Grabner-Kräuter & Bitter
[13] Palmer & Huo
[14] McKnight & Chervany
[15] Larzelere & Huston
[16] Ability
[17] Benevolence
[18] Integrity
[19] Predictability

پایبند به یک مجموعه اصول و استانداردهای پذیرفته شده می‌باشد اشاره دارد.

قابلیت پیش‌بینی- اشاره به باور اعتمادکننده دارد که اعتمادشونده به معامله وعده داده شده و همچنین به سیاست‌ها و خط مشی‌های تعامل پایبند است. قابلیت پیش‌بینی به شهرت درک شده فروشنده برای ارائه خدمات یکنواخت و پایدار اشاره دارد [۲۱، ۱۹، ۹]. مبنای این پژوهش، ۴ بعد شناسایی شده (توانایی، خیرخواهی، صداقت و قابلیت پیش‌بینی) است.

۲-۲- شبکه‌های اجتماعی

سایت‌های شبکه‌های اجتماعی را می‌توان به عنوان "خدمات مبتنی بر وب که امکان ایجاد پروفایل عمومی یا نیمه عمومی را برای افراد در سیستم محدود فراهم می‌کند، مجموعه ارتباطات با سایر کاربران در سیستم را نشان می‌دهد و لیست ارتباطات خود فرد و همچنین ارتباطات ایجاد شده توسط دیگران در سیستم را نشان می‌دهد" تعریف کرد [۲۳]. تعریف دیگر شبکه‌های اجتماعی شامل سرویس الکترونیکی، اپلیکیشن، بستر یا سایت مورد استفاده افرادی که علاقه، اعتقادات، نگرش، فرهنگ، فعالیتها و روابط واقعی زندگی مشترک دارند است و همچنین استفاده از رسانه‌های اجتماعی به عنوان ابزاری برای ارتباط مستقیم، تعامل با سایر افراد و در نتیجه داشتن رابطه واقعی با آنها است [۵]. محبوب‌ترین شبکه‌های اجتماعی شامل فیسبوک، مای اسپیس، لینکدین، توییتر و یوتیوب هستند.

چارچوب شبکه‌های اجتماعی، کاربران را به ارسال و اشتراک‌گذاری اطلاعات شخصی به عنوان بخشی از تعاملات اجتماعی تشویق می‌کنند [۲۴]. شرکت‌ها باید این را در ذهن داشته باشند که سایت‌های شبکه‌های اجتماعی محلی است "... که در آن مشتریان بالقوه و بالفعل در حال تعامل هستند و نحوه تفکر آنها را شکل می‌دهند" [۲۵].

شبکه‌های اجتماعی امکان ایجاد و حفظ ارتباط با خانواده، دوستان و شرکای کسب و کار را در یک محیط مجازی فراهم می‌کنند و همچنین به کاربران برای یافتن افراد دیگر با استفاده از واژه‌ها و تعاریف کلیدی کمک می‌کنند. سایت‌های شبکه‌های اجتماعی از سال ۲۰۰۴ به طور گسترده‌ای به شکل مشابه امروزی شناخته شده‌اند و بطور چشمگیری قابلیت اینترنت را تغییر داده‌اند. این شبکه‌ها پتانسیلی برای دسترسی مستقیم به مصرف‌کنندگان هستند و در یک محیط فردی و اجتماعی بدان معنی است که بازاریابان "مجبور" به تبلیغ در این رسانه جدید هستند [۲۶]. جوامع شبکه‌های اجتماعی آنلاین افراد را دور هم جمع می‌کند. بنابراین، شرکت‌ها می‌توانند ارتباط موثرتری را با گروه‌های هدف مصرف‌کننده در جوامع شبکه‌های اجتماعی نسبت به انواع دیگر رسانه‌ها برقرار کنند [۲۷].

۳-۲- استراتژی‌های بازاریابی الکترونیکی

استراتژی بازاریابی الکترونیکی، استراتژی است که از اینترنت و دیگر فناوری ارتباطی و اطلاعاتی دیجیتالی استفاده می‌کند. استفاده از اینترنت منجر به ظهور سایت‌های شبکه‌های اجتماعی شده است که به عنوان کانالی برای کسب‌وکار الکترونیکی برای تبلیغ محصولات و خدمات و یا عرضه استراتژی بازاریابی بکار می‌رود. بازاریابان همیشه به دنبال درک بهتر رفتار مشتریان خود هستند. شبکه‌های اجتماعی برای به دست آوردن اطلاعات بیشتر در مورد مصرف‌کنندگان بکار می‌رود چرا که این سایت‌ها حاوی ترجیحات کاربران هستند که برای شرکت‌هایی که به دنبال مشاهده و تعقیب روند رفتار مصرف‌کننده هستند مهم است [۲۸].

به طور کلی بر اساس مطالعات و پژوهشات صورت گرفته در زمینه بازاریابی الکترونیکی می‌توان استراتژی‌های بازاریابی الکترونیکی را به چهار دسته اصلی تقسیم‌بندی نمود:

* استراتژی بازاریابی مبادله‌ای[20] [۲۹، ۳۰، ۳۱، ۳۲، ۳۳، ۳۴]
* استراتژی بازاریابی رابطه‌ای[21] [۲۹، ۳۱، ۳۳، ۳۴، ۳۵، ۳۶]
* استراتژی بازاریابی پایگاه داده[22] [۳۰، ۳۲، ۳۳، ۳۴]
* استراتژی بازاریابی مبتنی بر دانش[23] [۳۱، ۳۳، ۳۴، ۳۷]

مدل و فرضیات پژوهش

در راستای این تحقیق ما به دنبال این موضوع هستیم که آیا استراتژی‌های مختلف بازاریابی در شبکه‌های اجتماعی

[20] Transactional Marketing Strategy
[21] Relationship Marketing Strategy
[22] Database Marketing Strategy
[23] Knowledge-Based Marketing Strategy

بر جلب اعتماد مشتریان تأثیرگذار هستند، تا شرکت‌هایی که از این فضا برای تبلیغات استفاده می‌کنند بتوانند از این استراتژی‌ها بهره گیرند. در واقع در این تحقیق ما به دنبال افزایش اعتماد مشتریان برای تبلیغات در شبکه‌های اجتماعی هستیم. لذا فرضیه اصلی عبارتست از:

فرضیه اصلی- نوع استراتژی بازاریابی در شبکه‌های اجتماعی بر جلب اعتماد مشتریان تاثیر معنی‌داری دارد.

انواع استراتژی‌های بازاریابی الکترونیکی

استراتژی بازاریابی مبادله‌ای

بازاریابی مبادله‌ای، استراتژی کسب‌وکاری است که بر معاملات فروش فردی متمرکز است و بر به حداکثر رساندن کارایی و حجم معامله فروش فردی تاکید دارد تا گسترش روابط طولانی‌مدت با مشتری [۳۰]. در این نوع استراتژی رابطه بین فروشنده و خریدار یک طرفه است، به طوری که فروشنده نقش فعال و مشتری نقش منفعل را در فرایند اعمال می‌کند. در این استراتژی، شرکت، کالاها و خدمات تولیدی را به بازار ارائه می‌دهد و خریداران بدون برقراری تقریبا هیچگونه تماس شخصی به صورت پیوسته اقدام به تامین نیازمندیهای خود از بازار می‌نماید. تمرکز اصلی بازاریابی بر محصول یا برند و آمیخته بازاریابی[۲۴] (محصول/ خدمت، قیمت، ترفیع و توزیع) می‌باشد. در این نوع بازاریابی، تمرکز بر ترفیع فروش و تلاش برای جذب مشتریان جدید بیشتر است [۳۲]. افق زمانی این استراتژی کوتاه‌مدت است و تاکید کمی بر خدمات به مشتریان دارد. به‌طور کلی، در این نوع استراتژی روابط مشتری و شرکت بسیار کم است. این به این دلیل است که ارتباط با مشتریان کم و بنابراین، تعهد مشتری نسبت به برند نیز رضایت‌بخش نیست [۲۹].

با توجه به مطالب بیان شده، در این پژوهش تاثیر استفاده از استراتژی بازاریابی مبادله‌ای در شبکه‌های اجتماعی بر جلب اعتماد مشتریان بررسی خواهد شد. با توجه به این امر فرضیه نخست این پژوهش به شرح زیر است:

فرضیه ۱. استراتژی بازاریابی مبادله‌ای در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارد.

استراتژی بازاریابی رابطه‌ای

با خلاصه کردن طیف گسترده‌ای از تعاریف بازاریابی رابطه‌ای که می‌توان در ادبیات یافت شاید بتوان بازاریابی رابطه‌ای را این‌گونه بیان کرد، بازاریابی رابطه‌ای؛ شناسایی و ایجاد، حفظ و افزایش و همچنین در صورت لزوم، خاتمه دادن به روابط سودمند با مشتریان و سایر ذینفعان است، به طوری که اهداف طرفین محقق گردد و این امر با تبادل متقابل و تحقق وعده‌ها انجام می‌شود [۳۵]. بازاریابی رابطه‌ای شامل کلیه فعالیت‌های بازاریابی در جهت ایجاد، توسعه و حفظ مبادلات رابطه‌ای موفق است. روابط آنلاین شامل مبادلات رابطه‌ای بین مشتری و شرکت است که به واسطه فناوری اینترنت انجام می‌شوند و در یک محیط غیر چهره به چهره (به عنوان مثال، تعاملات انسان با فناوری) صورت می‌گیرد. ویژگی‌های تعیین‌کننده که روابط آنلاین را از روابط آفلاین متمایز می‌کند شامل فاصله جغرافیایی، استقلال زمانی، در دسترس بودن گزینه جهانی و سطح بالاتری از ناشناس بودن، و همچنین یک محیط غنی از داده‌ها است [۳۶].

در واقع، تمرکز بازاریابی رابطه‌ای، شناسایی و حفظ بلندمدت مشتریان است به دلیل اینکه حفظ و جلب رضایت مشتریان موجود نسبت به جذب مشتریان جدید سودآورتر است. این استراتژی به جای مبادلات اقتصادی یا اطلاعاتی، بر روابط بین خریدار و فروشنده به صورت فردی تاکید دارد. استراتژی بازاریابی رابطه‌ای بلندمدت است و بر کیفیت ارائه خدمات به مشتریان تاکید دارد که منجر به حفظ مشتری می‌شود [۲۹].

در این پژوهش تاثیر استفاده از استراتژی بازاریابی رابطه‌ای در شبکه‌های اجتماعی بر جلب اعتماد مشتریان بررسی خواهد شد. با توجه به این امر فرضیه دوم پژوهش به این صورت شکل می‌گیرد:

فرضیه ۲. استراتژی بازاریابی رابطه‌ای در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارد.

استراتژی بازاریابی پایگاه داده

بازاریابی پایگاه داده "شکلی از بازاریابی مستقیم است که از پایگاه‌های داده مشتریان موجود و بالقوه برای ایجاد ارتباطات شخصی به منظور ترویج محصول یا خدمات برای اهداف بازاریابی استفاده می‌کند" [۳۸]. همچنین بازاریابی پایگاه داده را می‌توان به عنوان "استفاده از پایگاه داده مشتریان به منظور افزایش بهره‌وری بازاریابی از طریق کسب، حفظ و توسعه مشتریان به طور موثر تعریف کرد" [۳۹].

[24] 4P (product/service, price, promotion & place)

بازاریابی پایگاه داده توسط مجموعه‌ای از ابزارهای اطلاعاتی و تکنولوژیکی ایجاد شده است و هدف آن کسب اطلاعات (بخصوص اقتصادی) در مورد مشتریان و استفاده از آن به منظور افزایش سود، جلب رضایت مشتریان و کسب وفاداری بیشتر است [۳۲]. ایجاد و مدیریت موثر پایگاه داده مشتری، رویکرد شخصی به هر مشتری را امکان پذیر می‌سازد. بازاریابی پایگاه داده منبع مزیت رقابتی است. شرکت به ویژه در بازاریابی پایگاه داده دامنه وسیعی از اطلاعات در مورد مشتریان را داراست و بر این اساس می‌تواند بهترین خدمات یا محصولات را ارائه و امکان ترویج موثر محصولات و یا خدمات به مشتریان را نسبت به رقبا فراهم کند [۱].

لذا در این پژوهش تاثیر استفاده از استراتژی بازاریابی پایگاه داده در شبکه‌های اجتماعی بر جلب اعتماد مشتریان مورد بررسی قرار می‌گیرد. با توجه به این امر فرضیه سوم پژوهش عبارت است از:

فرضیه ۳. استراتژی بازاریابی پایگاه داده در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارد.

استراتژی بازاریابی مبتنی بر دانش

در این استراتژی، شرکت سعی می‌کند روابط بلندمدت دو جانبه سودمند با مشتریان ایجاد کند و از پروفایل مشتریان استفاده کند. زمانی که شرکتی بتواند یک پایگاه اطلاعاتی قوی از مشتریان، رقبا و سایر شرایط و متغیرهای محیطی ایجاد نموده و بتواند با استفاده از این پایگاه اطلاعاتی، با مشتریان خود رابطه یک به یک برقرار نماید، دارای استراتژی بازاریابی مبتنی بر دانش می باشد. در شرکت‌هایی که از استراتژی مبتنی بر دانش در بازاریابی استفاده می‌کنند، قوانین بازاریابی به طور مستمر مورد بازنگری واقع شده و تحت تاثیر پیشرفت فناوری می‌باشد. استراتژی شرکت، جذب مشتریان جدید توام با حفظ مشتریان موجود بوده و شرکت تعامل بالا و قابل توجهی با مشتریان آگاه خود دارد، مشتریان به صورت فردی نگریسته می‌شوند و محصولات سفارشی است [۳۱].

هوشمندی اصطلاح کلیدی در بازاریابی مبتنی بر دانش می‌باشد. هوشمندی بازاریابی[25] فرآیند شناخت و درک رقبا و قرار گرفتن یک گام بالاتر از آنان از طریق جمع آوری اطلاعات قابل کاربرد در مورد رقبا و بکارگیری آن در برنامه‌ریزی استراتژیک کوتاه و بلندمدت می‌باشد [۴۰، ۴۱].

داده کاوی[26] مفهوم دیگری است که در بازاریابی مبتنی بر دانش مطرح می‌شود. داده کاوی اغلب به کشف دانش در پایگاه داده که فرایند استخراج اطلاعات جالب توجه و یا الگوهایی از داده در پایگاه داده‌های بزرگ است اشاره دارد [۴۱]. بازاریابی مبتنی بر دانش، داده و مدل محور است که از ابزارها و روش‌های داده کاوی پیشرفته برای کشف دانش از پایگاه داده‌های مبادلاتی و مشتری استفاده می‌کند. در کل، این امر منجر به برقراری ارتباط کارا و اثربخش با خریداران بالقوه و افزایش سود می‌شود. رویکرد مهم برای بازاریابی مبتنی بر دانش، درک مشتری و الگوهای رفتاری آنهاست. این امر به ویژگی‌های مبادله‌ای مانند خریدهای اخیر، تعداد دفعات خرید، اندازه خرید، شناسایی گروه‌های مشتری و پیش‌بینی خرید نیاز دارد [42]. تکنیک‌های بسیاری برای استخراج داده‌ها وجود دارد و رایج‌ترین آنها، شبکه‌های عصبی مصنوعی[27]، درخت تصمیم[28] و روش نزدیکترین همسایه[29] می‌باشد. استفاده از تکنیک‌های داده کاوی به تصمیم‌گیرندگان کمک خواهد کرد تا در مورد ترجیحات و نیازهای مشتریان آگاهی یابند [۳۷].

بر این اساس در این پژوهش تاثیر استفاده از استراتژی بازاریابی مبتنی بر دانش در شبکه‌های اجتماعی بر جلب اعتماد مشتریان بررسی می‌شود. با توجه به این امر فرضیه چهارم پژوهش به شرح زیر است:

فرضیه ۴. استراتژی بازاریابی مبتنی بر دانش در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارد.

ویژگی‌های استراتژی‌های بازاریابی در شبکه‌های اجتماعی با توجه به مطالعات صورت گرفته در جدول ۲ ارائه شده است.

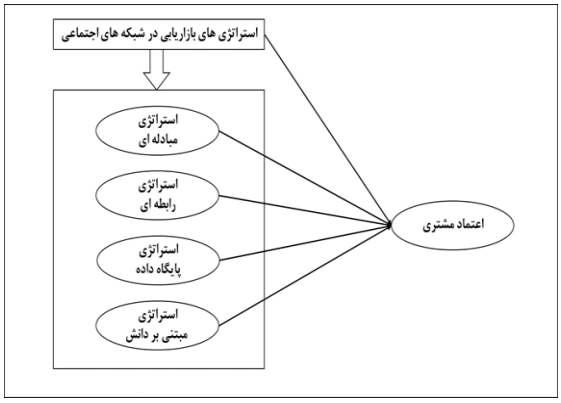
[25] Marketing intelligence

[26] Data mining
[27] Artificial neural networks
[28] Decision trees
[29] The nearest-neighbor method

جدول ۲. ویژگی‌های استراتژی‌های بازاریابی در شبکه‌های اجتماعی

استراتژی / ویژگی	مبادله‌ای	رابطه‌ای	پایگاه داده	مبتنی بردانش
تمرکز	محصول	ارتباطات میان فردی نزدیک بر اساس تعهد و اعتماد	اطلاعات	دانش
طرف‌های درگیر	شرکت و خریداران در بازار کل	شرکت و خریداران فردی (رابطه دو طرفه)	شرکت و خریداران در یک بازار هدف خاص	شرکت، خریداران و سایر شرکت‌ها
توازن قدرت	فروشنده فعال - خریداران منفعل	شرکت و خریدار متقابلا فعال و سازگار	فروشنده فعال - خریداران کمتر منفعل	شرکت و خریداران فعال و سازگار
هدف مدیریتی	جذب مشتری (جلب رضایت مشتری به سود)	تعامل با مشتری	شناخت مشتری	کسب دانش از مشتری
تمرکز تصمیم‌گیری	محصول / برند و 4Ps	روابط بین مشتریان و شرکت‌ها	محصول/ برند و مشتریان (در یک بازار هدف)	محصول/ برند و مشتریان (در یک بازار هدف) و فناوری اطلاعات
تمرکز بازاریابی	فروش	ذهن مشتری	رفتار مشتری	ذهن بازار (هوش رقابتی)
هدف از بکارگیری در بازاریابی	کاهش هزینه‌ها	ایجاد ارتباط با مشتریان	جمع آوری اطلاعات مشتریان	ایجاد سیستم بازاریابی هوشمند
عامل مزیت رقابتی	معاملات / حجم فروش و جذب مشتریان جدید، هزینه محصول / خدمت می‌باشد	حفظ مشتریان موجود، رضایت مشتری، خلاقیت، روابط نزدیک و دراز مدت، سازگاری، قرار دادن مشتری در مرکز سازمان	کسب اطلاعات صحیح، رضایت مشتری در بلندمدت	مبادلات، جذب مشتریان جدید و همچنین حفظ مشتریان موجود، رضایت مشتری، روابط نزدیک با مشتریان، نوآوری
تعاملات با مشتری	پایین	بالا	پایین	بالا و قابل توجه
ارتباطات بازاریابی	شرکت با بازار کل	شرکت با تک تک مشتریان	شرکت با بخش‌های بازار	شرکت با تک تک مشتریان
نوع ارتباط	غیرشخصی	چهره به چهره، غیر شخصی	شخصی	شخصی
گوشه‌های بازار	گوشه‌های بازار قابل شناسایی نیست	گوشه‌های بازار به آسانی قابل شناسایی است	گوشه‌های بازار به آسانی قابل شناسایی است	گوشه‌های بازار به طور کلی بزرگ است، همچنین این گوشه‌ها به راحتی قابل شناسایی هستند
هزینه تبلیغات	بسیار بالا	بالا	پایین	پایین
توسعه محصول	بازاریابی محرک توسعه محصول است	روابط بین شرکت‌ها محرک توسعه محصول است	گردآوری اطلاعات صحیح، محرک توسعه محصول است	سیستم‌های اطلاعاتی و بازاریابی محرک توسعه محصول است
دیدگاه شرکت درمورد رفتار مشتری	مشتریان رفتار اقتصادی دارند	مشتریان رفتار اجتماعی دارند	مشتریان رفتار پیچیده دارند	مشتریان آگاه و هوشمند هستند

با توجه به ادبیات و فرضیه‌های پژوهش، مدل مفهومی پژوهش در شکل ۱ نشان داده شده است.



شکل ۱. مدل مفهومی پژوهش

۱- روش تحقیق

این پژوهش بر مبنای هدف، کاربردی و بر مبنای نحوه گردآوری داده‌ها، پژوهش توصیفی - پیمایشی و از نوع پژوهش همبستگی می‌باشد. جامعه آماری این پژوهش تمامی کاربران ایرانی عضو سایت‌های شبکه‌های اجتماعی (اینستاگرام، تلگرام، واتس اپ) است که تحت تاثیر تبلیغات شرکت‌ها قرار می‌گیرند. با توجه به گستردگی جامعه آماری، از روش نمونه‌گیری غیراحتمالی گلوله برفی استفاده شده است و حجم نمونه آماری ۴۴۶ نفر است. نمونه انتخابی تحقیق، کاربرانی بوده‌اند که توالی خرید گاهی به بالا (خرید بیشتر از ۳ بار در سال) داشته‌اند. پرسشنامه آنلاین در قالب مقیاس ۵ امتیازی لیکرت طراحی شده است. این پژوهش مربوط به بازه زمانی بهار و تابستان ۱۳۹۵ می‌باشد. در این پژوهش، جهت تعیین روایی پرسشنامه از روش اعتبار محتوا و برای سنجش پایایی از ضریب آلفای کرونباخ استفاده شده است. مقدار آلفای کرونباخ به دست آمده در این پژوهش از ۰/۷ بیشتر بوده و بر این اساس پایایی پرسشنامه تایید می‌گردد. جدول ۳ نتایج محاسبه ضرایب آلفای کرونباخ متغیرهای پژوهش را نشان می‌دهد.

جدول ۳- نتایج محاسبه ضرایب آلفای کرونباخ متغیرهای پژوهش

متغیرها			تعداد نمونه	آلفای کرونباخ
اعتماد	X_1	توانایی	۲۰	۰/۹۰۷
		خیرخواهی	۲۰	۰/۸۸۱
		صداقت	۲۰	۰/۸۸۲
		قابلیت پیش‌بینی	۲۰	۰/۷۲۸
استراتژی‌های بازاریابی در شبکه‌های اجتماعی	X_2	استراتژی مبادله‌ای	۲۰	۰/۸۶۹
		استراتژی‌های رابطه‌ای	۲۰	۰/۸۴۶
		استراتژی پایگاه داده	۲۰	۰/۷۳۰
		استراتژی مبتنی بر دانش	۲۰	۰/۸۰۰
کل پرسشنامه			۲۰	۰/۹۵۲

۱. تحلیل داده‌ها و یافته‌های پژوهش

در این پژوهش جهت تعیین نرمال بودن توزیع متغیرها از آزمون کولموگروف – اسمیرنوف استفاده شده است. نتایج آزمون کولموگروف – اسمیرنوف در جدول ۴ برای هریک از متغیرهای پژوهش آورده شده است.

جدول ۴. آزمون کولموگروف- اسمیرنوف برای متغیرهای پژوهش

	اعتماد	مبادله‌ای	رابطه‌ای	پایگاه داده	مبتنی بر دانش
تعداد داده ها	۴۴۶	۴۴۶	۴۴۶	۴۴۶	۴۴۶
مقدار آماره Z	۳/۴۶۶	۴/۳۴۹	۴/۶۵۰	۴/۷۲۷	۳/۸۹۱
سطح معنی داری (Sig)	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰

با توجه به اینکه مقدار Sig برای همه متغیرهای پژوهش کمتر از ۰/۰۵ است، لذا نمی‌توان فرض نرمال بودن توزیع مربوط به این متغیرها را پذیرفت. در نتیجه برای انجام آزمون‌های مربوط به این متغیرها از آزمون‌های ناپارامتریک استفاده شده است.

۱-۴- آزمون فرضیات پژوهش

با توجه به اینکه مدل پژوهش حاضر، روابط بین چندین متغیر پنهان (متغیرهای اصلی پژوهش) را بطور همزمان

اندازه‌گیری می‌کند، بنابراین به منظور تحلیل داده‌ها و آزمون فرضیات از مدل‌سازی معادلات ساختاری استفاده شده است. از آنجا که فرض نرمال بودن توزیع برای اکثر متغیرهای پژوهش حاضر برقرار نبود، لذا در این پژوهش برای کسب نتایج دقیق‌تر، برای آزمون مدل مفهومی پژوهش، از روش پی ال اس[30] که یک تکنیک مدل‌سازی مسیر واریانس محور است و امکان بررسی نظریه و سنجه‌ها را به طور همزمان فراهم می‌سازد استفاده شده است [۴۳]. محاسبات مربوط به تحلیل داده‌ها و آزمون فرضیات بوسیله نرم‌افزار اسمارت پی ال اس[31] صورت گرفته است. در این روش دو مدل مورد بررسی قرار می‌گیرد: ۱) مدل بیرونی[32] که برای بررسی روابط میان شاخص‌ها (سؤالات پژوهش) با متغیرهای اصلی مربوط به خود استفاده می‌شود که در واقع معادل همان مدل اندازه گیری در روش‌های کوواریانس محور است. ۲) مدل درونی[33] که بخش ساختاری مدل را می‌سنجد و برای بررسی روابط میان متغیرهای پنهان (متغیرهای اصلی) که فرضیات پژوهش از روابط میان آنها شکل می‌گیرد، بکار می‌رود.

مدل بیرونی (اندازه‌گیری). برای بررسی مدل اندازه‌گیری از آزمون‌های روایی همگرا[34] و و روایی تشخیصی[35] استفاده می‌شود [۴۴]. سه معیار اصلی که با آن روایی همگرا برای مدل اندازه‌گیری می‌تواند مورد سنجش قرار گیرد عبارتند از: ۱. برای بررسی پایایی (همسانی درونی) هر یک از گویه‌ها، بارهای عاملی محاسبه شده در بخش اندازه‌گیری مدل بیشتر از ۰/۵ باشد. ۲. مقادیر محاسبه شده برای پایایی ترکیبی[36] (CR) (برای همه سازه‌ها) در اینجا متغیرهای پنهان بایستی بیشتر از مقدار معیار ۰/۷ باشد. پایایی ترکیبی به این معنا است که بین پاسخ دهندگان مختلف مورد مطالعه برداشت یکسانی از سوالات وجود داشته است. ۳. میانگین واریانس استخراج شده[37] (AVE)، برای سازه‌های پژوهش بایستی بالاتر از مقدار معیار ۰/۵ باشد [۴۴، ۴۳].

[30] Partial Least Squares
[31] SmartPLS
[32] Outer model
[33] Inner model
[34] Convergent validity
[35] Discriminant validity
[36] Composite Reliability (CR)
[37] Average Variance Extracted (AVE)

جدول ۵- مقادیر محاسبه شده برای نشانگرهای هر سازه در قالب مدل اندازه گیری

سازه	علامت نشانگر	بار عاملی	عدد معنی‌داری (T)	پایایی ترکیبی (CR)	مقدار میانگین واریانس استخراج شده (AVE)	ضریب تعیین (R2)
اعتماد	q1	۰/۸۷۳	۷۰/۶۰۹	۰/۹۶۳	۰/۷۰۶	۰/۶۱۳
	q2	۰/۸۴۸	۵۲/۵۸۷			
	q3	۰/۸۳۳	۵۱/۶۸۷			
	q4	۰/۸۷۴	۷۱/۹۹۲			
	q5	۰/۸۷۰	۶۹/۶۸۸			
	q6	۰/۸۳۶	۵۷/۱۹۲			
	q7	۰/۸۱۲	۴۹/۷۸۹			
	q8	۰/۸۳۶	۶۰/۲۵۰			
	q9	۰/۹۰۶	۱۰۹/۷۴۸			
	q10	۰/۷۸۷	۳۳/۹۷۷			
	q11	۰/۷۵۶	۲۸/۱۵۴			
مبادله‌ای	q12	۰/۶۶۹	۱۹/۸۹۳	۰/۹۳۳	۰/۷۸۰	-
	q13	۰/۹۴۲	۱۸۵/۰۹۶			
	q14	۰/۹۴۴	۱۵۰/۰۲۳			
	q15	۰/۹۴۷	۱۶۱/۵۹۵			
رابطه‌ای	q16	۰/۹۲۴	۱۰۴/۷۳۹	۰/۹۵۲	۰/۸۶۹	-
	q17	۰/۹۳۱	۱۳۹/۲۶۲			
	q18	۰/۹۴۳	۱۹۲/۶۰۴			
پایگاه داده	q19	۰/۹۱۴	۸۹/۲۲۸	۰/۹۱۳	۰/۷۷۸	-
	q20	۰/۸۲۱	۳۴/۲۳۹			
	q21	۰/۹۰۷	۱۱۲/۸۱۹			
مبتنی بر دانش	q22	۰/۹۱۹	۱۱۵/۹۵۴	۰/۹۴۸	۰/۸۵۹	-
	q23	۰/۹۳۴	۱۲۷/۴۸۹			
	q24	۰/۹۲۷	۱۴۷/۲۷۰			

همانطور که در جدول ۵ ملاحظه می‌شود، مقدار معنی داری کلیه نشانگرها دارای مقدار t بالاتر از ۱/۹۶ است. لذا این نشانگرها، ساختارهای عاملی مناسبی را جهت اندازه گیری ابعاد مورد مطالعه در مدل پژوهش فراهم می‌آورند.

از طرفی مقادیر محاسبه شده برای بارهای عاملی شاخص‌های مشاهده شده، بیشتر از حداقل مقدار معیار ۰/۵ است، مقادیر محاسبه شده برای پایایی ترکیبی (CR)، بیشتر از حداقل مقدار معیار در نظر گرفته شده ۰/۷ بوده و مقادیر محاسبه شده برای میانگین واریانس استخراج شده همه متغیرهای پژوهش بیشتر از ۰/۵ است، در نتیجه با توجه به مقادیر معیار در نظر گرفته شده و مقادیر محاسبه شده می‌توان گفت که مدل اندازه گیری پژوهش حاضر دارای روایی همگرا می‌باشد و ضریب تعیین (R^2) برابر با ۰/۶۱۳ است. مقدار شاخص برازش مدل ۰/۶۹۹ محاسبه شد که از مقدار ۰/۵ بزرگتر شده است و نشان از برازش مناسب مدل دارد. به بیان ساده‌تر داده‌های این پژوهش با ساختار عاملی و زیربنای نظری تحقیق برازش مناسبی دارد و این بیانگر همسو بودن سوالات با سازه‌های نظری است.

در پژوهش حاضر برای بررسی روایی مدل اندازه‌گیری، از روایی تشخیصی نیز استفاده شد. روایی تشخیصی بیانگر درجه‌ای است که در آن، شاخص‌های مشاهده شده مربوط به هر سازه (سوالات پرسشنامه مربوط به هر متغیر پژوهش) فقط سازه‌های مربوط به خود را می‌سنجند [۴۵]. در بررسی روایی تشخیصی سازه‌ها دو ملاک مورد بررسی قرار می‌گیرد: الف) بررسی بار تقاطعی گویه‌ها، ب) بررسی همبستگی بین متغیرهای پنهان.

الف) برای بررسی روایی تشخیصی از جدول بارهای عاملی متقاطع[38]، که از خروجی نرم افزار اسمارت پی ال اس قابل محاسبه است، استفاده شد. برای این کار میزان همبستگی هر نشانگر با تمام سازه‌های دیگر مدل محاسبه شد که مقادیر همبستگی باید برای سازه انتخابی پژوهش، بیشتر از سایر سازه‌ها باشد. لذا هر نشانگر باید بیشترین

[38] Cross Loadings

همبستگی را فقط با سازه خود نشان دهد و با سایر سازه‌ها نیز کمترین همبستگی را داشته باشد.

جدول ۶- بررسی روایی تشخیصی از طریق بارهای عاملی متقاطع

	اعتماد	مبادله ای	رابطه ای	پایگاه داده	مبتنی بر دانش
q1	۰/۸۷۳	-۰/۴۹۱	۰/۵۹۳	۰/۵۲۲	۰/۵۹۶
q2	۰/۸۴۸	-۰/۵۳۷	۰/۶۳۳	۰/۵۳۴	۰/۶۶۸
q3	۰/۸۳۳	-۰/۵۱۴	۰/۶۲۰	۰/۴۹۲	۰/۶۱۲
q4	۰/۸۷۴	-۰/۵۵۹	۰/۶۱۴	۰/۵۴۲	۰/۶۰۱
q5	۰/۸۷۰	-۰/۵۶۴	۰/۶۱۸	۰/۵۴۲	۰/۶۰۰
q6	۰/۸۳۶	-۰/۵۳۰	۰/۵۶۸	۰/۴۵۵	۰/۶۰۸
q7	۰/۸۱۲	-۰/۵۰۰	۰/۵۶۶	۰/۳۹۴	۰/۶۴۰
q8	۰/۸۳۶	-۰/۴۹۷	۰/۵۸۹	۰/۴۹۸	۰/۶۳۳
q9	۰/۹۰۶	-۰/۵۸۳	۰/۶۷۵	۰/۵۶۴	۰/۶۸۹
q10	۰/۷۸۷	-۰/۵۳۸	۰/۶۱۷	۰/۵۸۷	۰/۵۵۹
q11	۰/۷۵۶	-۰/۴۹۱	۰/۵۸۰	۰/۵۰۹	۰/۵۵۰
q12	-۰/۳۵۹	۰/۶۶۹	-۰/۵۰۵	-۰/۴۳۵	-۰/۳۷۷
q13	-۰/۶۲۷	۰/۹۴۲	-۰/۷۰۱	-۰/۵۲۵	-۰/۶۱۶
q14	-۰/۵۹۰	۰/۹۴۴	-۰/۶۹۸	-۰/۵۳۷	-۰/۶۰۸
q15	-۰/۵۹۹	۰/۹۴۷	-۰/۷۴۰	-۰/۵۴۷	-۰/۶۵۱
q16	۰/۶۶۴	-۰/۶۴۳	۰/۹۲۴	۰/۶۶۵	۰/۷۳۵
q17	۰/۶۶۸	-۰/۷۰۷	۰/۹۳۱	۰/۶۰۸	۰/۷۴۵
q18	۰/۶۹۰	-۰/۷۵۹	۰/۹۴۳	۰/۶۰۴	۰/۷۵۵
q19	۰/۵۲۶	-۰/۵۳۳	۰/۵۹۷	۰/۹۱۴	۰/۴۹۴
q20	۰/۳۷۱	-۰/۳۰۳	۰/۴۱۹	۰/۸۲۱	۰/۳۵۹
q21	۰/۶۵۵	-۰/۶۱۷	۰/۶۹۶	۰/۹۰۷	۰/۶۶۹
q22	۰/۶۶۵	-۰/۵۸۷	۰/۷۲۷	۰/۵۶۰	۰/۹۱۹
q23	۰/۶۵۶	-۰/۵۶۱	۰/۷۴۲	۰/۵۲۰	۰/۹۳۴
q24	۰/۷۱۲	-۰/۶۵۵	۰/۷۵۱	۰/۵۹۲	۰/۹۲۷

بررسی جدول بارهای عاملی متقاطع (جدول ۶) بیانگر قابل قبول بودن روایی تشخیصی مدل پژوهش است. چون مقادیر همبستگی مربوط به نشانگرها، بیشترین مقدار همبستگی را فقط با سازه خود دارند و با بقیه سازه‌ها مقدار همبستگی کمتری دارند.

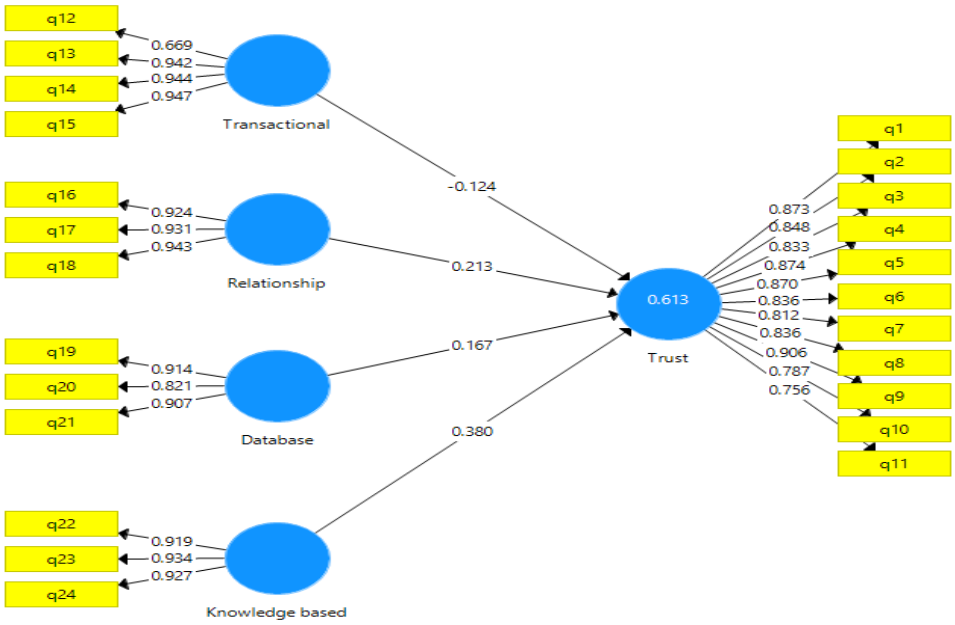
ب) برای بررسی همبستگی بین متغیرهای پنهان از جدول معیار فورنل و لارکر (جدول ۷) استفاده شده است

جدول ۷- بررسی همبستگی بین متغیرهای پنهان (معیار فورنل و لارکر)

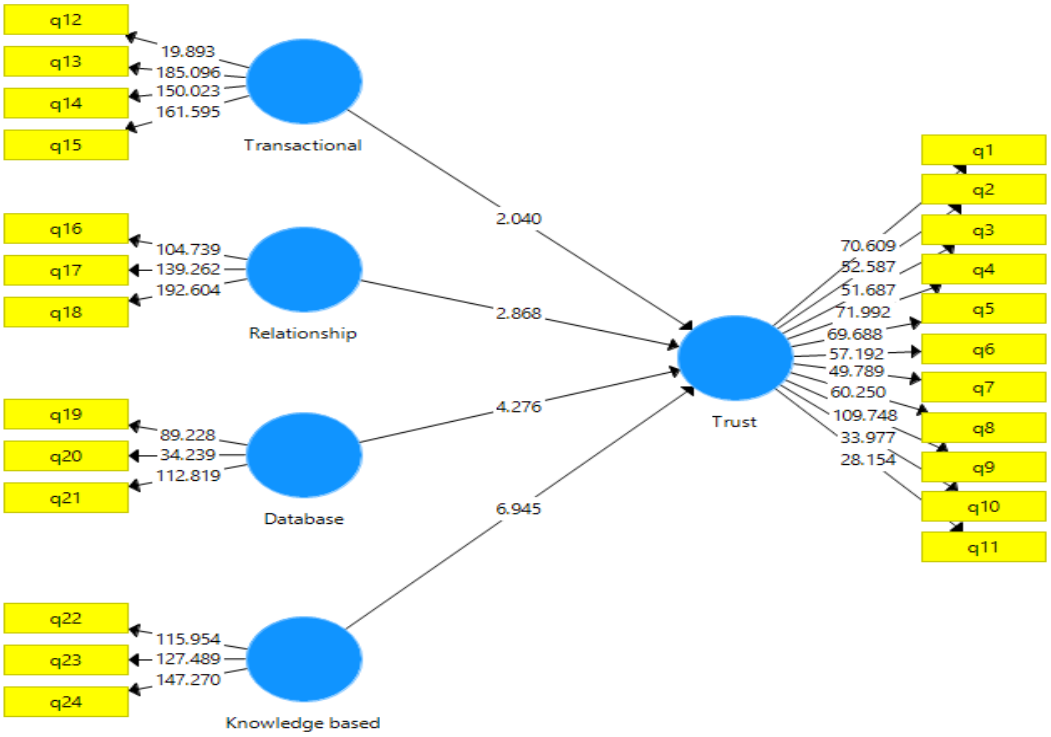
	اعتماد	استراتژی مبادله‌ای	استراتژی رابطه‌ای	استراتژی پایگاه داده	استراتژی مبتنی بر دانش
اعتماد	۰/۸۴۰				
استراتژی مبادله‌ای	-۰/۶۲۹	۰/۸۸۳			
استراتژی رابطه‌ای	۰/۷۲۳	-۰/۷۵۵	۰/۹۳۲		
استراتژی پایگاه داده	۰/۶۱۱	-۰/۵۷۸	۰/۶۷۱	۰/۸۸۲	
استراتژی مبتنی بر دانش	۰/۷۳۲	-۰/۶۵۰	۰/۷۹۹	۰/۶۰۲	**۰/۹۲۷**

جدول ۷ به بررسی ضرایب همبستگی و روایی تشخیصی می‌پردازد. اعداد روی قطر اصلی ماتریس ریشه دوم میانگین واریانس تبیین شده (AVE) را نشان می‌دهد. لازمه تایید روایی تشخیصی بیشتر بودن مقدار ریشه دوم میانگین واریانس تبیین شده از تمامی ضرایب همبستگی متغیر مربوطه با باقی متغیرها است. همان طور که در جدول مشخص است، مقدار ریشه دوم شاخص میانگین واریانس تبیین شده، برای تمامی متغیرها، از همبستگی آن متغیر با سایر متغیرها بیشتر می‌باشد. پایین قطر اصلی ضرایب همبستگی پیرسون نشان داده شده‌اند. ضریب مثبت نشان‌دهنده رابطه‌ی مثبت و ضریب منفی نشان‌دهنده رابطه منفی و معکوس بین دو متغیر می‌باشد.

مدل ساختاری. فرضیات پژوهش در قالب آزمون اسمارت پی ال اس مورد بررسی قرار گرفتند. با در نظر گرفتن نتایج حاصل از بررسی روابط میان متغیرهای مستقل و وابسته با استفاده از ضریب محاسبه شده، می‌توان به بررسی معنی‌داری اثرات میان متغیرهای پژوهش بر یکدیگر که در واقع فرضیات از روابط میان آنها شکل می‌گیرند، پرداخت. با توجه به شکل (۲) که بیانگر نمودار در وضعیت بارهای عاملی روابط بین متغیرها و شکل (۳) که بیانگر نمودار در وضعیت اعداد معنی‌داری روابط بین متغیرهای پژوهش است به بررسی فرضیات پژوهش می‌پردازیم.



شکل ۲- مدل پژوهش در وضعیت بارهای عاملی



شکل ۳- مدل پژوهش در وضعیت اعداد معنی‌داری (T)

در جدول ۸، خلاصه نتایج مربوط به نمودارهای ۲ و ۳ ارائه گردیده است.

جدول ۸- خلاصه نتایج مربوط به بارهای عاملی و معنی‌داری بخش ساختاری مدل

نتیجه	مقدار T	بار عاملی	روابط بین متغیرها
قبول فرضیه	۲/۰۴۰	-۰/۱۲۴	اعتماد مشتریان ..... استراتژی بازاریابی مبادله‌ای
قبول فرضیه	۲/۸۶۸	۰/۲۱۳	اعتماد مشتریان ..... استراتژی بازاریابی رابطه‌ای
قبول فرضیه	۴/۲۷۶	۰/۱۶۷	اعتماد مشتریان ..... استراتژی بازاریابی پایگاه داده
قبول فرضیه	۶/۹۴۵	۰/۳۸۰	اعتماد مشتریان ..... استراتژی بازاریابی مبتنی بر دانش

۲-۴- بررسی نتایج حاصل از آزمون فرضیات پژوهش

H_1: استراتژی بازاریابی مبادله‌ای در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارد.

نتایج بدست آمده نشان می‌دهد که این اثر دارای مقدار بار عاملی ۰/۱۲۴- می‌باشد و با توجه به اینکه مقدار T برای این رابطه (۲/۰۴۰) بیشتر از ۱/۹۶ محاسبه شده است، لذا در سطح اطمینان ۹۵ درصد، تأثیر استراتژی بازاریابی مبادله‌ای در شبکه‌های اجتماعی بر جلب اعتماد مشتریان معکوس و معنی‌دار است.

H_2: استراتژی بازاریابی رابطه‌ای در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارد.

نتایج بدست آمده نشان می‌دهد که این اثر دارای مقدار بار عاملی ۰/۲۱۳ می‌باشد و با توجه به اینکه مقدار T برای این رابطه (۲/۸۶۸) بیشتر از ۱/۹۶ محاسبه شده است، لذا در سطح اطمینان ۹۵ درصد، تأثیر استراتژی بازاریابی رابطه‌ای در شبکه‌های اجتماعی بر جلب اعتماد مشتریان مستقیم و معنی‌دار است.

H_3: استراتژی بازاریابی پایگاه داده در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارد.

نتایج بدست آمده نشان می‌دهد که این اثر دارای مقدار بار عاملی ۰/۱۶۷ می‌باشد و با توجه به اینکه مقدار T برای این رابطه (۴/۲۷۶) بیشتر از ۱/۹۶ محاسبه شده است، لذا در سطح اطمینان ۹۵ درصد، تأثیر استراتژی بازاریابی پایگاه داده در شبکه‌های اجتماعی بر جلب اعتماد مشتریان مستقیم و معنی‌دار است.

H_4: استراتژی بازاریابی مبتنی بر دانش در شبکه‌های اجتماعی تاثیر معنی‌داری بر جلب اعتماد مشتریان دارد.

نتایج بدست آمده نشان می‌دهد که این اثر دارای مقدار بار عاملی ۰/۳۸۰ می‌باشد و با توجه به اینکه مقدار T برای این رابطه (۶/۹۴۵) بیشتر از ۱/۹۶ محاسبه شده است، لذا در سطح اطمینان ۹۵ درصد، تأثیر استراتژی بازاریابی مبتنی بر دانش در شبکه‌های اجتماعی بر جلب اعتماد مشتریان مستقیم و معنی‌دار است.

۳-نتیجه‌گیری و پیشنهادها

هدف از این پژوهش بررسی تاثیر استفاده از انواع استراتژی‌های بازاریابی در شبکه‌های اجتماعی بر جلب اعتماد مشتریان بوده است. تحقیقات بسیاری در خصوص اعتماد و اهمیت آن در بستر شبکه‌های اجتماعی و تجارت الکترونیک انجام شده است اما با توجه به ادبیات علمی، شواهد و تحقیقاتی درباره استراتژی‌های بازاریابی در بستر شبکه‌های اجتماعی موجود نیست. اینکه چه استراتژی‌هایی در شبکه‌های اجتماعی که امروزه با سرعت فراوان و ضریب نفوذ گسترده زندگی مردم و کسب و کارهای مختلف را احاطه کرده کاربرد دارند و اعتماد مشتریان که موضوع با اهمیتی در بستر شبکه‌های اجتماعی و تعاملات آنلاین مطرح بوده و تاثیر مثبتی بر قصد خرید مشتریان و میزان تعاملات آنها می‌تواند داشته باشد با چه راهکارهایی می‌توان افزایش داد. بنابراین، این مطالعه به درک ما در این زمینه که کدام استراتژی‌های بازاریابی الکترونیکی به بهترین وجه می‌توانند اعتماد مشتریان در بستر شبکه‌های اجتماعی را تقویت کنند، کمک می‌کنند. شرکت‌ها نیز می‌توانند این استراتژی‌ها را پیاده‌سازی کنند. یافته‌های پژوهش نشان می‌دهد که استراتژی‌های بازاریابی مدنظر پژوهش (مبادله‌ای، رابطه‌ای، پایگاه داده، مبتنی بر دانش)، تاثیر معنی‌داری بر جلب اعتماد مشتریان در شبکه‌های اجتماعی دارند و همچنین نتایج بیانگر آن است که استراتژی بازاریابی مبتنی بر دانش بیشترین تاثیر و استراتژی بازاریابی مبادله‌ای کمترین تاثیر را بر جلب اعتماد و جذب مشتریان در شبکه‌های اجتماعی دارد. چرا که در استراتژی بازاریابی

مبتنی بر دانش، شرکت‌ها سعی بر آن دارند تا علاوه بر استفاده از پایگاه داده مشتریان، با در نظر گرفتن رقبا و سایر شرایط و متغیرهای محیطی، روابط بلندمدت دوجانبه سودمند و همچنین تعامل بالا و قابل توجهی با مشتریان خود برقرار نمایند. برقراری روابط متقابل سودمند و تعاملات بالا با مشتریان در استراتژی بازاریابی مبتنی بر دانش، منجر به افزایش اعتماد مشتریان به فروشگاه‌های آنلاین فعال در شبکه‌های اجتماعی خواهد شد و برعکس در استراتژی بازاریابی مبادله‌ای، تعامل پایین با مشتریان و عدم برقراری ارتباط با مشتری و توجه صرف به حجم فروش بیشتر و کاهش هزینه منجر به کاهش اعتماد مشتریان خواهد شد، چرا که امروزه نیازهای ارتباطی و خلاقیت مشتریان در سطوح بالایی قرار دارد. بنابراین در این راستا پیشنهادات زیر در جهت جلب اعتماد و جذب مشتریان در شبکه های اجتماعی پیشنهاد می‌گردد:

با بررسی تاثیر استراتژی بازاریابی مبادله‌ای بر جلب اعتماد مشتریان پیشنهاد می‌شود که فروشگاه‌های آنلاین فعال در شبکه‌های اجتماعی، از این استراتژی برای جلب اعتماد مشتریان استفاده نکرده (با توجه به منفی بودن تاثیر) و سعی کنند در هنگام معرفی محصول، اطلاعات کامل‌تری درباره محصول ارائه داده و پاسخگوی سوالات مشتریان باشند، با مشتریان پس از انجام خرید تعامل مجدد برقرار کرده، تنها بر میزان فروش تمرکز نداشته و حفظ مشتریان نیز برای آنها با اهمیت باشد چرا که در اینصورت می‌توان اعتماد مشتریان را در شبکه‌های اجتماعی جلب کرد.

با بررسی تاثیر استراتژی بازاریابی رابطه‌ای بر جلب اعتماد مشتریان پیشنهاد می‌شود که فروشگاه‌های آنلاین فعال در شبکه‌های اجتماعی با توجه به تاثیر مثبت این استراتژی بر جلب اعتماد مشتریان، در بکارگیری آن به مواردی همچون ارتباطات میان فردی نزدیک براساس تعهد و اعتماد، ارائه محصولات مبتنی بر نیازهای مشتریان، تماس مجدد با مشتریان پس از انجام خرید و حفظ ارتباط با آنها توجه داشته تا بتوانند اعتماد مشتریان را جلب نمایند. برقراری و حفظ ارتباط با مشتری منجر به افزایش جلب اعتماد مشتریان خواهد شد.

با بررسی تاثیر استراتژی بازاریابی پایگاه داده بر جلب اعتماد مشتریان پیشنهاد می‌شود که فروشگاه‌های آنلاین فعال در شبکه‌های اجتماعی با بکارگیری استراتژی بازاریابی پایگاه داده (با توجه به تاثیر مثبت بر جلب اعتماد مشتریان) به جمع آوری اطلاعات در خصوص مشتریان به عنوان مثال آنچه که برای مشتریان جالب و مهم است، نیازها، تمایلات، عادت‌ها و مقدار صرف هزینه مشتریان تمرکز داشته و به دنبال کسب رضایت آنها باشند تا بتوانند اعتماد مشتریان را جلب کنند.

با بررسی تاثیر استراتژی بازاریابی مبتنی بر دانش بر جلب اعتماد مشتریان پیشنهاد می‌شود که فروشگاه‌های آنلاین فعال در شبکه‌های اجتماعی با توجه به تاثیر مثبت این استراتژی بر جلب اعتماد مشتریان با استفاده از تکنیک‌های داده کاوی، نیازها و ترجیحات، سلایق و خواسته‌های مشتریان را شناسایی کرده، برای هر مشتری پروفایلی تشکیل دهند تا از تعداد دفعات خریداری محصول یا تعداد دفعات بازدید از فروشگاه، حجم خرید و میزان صرف هزینه مشتریان آگاهی داشته و گروه‌های بارز مشتریان را شناسایی کنند و در بکارگیری استراتژی بازاریابی مبتنی بر دانش به مواردی از قبیل روابط بسیار نزدیک و یک به یک با مشتری، سفارشی‌سازی و منطبق کردن محصول با نیاز مشتری و پاسخگویی بهتر از رقبا نسبت به نیازهای مشتریان، جذب مشتریان جدید توام با حفظ مشتریان موجود و تعامل بالا و قابل توجه با آنها توجه داشته و با برقراری روابط یک به یک و بلندمدت با مشتریان بتوانند اعتماد آنها را جلب نمایند.

با در نظرگرفتن این موضوع که تحقیقاتی در زمینه شناسایی استراتژی‌های بازاریابی کلیدی در بستر شبکه های اجتماعی انجام نشده، پیشنهاد می‌شود با توجه به نفوذ گسترده این شبکه‌ها در کسب و کارها و اهمیت آن، محققین آتی مطالعات بیشتری در این خصوص داشته باشند، از روش‌های جدیدتری مانند رویکرد هوش مصنوعی، شبکه عصبی و ... برای شناسایی این استراتژی‌ها و نیز بررسی عوامل موثر بر جلب اعتماد مشتریان در این بستر استفاده کنند. همچنین با توجه به اینکه در این تحقیق استراتژی‌های بازاریابی مبادله‌ای، رابطه‌ای، پایگاه داده و مبتنی بر دانش بر جلب اعتماد مشتریان در شبکه‌های اجتماعی بررسی گردید،

محققین آتی می‌توانند استراتژی‌های دیگری که بر جلب اعتماد مشتریان در شبکه‌های اجتماعی تاثیر دارند را نیز مورد شناسایی قرار دهند. پیشنهاد می‌شود در مطالعات بعدی هر کدام از استراتژی‌های بازاریابی مبادله‌ای، رابطه‌ای، پایگـاه داده و مبتـنی بر دانـش را در صـنایع مخـتلف مورد بررسی قرار داده و استراتژی‌های متناسب با صنایع را شناسایی کنند و از آنجایی که در این پژوهش، بررسی انواع استراتژی‌های بازاریابی در سایت‌های شبکه‌های اجتماعی مطرح بوده، پیشنهاد می‌گردد در تحقیقات بعدی، بررسی استراتژی‌ها بصورت مطالعه موردی انجام گیرد.

منابع

۱. اکبری، محسن، زاهدفر، کامران، ایاغ، زهرا، "درک اعتماد اولیه کاربران به شبکه‌های اجتماعی"، فصلنامه علمی-پژوهشی فناوری اطلاعات و ارتباطات ایران، شماره‌های ۲۷ و ۲۸، صفحات ۹۴-۸۱، ۱۳۹۵

1. emarketer.com. Available: https://www.emarketer.com/content/global-social-network-users.2020.
2. G. Appel1, L. Grewal, R. Hadi, and A. T. Stephen, "The future of social media in marketing," *Journal of the Academy of Marketing Science*, 48, 79-95, 2020. https://doi.org/10.1007/s11747-019-00695-1.
3. A. Saxena, and Kh. Uday, "Advertising on Social Nework Sites: A Structural Equation Modelling Approach," *Vision*, 17(1), 17-25, 2013.
4. A. Alalwan, N. P. Rana, Y. K. Dwivedi, and R. Algharabat, "Social media in marketing: A review and analysis of the existing literature," *Telematics and Informatics*, 2017. http://dx.doi.org/10.1016/j.tele.2017.05.008.
5. J. Fogel, and E. Nehmad, "Internet social network communities: risk taking, trust, and privacy concerns," *Computers in Human Behavior*, 25(1), 153-160, 2009.
10. I. Gregurec, T. Vranesevic, and D. Dobrinic, "The importance of database marketing in social network advertising," *In Proceedings of the International Journal of Management Cases, 27th-29th April, (pp.165-172). University of Dubrovnik, Croatia: Access Press UK*, 13(4), 2011.

1. C-C. Wu, Y. Huang, and C-L. Hsu, "Benevolence trust: a key determinant of user continuance use of online social networks," *Information Systems and e-Business Management,* 12(2), 189-211, 2014.
2. N. Hajli, J. Sims, A. H. Zadeh, and M-O. Richard, "A social commerce investigation of the role of trust in a social networking site on purchase intentions," *Journal of Business Research*, 71, 133-141, 2017. http://dx.doi.org/10.1016/j.jbusres.2016.10.004.
3. T. Oliveira, M. Alhinho, P. Rita, and G. Dhillon, "Modelling and Testing Consumer Trust Dimensions in E-commerce," *Computers in Human Behavior,* 2017. doi: 10.1016/j.chb.2017.01.050.
4. L-Y. Leong, T-S. Hewb, K-B. Ooic, A. Y-L. Chongd, "Predicting the antecedents of trust in social commerce – A hybrid structural equation modeling with neural network approach," *Journal of Business Research*, 110, 24-40, 2020.
5. J. Yeon, I. Park, D. Lee, "What creates trust and who gets loyalty in social commerce," *Journal of Retailing and Consumer Services*, 50, 138-144, 2019.
6. S. Grabner-Kräuter, and S. Bitter, "Trust in online social networks: A multifaceted perspective," *Forum for Social Economics*, 44(1), 48-68, 2015. DOI:10.1080/07360932.2013.781517.
7. A. Palmer, and Q. Huo, "A study of trust over time within a social mediated

environment," *Journal of Marketing Management*, 29, 15-16, 1816-1833, 2013. DOI:10.1080/0267257X.2013.803143.

۱۵.رحمانی، مرضیه، فتحیان، محمد، یعقوبی، سعید، "ارائه یک مدل مفهومی جهت شناسایی کیفیت خدمات الکترونیکی در فروشگاه‌های اینترنتی کشور"، *فصلنامه علمی- پژوهشی فناوری اطلاعات و ارتباطات ایران*، شماره‌های ۲۹ و ۳۰، صفحات ۱۴-۱، ۱۳۹۵.

۱۶.مصممی، شیلا، مرادی، محمود، شاه بهرامی، اسدالله، "بررسی عوامل موثر بر موفقیت تجارت الکترونیک B2C از دیدگاه مدیریتی"، *فصلنامه علمی- پژوهشی فناوری اطلاعات و ارتباطات ایران*، شماره‌های ۱۷ و ۱۸، صفحات ۷۲-۵۷، ۱۳۹۲.

17. Y.D. Wang, and H.H. Emurian, "An Overview of Online Trust: Concepts, Elements, and Implications," *Computers in Human Behavior*, 21(1), 105-125, 2005.

18. R. C. Mayer, J. H. Davis, and F. C. Schoorman, "An integrative model of organizational trust," *Academy of Management Review*, 20(3), 709-734, 1995.

19. F. Tan, and P. Sutherland, "Online Consumer Trust: A Multi-Dimensional Model," *The Journal of Electronic Commerce in Organizations*, 2(3), 40-58, 2004.

20. R. Larzelere, and T. Huston, "The dyadic trust scale: Toward understanding interpersonal trust in close relationships," *Journal of Marriage and the Family*, 42, 595-604, 1980.

21. D. H. McKnight, and N. L. Chervany, "What trust means in E-commerce customer relationships," *International Journal of Electronic Commerce*, 6(2), 35-59, 2002.

22. J.-J. Wu, Y.-H. Chen, and Y.-S. Chung, "Trust factors influencing virtual community members: A study of transaction communities," *Journal of Business Research*, 63, 1025-1032, 2010.

23. d. Boyd, and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship," Journal of Computer-Mediated Communication, 13(1), 210–230, 2007.

24. M.G. Hoy, and G. Milne, "Gender differences in privacy-related measures for young adult Facebook users," *Journal of Interactive Advertising*, 10(2), 28-45, 2010.

25. T. Fisher, "ROI in social media: A look at the arguments," *Journal of Database Marketing & Customer Strategy Management*, 16(3), 189-195, 2009.

26. L. Kelly, G.F. Kerr, and D. Judy, "Avoidance of Advertising in Social Networking Sites: The Teenage Perspective," *Journal of Interactive Advertising*, 10(2), 16-27, 2010.

27. G. Shen, J-S. Chiou, C-H. Hsiao, C-H. Wanga, and H-N. Li, "Effective marketing communication via social networking site: The moderating role of the social tie," *Journal of Business Research*, 69(6), 2265-2270, 2015.

28. P. Chailom, "Antecedents and consequences of e-marketing strategy: evidence from e-commerce business in Thailand," *International Journal of Business Strategy*, 12, 75-87, 2012.

29. T. Šonkova, & M. Grabowska, "Customer engagement: transactional vs. relationship marketing," *Journal of International Studies*, 8(1), 196-207, 2015.

30. N. Shaik, and SH. Ritter, *Social Media Based Relationship Marketing*. IGI Global, Chapter 9, 119–139, 2011.

31. M. Zineldin, "Beyond relationship marketing: technologicalship marketing," *Marketing Intelligence & Planning*, 18(1), 9-23, 2000.

32. R. J. Brodie, N. E. Coviello, R. W. Brookes, and V. Little, "Towards a Paradigm Shift in Marketing? An Examination of Current Marketing Practices," *Journal of Marketing Management*, 13(5), 383-406, 1997.

۳۳. سیدجوادین، سید رضا، اسفیدانی، محمدرحیم، آقازاده، هاشم، "بررسی استراتژی‌های بازاریابی در بازارهای الکترونیکی مورد مطالعه: شرکت‌های تولیدی برتر ایران"، *فصلنامه پژوهشنامه بازرگانی*، شماره ۴۲، صفحات ۱۱۷-۸۳، ۱۳۸۶.

34. H. Aghazadeh, and M.R. Esfidani, "Internet Marketing Strategies", *Iranian Economic Review*, 12(1), 179-191, 2007.
35. A. Moretti, and A. Tuan, "Social media marketing and relationship marketing: revolution or evolution? A first step analysis," *sinergie, rivista di studi e ricerche*, 93, 115-137, 2014.
36. L. Steinhoff, D. Arli, S. Weaven, and I. V. Kozlenkova, "Online relationship marketing," *Journal of the Academy of Marketing Science*, 2018. https://doi.org/10.1007/s11747-018-0621-6.
37. T. Zatari, "Data mining in marketing," *International Journal of Scientific & Engineering Research*, 5(10), 1223-1226, 2014.
38. B.C. Brown, *How to Use the Internet to Advertise, Promote, and Market Your Business or Web Site – With Little or No Money*. Atlantic Publishing Group, Ocala, Florida, 2006.
39. R.C. Blattberg, B. Kim, and S.A. Neslin, *Database Marketing Analyzing and Managing Customers*. Springer, New York, 2008
40. P. Kotler, and G. Armstrong, *Principles of Marketing*. USA: Pearson Prentice Hall, 2011.
41. N. Gladson-Nwokah, and J. Gladson-Nwokah, "Strategic Use of Database Marketing for Marketing Decicion-Making," *Research Journal of Information Technology*, 4(2), 27-37, 2012.
42. D. L. Olson, B. Chae, "Direct marketing decision support through predictive customer response modelling," *Decision Support Systems*, 54, 443-451, 2012. DOI:10.1016/j.dss.2012.06.005.
43. C. Fornell, and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *Journal of marketing research*, 18(1), 39-50, 1981.
44. A. C. Teo, G. W. H. Tan, K. B. Ooi, T. S. Hew, and K. T. Yew, "The effects of convenience and speed in m-payment," *Industrial Management & Data Systems*, 115(2), 311-331, 2015.
45. T. Ramayah, and E. Rahbar, "Greening the environment through recycling: an empirical study," *Management of Environmental Quality: An International Journal*, 24(6), 782-801, 2013.

سال یازدهم، شماره‌های ۳۹و۴۰، بهار و تابستان ۱۳۹۸
صص:۱۰۹ -۱۲۲

پژوهشی فصلنامهٔ علمی-
فناوری اطلاعات و ارتباطات ایران

معیارهای ارزیابی ارزش اثرگذاری کاربران رسانه‌های اجتماعی- چارچوبی براساس کاوش رسانه‌های اجتماعی

*روجیار پیرمحمدیانی **شهریار محمدی
* استادیار، دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه کردستان، ایران
** استادیار، دانشکده مهندسی صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی، ایران
تاریخ دریافت: ۱۳۹۸ /۰۹/۱۶ تاریخ پذیرش: ۱۳۹۹/۰۴/۱۸

چکیده

امروزه رفتارهای تعاملی کاربران در رسانه‌های اجتماعی به یک منبع مهم و اثرگذار بر فعالیت‌های حوزه‌ی بازاریابی تبدیل شده است. مفاهیم موجود در کاوش رسانه‌های اجتماعی، تکنیک‌های لازم برای محاسبه‌ی معیارهای مربوط به ارزیابی رفتارهای اثرگذار کاربران را فراهم می‌آورد. علی‌رغم اهمیت کاوش رسانه‌های اجتماعی برای تحلیل رفتارهای تعاملی کاربران، فقدان یک بازنگری جامع و طرح کلاس‌بندی در این زمینه وجود دارد. به این منظور در این تحقیق در قدم اول با ارائه‌ی یک دسته‌بندی شامل سه حوزه‌ی، تحلیل مبتنی بر کاربر، تحلیل مبتنی بر ارتباط و تحلیل مبتنی بر محتوا، تکنیک‌های کاوش رسانه‌های اجتماعی برای تحلیل رفتارهای اثرگذار کاربران مورد بررسی قرار گرفته است. در ادامه با توجه به مرور ادبیات صورت گرفته، یک چارچوب نوآورانه و ترکیبی شامل دو بعد اصلی" پتانسیل اثرگذاری" و "سطح اثرگذاری" ارائه گردیده است و معیارهای "تعداد کاربران فعال"، "رتبه‌ی کاربر"، "کیفیت و میزان تحلیلی و قضاوتی بودن متون تولید شده توسط کاربران" برای محاسبه‌ی هر یک از این ابعاد تعریف شده است. در واقع این مقاله اولین کلاس‌بندی جامع و آکادمیک در خصوص تکنیک-های کاوش رسانه‌های اجتماعی مثمرثمر در تحلیل رفتارهای کاربران می‌باشد که با ارائه‌ی یک چارچوب امکان ارزیابی ارزش اثرگذاری کاربران را برای کسب‌وکارها فراهم می‌نماید.

واژگان کلیدی: ارزش اثرگذاری، کاوش رسانه‌های اجتماعی، رفتارهای تعاملی کاربران

۱- مقدمه

درسال‌های اخیر با گسترش مفاهیم وب ۲،۰ زیرساخت‌هایی در فضای اینترنت ایجاد شده تا کاربران بتوانند محتوای مورد نظر خود را ایجاد کنند، با یکدیگر به اشتراک گذاشته و به راحتی با دیگر کاربران ارتباط پیدا کنند.از جمله رسانه‌های اجتماعی می‌توان به بلاگ‌ها و ویکی‌ها، سایت‌های شبکه‌های اجتماعی، فروم‌ها و انجمن‌های گفتگو اشاره کرد. به مجموعه‌ی این برنامه‌های کاربردی مبتنی بر اینترنت، رسانه‌های اجتماعی گفته می‌شود [۱۹]. اهمیت استفاده از رسانه‌های اجتماعی در حوزه‌های تجاری به سرعت در حال رشد است. ارزش تجاری و تعداد کاربران این نوع وب‌سایت‌ها بیانگر این اهمیت می‌باشد. این اعتقاد وجود دارد که محتوایی که از طریق کاربران در قالب رتبه‌بندی و ارزیابی محصولات در رسانه‌های اجتماعی ایجاد می‌شود، نقش قابل توجهی بر فعالیت‌های حوزه‌ی بازاریابی کسب وکارهای مختلف ایفا می‌کند [۸].تبلیغات کلامی قبل از ظهور رسانه‌های اجتماعی نیز وجود داشت ولی رشد رسانه‌های اجتماعی این امکان را برای کاربران فراهم کرده است که گستره‌ی

نویسنده عهده‌دار مکاتبات : روجیار محمدیانی pirmohammadiani.r@kntu.ac

ارتباطات خود را از طریق ایجاد شبکه‌های مختلف، تعامل حرفه‌ای و به اشتراک گذاشتن ایده‌ها با دیگران، وسیع‌تر نمایند. این اثرگذاری‌ها موجب پررنگ شدن مفاهیمی تحت عنوان رفتارهای تعاملی۱ کاربران شده است [۱۵].

مفهوم تعامل مشتری، به عنوان رفتار یک مشتری نسبت به یک نام تجاری و یا یک دسته‌بندی از محصول، خدمات و غیره می‌باشد که ازطریق یک کانال ارتباطی به عنوان مثال، رسانه‌های جمعی ایجاد شده است [۱۴]. تعامل در حوزه‌ی رسانه‌های اجتماعی می‌تواند به رفتارهایی مانند تبلیغات کلامی، رفتارهای توصیه‌گر، کمک به سایر کاربران، بلاگ، نوشتن مرورها و نظرات تبدیل شود. این تعاملات، اثرات آگاه‌سازی و یا اطلاع‌رسانی را بر سایر کاربران شامل می‌شود. اثر آگاه‌سازی کاربر را از وجود محصول خبردار می‌کند. اثر اطلاع‌رسانی نیز منجر به تصمیم به خرید محصول در سایر کاربران می‌شود. از این طریق مشتریان می‌توانند با توجه به سطح دانش و تعامل و ارتباطات خود ارزش‌هایی ورای خرید مستقیم کالا برای کسب‌وکارها فراهم می‌کنند [۹]. به همین خاطر برای تحلیل رفتارهای تعاملی کاربران پارامترهای مختلفی مانند ویژگی‌های ساختاری و رابطه‌ای شبکه، ویژگی‌های کیفی و جهت‌گیری محتوایی که بیانگر توانایی کاربر برای تولید و انتشار تبلیغات مثبت و منفی و اثرگذاری بر سایر کاربران می باشد، تعریف شده است[۶]. داده‌های موجود در رسانه‌های اجتماعی حجیم، ناکامل، نویزی، غیر ساختاریافته و شامل ارتباطات اجتماعی می‌باشند. مقیاس و خواص داده‌های رسانه‌های اجتماعی با داده‌های موجود در داده‌کاوی سنتی متفاوت هستند[۲۸]. این امر موجب شده است تحلیل داده‌های رسانه‌های اجتماعی در رابطه با رفتارهای تعاملی کاربران به یکی از چالش‌های اصلی مربوط به کسب‌وکارهای مختلف تبدیل شود. ولی پیشرفت‌های اخیر در علوم کامپیوتر یک زمینه‌ی تحقیقاتی جدید به نام کاوش رسانه‌های اجتماعی توسعه داده است. کاوش رسانه‌های اجتماعی با توسعه مفاهیم داده‌کاوی و استفاده از مفاهیم نظریه‌های اجتماعی، ابزارها و تکنیک‌های محاسباتی لازم به منظور بررسی داده‌های رسانه‌های اجتماعی با مقیاس بزرگ را برای ما فراهم می‌کند[۱۸]. مفاهیم مربوط به کاوش رسانه‌های اجتماعی، برای محاسبه‌ی معیارهای ارزیابی رفتارهای تعاملی کاربران نیز بسیار مثمرثمر خواهد بود. اما علی‌رغم اهمیت کاوش رسانه‌های اجتماعی برای تحلیل رفتارهای تعاملی کاربران، فقدان یک بازنگری جامع و طرح کلاس‌بندی در این زمینه وجود دارد. به همین خاطر هدف اصلی این مقاله، فراهم کردن یک چارچوب و راهنمای کلی برای محاسبه‌ی ارزش تعاملی کاربران با استفاده از روش‌های کاوش رسانه‌های اجتماعی، می‌باشد. برای این منظور ابتدا با ارائه‌ی یک دسته‌بندی، تکنیک‌های کاوش رسانه‌های اجتماعی برای تحلیل رفتارهای تعاملی کاربران در سه حوزه‌ی تحلیل مبتنی بر کاربر، تحلیل مبتنی بر ارتباط و تحلیل مبتنی بر محتوا دسته‌بندی شده است. در ادامه تلاش شده است با توجه به دسته‌بندی انجام شده، معیارهایی برای ارزیابی ارزش اثرگذاری کاربران رسانه‌های اجتماعی ارائه گردد تا به این پرسش‌ها پاسخ داده شود:

ابعاد اصلی موردنیاز برای ارزیابی رفتارهای اثرگذار کاربران چیست؟

چگونه می‌توان اثرات و مزایای ناملموس کاربران برای برند و سازمان که از طریق ارائه‌ی نظرات جامع و مثبت، ارتباطات و اثرگذری‌های مستقیم و غیرمستقیم ایجاد می‌شود را اندازه‌گیری نمود؟

کاوش رسانه‌های اجتماعی چگونه می‌توانند به ما در استخراج این خصیصه‌ها کمک کند؟

در سیستم پیشنهاد شده دو بعد اصلی "پتانسیل اثرگذاری" و "سطح اثرگذاری"، برای ارزیابی ارزش کاربران تعریف شده است. در نهایت یک چارچوب نوآورانه و ترکیبی ارائه و معیارهای تعداد کاربران فعال، رتبه‌ی کاربر، کیفیت و میزان تحلیلی و قضاوتی بودن متون تولید شده توسط کاربر، برای محاسبه‌ی بعدهای ارزش اثرگذاری کاربران رسانه‌های اجتماعی معرفی گردیده است. این مقاله اولین بازنگری جامع و آکادمیک در خصوص تکنیک‌های کاوش رسانه‌های اجتماعی بکار رفته در تحلیل ارزش اثرگذاری مشتریان می‌باشد. و از طریق چارچوب پیشنهادی می‌توان کلیه‌ی اثرات و مزایای ناملموس کاربران برای برند و سازمان که از طریق ارائه‌ی نظرات جامع و مثبت، ارتباطات و اثرگذری‌های مستقیم و غیرمستقیم ایجاد می‌شود را اندازه‌گیری نمود.

1Customer Engagement Behaviour

مقاله‌ی موجود در سه بخش سازماندهی شده است. در بخش ۲ روشی برای دسته‌بندی تکنیک‌های کاوش رسانه‌های اجتماعی برای تحلیل رفتارهای تعاملی کاربران ارائه شده است. در بخش ۳ پژوهش‌هایی که در این زمینه انجام شده و با موضوع تحقیق مرتبط بوده، در قالب این دسته‌بندی مورد بررسی قرار گرفته است. در بخش ۴ چارچوبی برای تحلیل رفتارهای تعاملی کاربران ارائه شده است و معیارهای کمی و روش‌های کاوش رسانه‌های اجتماعی مثمرثمر برای محاسبه‌ی هر یک از ابعاد ارزش اثرگذاری کاربران رسانه‌های اجتماعی بیان شده است. در بخش ۵ نتیجه‌گیری و پیشنهاداتی برای کارهای آتی ارائه گردیده است.

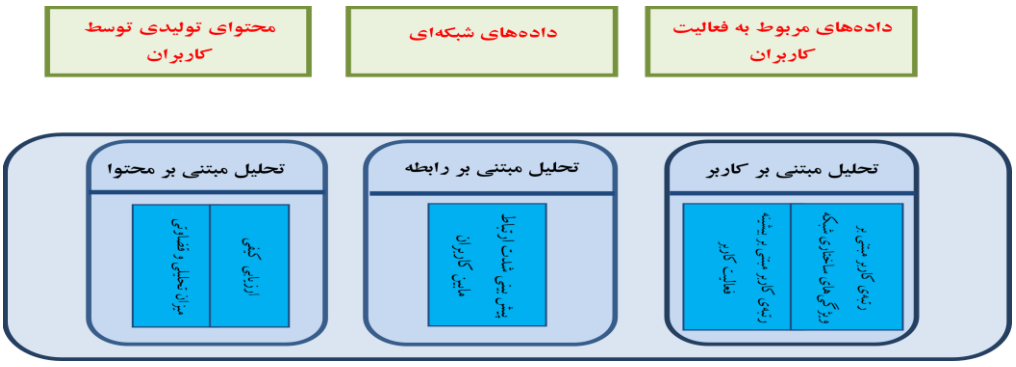
۲- مبانی پژوهش

برای ایجاد نگرش جامع و درک کامل رفتارهای تعاملی کاربران، بایستی کلیه‌ی داده‌های جدا از هم مرتبط با کاربران را در نظر گرفت. در رسانه‌های اجتماعی کاربران محتوای خود را تولید می‌کنند. این داده‌ها تحت عنوان Content User Generatedشناخته می‌شود[۲۳]. سایر کاربران می‌توانند به ارزیابی این محتوا بپردازند و آن را از جنبه‌ی مفید بودن مورد بررسی قرار دهند یا آن را در داخل شبکه به اشتراک بگذارند. مکانیزم رتبه‌بندی به کاربران امکان می‌دهد تا ارزیابی و رضایت خود را از تعاملات بین کاربران بیان کنند. سیستم‌های رتبه‌بندی و فیدبک‌دهی از جمله مهمترین منابع داده‌های زمینه‌ای[۲] در وب اجتماعی هستند. همچنین این وب‌سایت‌ها امکاناتی را فراهم آورده‌اند تا کاربران لیست افراد قابل اعتماد و یا دوستان خود را بیان کنند. ارتباطات میان کاربران در اجتماعات آنلاین باعث شکل‌گیری شبکه بین کاربران آن اجتماع شده است. در ادبیات مربوط به شبکه‌های اجتماعی این داده‌ها تحت عنوان داده‌های ساختاری[۳] نیز شناخته می‌شوند[۳۲] .

داده‌های موجود در رسانه‌های اجتماعی در سه گروه محتوای تولیدی توسط کاربران و داده‌های شبکه‌ای و داده‌های مربوط به فعالیت کاربران وجود دارد. بر اساس داده‌های موجود در این وب‌سایت‌ها، روش‌های کاوش رسانه‌های اجتماعی در رابطه با ارزیابی اثرگذاری کاربران را به سه گروه تحلیل مبتنی بر کاربر، تحلیل مبتنی بر ارتباط و تحلیل مبتنی بر محتوا دسته‌بندی شده است [۱۸]. در شکل ۱ این دسته‌بندی کلی و وظایف مربوط به هر دسته بیان شده است. در ادامه به توضیح وظایف مربوط به هر یک از این گروه‌ها می‌پردازیم و مقالات مختلفی که از این روش‌ها استفاده کرده‌اند، مورد بررسی قرار می‌گیرد.

۲-۱-تحلیل مبتنی بر کاربر

تحلیل‌های مبتنی بر کاربر روش‌های مؤثر و ضروری برای تحلیل رفتار کاربر در کل سیستم را فراهم می‌آورد.یکی از این وظایف، اندازه‌گیری رتبه‌ی کاربر می‌باشد. رتبه‌ی کاربر با توجه به داده‌های ساختاری و جایگاه قرارگیری فرد در شبکه یا پیشینه‌ی فعالیت هر کاربر قابل محاسبه است.



شکل۱- دسته‌بندی روش‌های کاوش رسانه‌های اجتماعی

2.Contextual data

3Structural data

۲-۱-۱- رتبه‌ی کاربر مبتنی بر ویژگی‌های ساختاری شبکه

ویژگی‌های ساختاری از جمله معیارهای مهم برای استنتاج موقعیت کابران در داخل رسانه‌های اجتماعی می‌باشد. ویژگی‌های ساختاری با توجه به مشخصات و توپولوژی شبکه و جایگاه قرارگیری فرد در شبکه اهمیت و محبوبیت فرد در ارتباط با افراد دیگر را بیان می‌کند. ویژگی‌های ساختاری استفاده شده برای محاسبه رتبه‌ی فرد، معیارهای قدرت و مرکزیت می‌باشد [۲۴].

در ساختارهای مبتنی بر گراف، معیارهای "درجه[۴]"، "نزدیکی[۵]"، "بینابینی[۶]" و "رتبه‌بندی صفحه[۷]" از روش های اصلی در محاسبه رتبه گره به لحاظ محبوبیت هستند. در معیار "درجه"، اهمیت هر کاربر با توجه به تعداد همسایه‌های آن کاربر در شبکه محاسبه می‌شود. معیار "بینابینی"، برای هر کاربر در واقع بدین معنی می‌باشد که مسیر بین هر دو فرد دیگر، لزوماً از این فرد خاص بگذرد. تعداد این مسیرها برای هر کاربر، بیانگر اهمیت آن کاربر می‌باشد. در معیار "نزدیکی" گره‌ای که مجموع فواصلش از سایر گره‌ها در مقایسه با بقیه گره‌ها کمتر باشد از اهمیت بیشتری برخوردار است. مبتنی بر معیار "رتبه صفحه"، گره‌ای مهم و معتبر است که از گره‌های مهم و معتبر دیگر، به آن لینک شده باشد [۳۰].

در شبکه‌ی لحاظ شده برای رسانه‌های اجتماعی، این فاکتورها بیانگر این مطلب است که هر چقدر افراد بیشتری با کاربر هدف در ارتباط باشند و فرد بتواند سریع‌تر با سایر افراد در شبکه مراوده داشته باشد، شهرت فرد در شبکه بیشتر بوده و انتظار می‌رود که فرد قابلیت اثرگذاری بیشتری داشته باشد.

۲-۱-۲-رتبه‌ی کاربر مبتنی بر پیشینه فعالیت کاربر

با توجه به فعالیت‌ها و رفتارهای کاربر در داخل رسانه‌های اجتماعی شامل پست‌ها، پاسخ‌ها، اشتراک‌گذاری‌ها و غیره محاسبه می‌شوند. برای محاسبه‌ی رتبه‌ی کاربر مبتنی بر فعالیت دو معیار "سطح فعالیت[۸]" و "میزان تاثیرگذاری فعالیت[۹]" تعریف شده است [۳]. در "سطح فعالیت" سطح تمایل کاربر برای به اشتراک‌گذاری و نوشتن نظرات، در نظر گرفته می‌شود. بیشتر بودن تعداد نظرات نوشته شده و پست‌های ارسال شده‌ی یک کاربردر یک دوره‌ی زمانی خاص بیانگر این مطلب می‌باشد که کاربر اصطلاحا talkative می‌باشد و احتمال پخش اطلاعات از سمت آن‌ها بیشتر است[۵]. "سطح اثرگداری فعالیت کاربر"، میزان محبوبیت محتوای تولید شده توسط کاربر را در نظر می‌گیرد. هر پاسخ در قالب لایک، نوشتن نظر و به اشتراک‌گذاری به یک پیام یا سند تولید شده از طرف یک کاربر مانند یک رای می‌باشد و این دلالت بر اثرگذاری آن پیام می‌باشد. با توجه به تعداد امتیازاتی که برای یک پیام دریافت شده است میزان محبوبیت پیام محاسبه می‌گردد [۱۰]. با توجه به این معیارها کاربرانی که به طور پیوسته فعالیت داشته‌اند رتبه‌ی بالاتری دارند و تفاوت‌های قابل توجهی درمیزان اثرگذاری کاربران با توجه به میزان فعالیت انجام شده از طرف آن‌ها وجود دارد.

۲-۲- تحلیل مبتنی بر ارتباط

تحلیل‌های مبتنی بر ارتباط روش‌های مؤثر و ضروری برای تحلیل ارتباطات کاربران را فراهم می‌کند. وظیفه‌ی مهم تحلیل مبتنی بر ارتباط در حوزه‌ی تحلیل رفتارهای تعاملی کاربران عبارتند از: پیش‌بینی شدت ارتباط ما بین کاربران مبتنی بر شبکه‌ی ارتباطات موجود. وجود لینک و ارتباط بین کاربران تنها دلیلی برای اطمینان از دیده شدن پیام‌ها و تبلیغات کلامی آن‌ها توسط یکدیگر می‌باشد و میزان اثرگذاری کاربران به قدرت ارتباط بستگی دارد [۲۷]. بنابراین وظیفه‌ی اصلی برای تحلیل‌های مبتنی بر رابطه پیش‌بینی قدرت اتصال می‌باشد. با توجه به داده‌های ساختاری و زمینه‌ای سه گروه از معیارها برای اندازه‌گیری شدت ارتباط بین کاربران در نظر گرفته شده است: معیارهای مبتنی بر شباهت، معیارهای مبتنی بر تراکنش و معیارهای مبتنی بر ارتباط [۱۸].

۲-۲-۱- معیارهای مبتنی بر شباهت

تئوری هموفیلی در ارتباطات اجتماعی بیان می‌کند که کاربران مشابه به احتمال زیاد روابط اجتماعی برقرار می‌کنند. هر چقدر میزان شباهت بیشتر باشد شدت ارتباط نیز بیشتر خواهد بود. این شباهت می‌تواند از جنبه‌ی زمینه‌ای یا ساختاری مطرح شود. در محاسبه‌ی شباهت زمینه‌ای، هدف بررسی و محاسبه میزان اهمیت ویژگی‌های مشترک بین کاربران نظیر ترجیحات و

[4]Degree
[5]Closeness
[6]betweness
[7]Page Rank
[8]Activity Level
[9]Activity Impact

روش‌های مشترک در قضاوت دیگران بر شکل‌گیری و شدت ارتباط ما بین کاربران است. در این حالت زوج کاربر مورد نظر شبیه تلقی می‌شوند اگر نحوه‌ی ارزیابی دو کاربر بر روی مجموعه موضوعات مشترک شامل کالاهای مشترک، اطلاعات و یا سرویس‌های مشترک، مشابه باشد [۳۲].

در محاسبه‌ی شباهت ساختاری هدف بررسی و محاسبه میزان اهمیت ویژگی‌های مشترک بین کاربران نظیر ویژگی‌های شبکه‌ای برشکل‌گیری و شدت ارتباط ما بین کاربران است. شباهت ساختاری به شیوه‌ای از محاسبه شباهت بین زوج کاربران اشاره می‌کند که مبتنی بر توپولوژی مشابه در شبکه تعریف می‌شود. در این حالت زوج کاربر مورد نظر شبیه تلقی می‌شوند اگر با مجموعه یکسانی از افراد در شبکه ارتباط داشته باشند [۲۷].

۲-۲-۲. معیارهای مبتنی بر ارتباط

در معیارهای مبتنی بر ارتباط، موقعیت کاربران نسبت به یکدیگر مبتنی بر روابط موجود در شبکه، بر ایجاد اثرگذاری موثر است. به این صورت که به هر میزان که دو کاربر در ساختار شبکه نسبت به یکدیگر دسترس پذیرتر باشند، امکان تشکیل یال ارتباطی میان آن دو مبتنی بر خاصیت تعدی بیشتر می‌شود. واضح است که در این حالت هر چقدر مسیر استنتاجی کوتاهتر باشد، امکان تشکیل یال قویتر و میزان اثرگذاری بیشتر است [۳۲].

۲-۲-۳- معیار مبتنی بر تراکنش

در معیار مبتنی بر تراکنش شدت ارتباط بین دو کاربر از طریق میزان آشنایی دو کاربر با توجه به تعداد تعاملات ما بین آن‌ها، سنجیده می‌شود. در برنامه‌های کاربردی وب اجتماعی، این فاکتور به راحتی از داده‌های زمینه‌ای موجود در سیستم بویژه داده‌های فیدبک و رتبه‌بندی قابل محاسبه است. انتظار می‌رود که فیدبک‌ها و میانگین امتیازات داده شده بیانگر نگرش و میزان رضایت کاربران نسبت به یکدیگر باشد. واضح است که هر چقدر میزان رضایت به ازای تعداد تعامل بیشتری محاسبه شده باشد، سطح رضایت و شدت ارتباط بیش‌تر خواهد بود [۲۷].

هر چه روابط بین کاربران قوی‌تر باشد احتمال انتشار اطلاعات در داخل شبکه نیز بیشتر خواهد شد. در تعداد زیادی از مقالات خصیصه‌های مربوط به کاربران اثرگذار از طریق اندازه‌گیری تاثیرات آن‌ها بر پدیده‌ی انتشار مورد بررسی قرار گرفته است. مسئله‌ی بیشینه‌سازی انتشار در رسانه‌های اجتماعی به بیان این موضوع می‌پردازد که چگونه یک زیرمجموعه‌ی K راسی از یک گراف داده شده را انتخاب کرده و فعال سازیم به طوری که بیشترین تعداد نود را در داخل شبکه تحت تاثیر قرار دهیم. مدل‌های ارائه شده برای مسئله‌ی انتشار را می‌توان به دو دسته‌ی آستانه ای و آبشاری تقسیم کرد [۲۴]. همچنین بایستی با توجه به ساختار شبکه‌ای و داده‌های زمینه‌ای، اثرات بعدی شبکه در محاسبه‌ی ارزش و تحلیل کاربران مد نظر قرار بگیرد چرا که این احتمال وجود دارد که اکنون لینکی بین کاربر وجود نداشته باشد ولی احتمال شکل‌گیری این لینک در آینده وجود دارد.

۲-۳. تحلیل مبتنی بر محتوا

یک بخش مهم دیگر برای ارزیابی کاربران علاوه بر ویژگی‌های ساختاری مبتنی بر شبکه و فعالیت‌های کاربر، نظرات و متون تولید شده توسط کاربران می‌باشد. تحلیل‌های مبتنی بر محتوا، روش‌های مؤثر و ضروری برای تحلیل محتوای تولید شده توسط مشتریان را فراهم می‌آورد. کاربران مختلف دارای تخصص و ترجیحات مختلفی در مورد محصولات مختلف هستند. این ترجیحات و تخصص‌ها از طریق نظراتی که نوشته می‌شود آشکارمی‌شود و در میزان اثرگذاری کاربر بر سایر کاربران نقش مهمی را ایفا می‌کند [۱۰]. دو وظیفه‌ی مهم تحلیل مبتنی بر محتوا در حوزه‌ی تحلیل رفتارهای تعاملی کاربران عبارتند از: ارزیابی کیفی و میزان تحلیلی و قضاوتی بودن[10].

۲-۳-۱- ارزیابی کیفی

ارزیابی کیفی میزان مفید بودن محتوای تولید شده توسط کاربران را مورد بررسی قرار می دهد. از جمله معیارهای مفید بودن می‌توان به موارد زیر اشاره کرد:

- ✓ بسیاری از جنبه‌های آن محصول یا خدمت را پوشش دهد. به عبارتی جامع باشد.
- ✓ انعکاس دهنده‌ی تجربیات شخصی باشد صرف اینکه بازتابی از تجربیات دیگران باشد.
- ✓ از لحاظ ساختار گرامری و نوشتاری قابل قبول باشد.
- ✓ بازتاب خوبی از سایر کاربران در داخل شبکه نسبت به این نظر ارائه شده باشد.

[10] subjective degree

این ویژگی‌ها از طریق خصیصه‌های متنی استخراج شده، قابل محاسبه است [۳۱]. کیفیت متون تولید شده توسط کاربران بر میزان اثرگذاری آن‌ها موثر است. چرا که هر چقدر که کیفیت متون تولید شده توسط یک کاربر در متن‌هایی که از خود به جای می‌گذارد، بیشتر باشد، احتمال این که وی در آن حوزه فرد اثرگذارتری باشد، بیشتر است [۱۰].

۲-۳-۲- میزان تحلیلی و قضاوتی بودن

میزان تحلیلی و قضاوتی بودن بیانگر توانایی کاربر برای تولید و انتشار تبلیغات مثبت و منفی می‌باشد. این ارزیابی از طریق فراوانی کلمات حاوی احساس به کار برده شده در یک متن تعیین می‌گردد. برای مثال کلمات خوب، شگفت‌انگیز، و متحیرکننده کلمات حاوی احساس مثبت هستند و کلماتی مانند بد، ضعیف جز کلمات حاوی احساسات منفی هستند. از نقطه نظر مصرف‌کنندگان، یک نظر تاثیرگذار نه تنها مزایا و منافع مربوط به یک کالا را بیان می‌کند بلکه به معایب و نقص‌های محصول نیز اشاره دارد. بنابراین اگر نظری مزایا و معایب مربوط به یک کالا را بیان کند، حاوی کلمات احساسی مثبت و منفی می‌باشد و بیشتر قابل اعتماد است. به همین خاطر میزان تحلیلی و قضاوتی بودن برای استنتاج در مورد اثرگذاری کاربران مثمرثمر می‌باشد[۱۰].

۳- پیشینه‌ی تحقیق با توجه به دسته‌بندی ارائه شده برای روش های کاوش اجتماعی

کارهای زیادی هم در این حوزه انجام شده است. در این کارها سعی شده است با لحاظ کردن ویژگی‌های خاص مربوط به داده‌ها و محیط رسانه‌های اجتماعی، معیارهای مناسبی برای تحلیل و اندازه‌گیری رفتارهای تعاملی کاربران ارائه شود. در جدول ۱ بیان شده است که هر کدام از این مقالات بر روی چه مفاهیمی تمرکز داشته‌اند. مرور مطالعات انجام شده در حوزه‌ی رسانه‌های اجتماعی نشان می‌دهد که پژوهش‌های زیادی در زمینه‌ی ارزیابی اثرگذاری کاربران، جهت تعیین افراد بانفوذ در داخل رسانه‌های اجتماعی انجام گرفته است.

جدول ۱-روش‌های کاوش رسانه‌های اجتماعی استفاده شده در مقالات انتخاب شده‌ی اخیر

مرجع	حوزه‌ی مربوط به کاوش رسانه‌های اجتماعی	وظایف مربوط به کاوش رسانه‌ها اجتماعی	معیارها	گراف	نوع رسانه‌ی اجتماعی
گلدنبرگ و همکاران(۲۰۰۹) [۱۳]	تحلیل مبتنی بر کاربر	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه	درجه‌ی ورودی و خروجی	گراف بدون جهت لینک‌های مبتنی بر فعالیت	سی ورلد
هیدمن و همکاران(۲۰۱۰)، [۱۶]	تحلیل مبتنی بر کاربر	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه	رتبه‌بندی صحفه	گراف وزن‌دار بر اساس لینک-های فعالیت	فیس بوک
هینز و همکاران [۲۶]، (۲۰۱۱)	تحلیل مبتنی بر کاربر	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه	درجه‌ی ورودی و خروجی و بینابینیی	گراف مبتنی بر فعالیت‌های اجتماعی	شبکه‌ی ارتباطات مخابراتی
الیاس و همکاران [۱۷], (۲۰۱۱)	تحلیل مبتنی بر کاربر	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه	پیشنهاد مرکزیت مؤلفه ی اصلی	گراف غیرجهت دار مبتنی بر لینک های اجتماعی	اورکات و فیس بوک
کیستاک و همکاران [۴]، (۲۰۱۰)	تحلیل مبتنی بر کاربر	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه	توسعه متدی به نام K-shell		منابع داده ای متعدد مانند شبکه ایمیل، بازیگران و غیره
کیس و بیچلر (۲۰۰۸)، [۲۱]	تحلیل مبتنی بر کاربر	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه	دو معیار، درجه ی خروجی و معیار توسعه داده		داده‌های تماس های مشتریان یک شبکه‌ی مخابراتی

			شده‌ی		
لرمن و گوش(۲۰۱۰)، [۷]	تحلیل مبتنی بر کاربر	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه	معیار αمرکزیت (نرمالیزه شده)	گراف(وزن دار) جهت دار مبتنی بر لینک های اجتماعی	دیگ
مانگر و ژاو (۲۰۱۵)، [۲۲]	تحلیل مبتنی بر کاربر و محتوا	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه، ارزیابی کیفی و میزان قضاوتی بودن	رتبه بندی صحفه	گراف وزن دار مبتنی بر فعالیت	داده‌های جمع‌آوری شده از تالار گفتگوی تخصصی سیسکو
لی و همکاران(۲۰۱۱)، [۱۱]	تحلیل مبتنی بر کاربر و محتوا	رتبه‌بندی کاربران مبتنی بر ویژگی‌های ساختاری شبکه و بیشینه ی فعالیت، ارزیابی کیفی و میزان قضاوتی بودن			داده‌های جمع‌آوری شده از یک بلاگ
لی و همکاران(۲۰۱۰)، [۱۲]	تحلیل مبتنی بر کاربر و محتوا	رتبه‌بندی کاربران مبتنی بر پیشینه ی فعالیت ارزیابی کیفی و میزان قضاوتی بودن			اپینیونز
لی و همکاران (۲۰۱۳)، [۱۰]	تحلیل مبتنی بر کاربر و محتوا	رتبه‌بندی کاربران مبتنی بر بیشینه‌ی فعالیت و ارزیابی کیفی			داده‌های جمع‌آوری شده از یک بلاگ
راماکیرتینان و لیوینگستون(۲۰۱۶)،] ۲۵[	تحلیل مبتنی بر کاربر و محتوا	رتبه‌بندی کاربران مبتنی بر ویژگی های ساختاری و میزان قضاوتی بودن			داده‌های جمع‌آوری شده از بلاگ
کومار و همکاران [۵] ,(۲۰۱۳)	تحلیل مبتنی بر کاربر و رابطه	رتبه‌بندی کاربران مبتنی بر بیشینه ی فعالیت و میزان اثرگذاری با توجه به شدت ارتباط ما بین کاربران			داده‌های جمع‌آوری شده از شبکه های اجتماعی
یونگ و تاران(۲۰۱۳)، [۲۷]	تحلیل مبتنی بر رابطه	رتبه‌بندی کاربران مبتنی بر میزان اثرگذاری با توجه به شدت ارتباط ما بین کاربران		گراف جهت‌دار مبتنی بر رابطه‌ی اعتماد	اپینیونز
آگاروال و همکاران(۲۰۱۱)، [۱]	تحلیل مبتنی بر کاربر و محتوا	رتبه‌بندی کاربران مبتنی بر پیشینه ی فعالیت و ارزیابی کیفی			داده های مربوط به رسانه های اجتماعی مختلف
کاتسیمپراس و همکاران (۲۰۱۵)، [۲۰]	تحلیل مبتنی بر کاربر و محتوا	رتبه‌بندی کاربران مبتنی بر ویژگی های ساختاری شبکه و متون		گراف با توجه به مشابهت کاربران در محتوای تولید شده	تویتر
سایتو و همکاران(۲۰۱۲)، [۲۶]	تحلیل مبتنی بر رابطه	حل مسئله‌ی بیشینه‌سازی انتشار با لحاظ کردن شدت ارتباط بین کاربران		گراف جهت دار مبتنی بررابطه‌ی کاربران	بلاگ

ژانگ و همکاران(۲۰۱۰)، [۲۹]	تحلیل مبتنی بر رابطه	حل مسئله‌ی بیشینه‌سازی انتشار با لحاظ کردن شدت ارتباط بین کاربر ان مبتنی بر شباهت	گراف وزن دار مبتنی بر داده های اعتماد و امتیازبندی	اپینیونز
ژانگ و همکاران(۲۰۱۱)، [۳۰]	تحلیل مبتنی بر رابطه	حل مسئله‌ی بیشینه‌سازی انتشار با لحاظ کردن شدت ارتباط بین کاربران مبتنی بر میزان علاقه مندی کاربران در یک حوزه		شبکه‌ی نویسندگان همکار
ایریناکی و همکاران(۲۰۱۲)، [۳]	تحلیل مبتنی بر کاربر	رتبه‌بندی کاربران مبتنی بر ویژگی‌های بیشینه‌ی فعالیت و خصیصه‌های شبکه‌ای	رتبه‌بندی صحفه بینابینیو سطح فعالیت	مای اسپیس و داده-های ساختگی
بی و همکاران (۲۰۱۴)، [۲]	تحلیل مبتنی بر کاربر و محتوا	رتبه‌بندی کاربران مبتنی بر ویژگی های ساختاری شبکه و ارزیابی کیفی متون	گراف با توجه به مشابهت کاربران در محتوای تولید شده	تویتر

در این مقالات این مقوله که رهبران عقیده یا افراد با نفوذ چه خصیصه‌های بارزی دارند و محرک‌های موثر بر اثرگذاری آن‌ها مورد بررسی قرار گرفته است. هر کدام از پژوهش‌های انجام شده، معیارهای مجزایی را برای انتخاب افراد با نفوذ در نظر می‌گیرند. نکته‌ی قابل توجه در تحقیقات گذشته، عدم تمرکز بر فاکتورهای موثر بر ارزیابی ارزش اثرگذاری کاربران در ساخت ویژگی‌هاست، بگونه‌ای که فاکتورهای پیش‌بینی تنها مبتنی بر منابع داده‌ای موجود در سیستم استخراج شده‌اند و لذا عموما وابسته به مجموعه داده مورد نظر هستند، در حالی که در این تحقیق، سعی شده است از طریق ارائه‌ی یک چارچوب کلی فاکتورهای اصلی در ارزیابی ارزش اثرگذاری تعیین گردد. سپس این فاکتورهای کیفی مبتنی بر منابع داده‌ای به صورت سیستماتیک به مجموعه از ویژگی‌های کمی نگاشت شده‌اند که قابلیت تعمیم به شرایط مشابه را دارند.

۴- چارچوب کلی مدل پیشنهادی برای ارزیابی ارزش اثرگذاری کاربران رسانه‌های اجتماعی

همان‌طور که اشاره شد تعامل مشتریان در رسانه‌های اجتماعی می‌تواند بر ارزش مشتری در این رسانه‌ها برای کسب وکارها اثرگذار باشد. ارزش اثرگذاری مشتری[11]، میزان تاثیرگذاری یک مشتری بر دیگران را در رسانه‌های اجتماعی مورد ارزیابی قرار می‌دهد. چندین عامل می‌تواند میزان این اثرگذاری را تحت تأثیر قرار دهد از جمله تعداد اتصالات یا ارتباطات، جایگاه و سطح فعالیت مشتریان در رسانه های اجتماعی، مقدار و کیفیت محتوای تولید شده توسط آن‌ها. مثلاً کاربر با اتصالات زیاد می‌تواند میزان اثرگذاری بیشتری نسبت به یک کاربر با اتصالات پایین داشته باشد [۵].

با توجه به مرور ادبیات انجام شده چارچوب پیشنهادی در این مقاله به طور خاص برای ارزیابی ارزش اثرگذاری کاربران رسانه‌های اجتماعی دو بعد را پیشنهاد داده است که می‌تواند برای پشتیبانی از تصمیمات بازاریابی مورد استفاده قرار گیرد: "پتانسیل اثرگذاری" و "سطح اثرگذاری". هر یک از این بعدهای تعریف شده از طریق پاسخ‌گویی به سوالات بیان شده در شکل ۲ مورد بررسی قرارمی‌گیرد.

با توجه به دسته‌بندی که در مورد وظایف مربوط به کاوش رسانه‌های اجتماعی انجام دادیم، معیارهای لحاظ شده در بخش تحلیل مبتنی بر کاربر و تحلیل مبتنی بر محتوا می-توانند برای ارزیابی پتانسیل اثرگذاری و پاسخ‌گویی به سوال اول مثمرثمر باشند. این معیارها با توجه به رفتار کاربر در کل سیستم یک مقدار واحد به کاربر نسبت می‌دهند. از این خصیصه‌ها می‌توان برای ارزیابی میزان قابلیت کاربران برای تبدیل

[11] Customer Influence Value (CIV)

شدن به رهبر عقیده استفاده کرد. درتحلیل‌های مبتنی بر کاربر هر چه مقدار معیارهای قدرت و مرکزیت یا معیارهای مبتنی بر فعالیت کاربر بیشتر باشد، این افراد در مرکز تعاملات قرار داشته، به خوبی متصل بوده و با اکثریت افراد در شبکه ارتباط دارند. از این رو قابلیت کاربر برای تبدیل شدن به رهبر عقیده بیشتر خواهد بود. در تحلیل مبتی بر محتوا هم هر چقدرسطح کیفیت و میزان تحلیلی و قضاوتی بودن متون تولید شده توسط یک کاربر در متن‌هایی که از خود به جای می‌گذارد، بیشتر باشد، احتمال این که وی در آن حوزه فرد متخصص و قابل اعتمادتری باشد بیش‌تر است در نتیجه میزان اثرگذاری آن‌ها نیز بیشتر خواهد بود.

معیارهای لحاظ شده در بخش تحلیل‌های مبتنی بر ارتباط در محاسبه‌ی درصد و نسبت افرادی که در داخل شبکه تحت تاثیر کاربر مورد نظر قرار گرفته‌اند، نقش دارد. پیشنهادی برای محاسبه‌ی تعداد کاربرانی که تحت تاثیر قرار می‌گیرند مفهوم

پتانسیل اثرگذاری

آیا این کاربر از پتانسیلی برای اثرگذاری برخوردار است؟

با لحاظ کردن ویژگی‌های مختلف، قابلیت و توانایی‌های کاربر را برای اثرگذاری اندازه گیری می‌شود

سطح اثرگذاری

میزان این اثرگذاری تا چه اندازه می‌باشد؟

از طریق لینک‌های ارتباطی، درصد و نسبت افرادی که در داخل شبکه تحت تاثیر کاربر مورد نظر قرار گرفته‌اند اندازه‌گیری می‌شود.

شکل۲- ابعاد اصلی ارزش اثرگذاری کاربران رسانه‌های اجتماعی

تعداد کاربران فعال[12] تعریف شده است. برای محاسبه‌ی این پارامتر یک آستانه در نظر گرفته شده است که اگر تاثیری که کاربر a بر کاربر b ایجاد می‌کند از این آستانه بیشتر شود، فرد b به عنوان یکی از افرادی که توسط کاربر a فعال شده است در نظر گرفته می‌شود. این اثرگذاری‌ها می‌تواند فرد را از مصرف‌کننده‌ی کالای یک شرکت به مصرف‌کننده‌ی کالای شرکت دیگر تبدیل کند یا باعث شود فرد کالایی را خریداری کند، در حالیکه تا کنون تمایلی به خرید آن کالا نداشته است. خصیصه‌های مبتنی بر شباهت، خصیصه‌های مبتنی بر دانش و خصیصه‌های مبتنی بر ارتباط که در بخش مرور ادبیات برای اندازه‌گیری شدت ارتباط بین کاربران در نظر گرفته شده است بیانگر میزان اثرگذاری مابین کاربران می‌باشد و در محاسبه‌ی تعداد کاربران فعال مورد استفاده قرار می‌گیرد. در جدول ۲ معیارهای کمی و روش-های کاوش رسانه‌های اجتماعی مثمرثمر برای محاسبه‌ی هر یک از ابعاد ارزش اثرگذاری کاربران رسانه‌های اجتماعی بیان شده است.

سیستم ارزیابی پیشنهاد شده با لحاظ کردن دو بعد اصلی پتانسیل اثرگذاری و میزان اثرگذاری، سعی در تعریف ارزش اثرگذاری کاربران و درک ماهیت اثرگذاری‌ها را دارد. معیارهای تعداد کاربران فعال، رتبه‌ی کاربر، کیفیت و میزان تحلیلی و قضاوتی بودن متون تولید شده توسط کاربر، برای محاسبه‌ی بعدهای ارزش اثرگذاری کاربران رسانه‌های اجتماعی معرفی گردیده است. معیارهای تعریف شده در قالب دسته بندی ارائه شده در شکل ۳ نمایش شده است.

در این بررسی‌ها سعی شده است با ارائه‌ی یک چارچوب کلی فاکتورهای اصلی در ارزیابی ارزش اثرگذاری تعیین گردد. سپس این فاکتورهای کیفی مبتنی بر منابع داده‌ای به صورت سیستماتیک به مجموعه از ویژگی‌های کمی نگاشت شده‌اند. از طریق چارچوب و راهنمای کلی، کسب و کارها می‌توانند با توجه به داده‌های در دسترس معیارهای مناسب برای ارزیابی ارزش اثرگذاری کاربران را در نظر بگیرند.

در ادامه در قالب یک مثال تفاوت ارزش اثرگذاری کاربران رسانه‌های اجتماعی با توجه به معیارها و ابعاد پیشنهاد شده مورد بررسی قرار گرفته است. به عنوان مثال فرض کنید مقادیر مربوط به هر کدام از ابعاد "پتانسیل اثرگذاری" و "سطح اثرگذاری" دارای دو سطح "کم[13](L)" و "زیاد[14](H)" باشند. در این حالت می‌توان

[12] Activated user

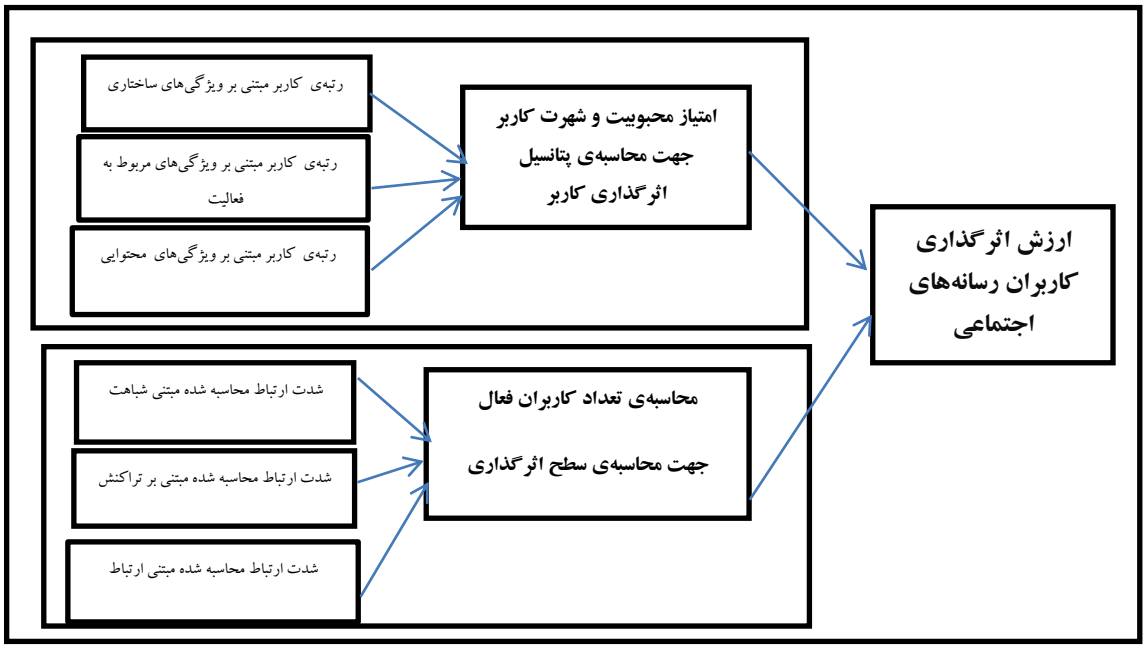
[13] Low

[14] High

این گونه استدلال کرد که ارزش اثرگذاری مشتریان با پتانسیل اثرگذاری و سطح اثرگذاری بالاتر؛ کاربر با طبقه‌ی (H,H)؛ بیش‌تر از کاربران با یک بعد L و یک بعد H می‌باشد. همچنین ارزش اثرگذاری کاربران با یک بعد L و یک بعد H بیشتر از کاربران با طبقه‌ی (L,L) می‌باشد. در این حالت می‌توان با توجه به طبقه‌ی کاربر،ارزش اثرگذاری کاربر در آن رسانه را به یکی از موارد پایین (L)، متوسط (M)، یا بالا (H) نسبت داد.

جدول۲- معیارهای لحاظ شده برای ارزیابی هر یک از ابعاد ارزش اثرگذاری کاربران رسانه‌های اجتماعی

معیارهای کمی	روش‌های کاوش رسانه-های اجتماعی	ابعاد اصلی ارزش اثرگذاری
رتبه‌ی کاربر مبتنی بر ویژگی‌های ساختاری	تحلیل مبتنی بر کاربر	پتانسیل اثرگذاری کاربر
رتبه‌ی کاربر مبتنی بر بیشینه‌ی فعالیت		
ارزیابی کیفی متون تولید شده توسط کاربر	تحلیل مبتنی بر محتوا	
میزان قضاوتی بودن متون تولید شده توسط کاربر		
محاسبه‌ی تعداد کاربران فعال با توجه به شدت ارتباط محاسبه شده مبتنی بر شباهت	تحلیل مبتنی بر رابطه	سطح اثرگذاری
محاسبه‌ی تعداد کاربران فعال با توجه به شدت ارتباط محاسبه شده مبتنی بر ارتباط		
محاسبه‌ی تعداد کاربران فعال با توجه به شدت ارتباط محاسبه شده مبتنی بر دانش		



شکل ۳- چارچوب پیشنهادی برای ارزیابی ارزش اثرگذاری کاربران

جدول ۳-مقادیر SM_{ij} برای هر یک از کاربران رسانه های اجتماعی

	رسانه ی اجتماعی			
	فیس بوک	توییتر	بلاگ	جوامع
کاربرA	(H,L)	(L,L)	(L,H)	(H,H)
کاربر B	(H,H)	(H,L)	(L,L)	(H,L)
کاربرC	(L,L)	(H,L)	(H,H)	(H,H)

به عنوان مثال فرض کنید که مقادیر SM_{ij} بیان‌کننده‌ی ارزش اثرگذاری مشتریان در رسانه‌های اجتماعی است، $i=1....I$ نشان‌دهنده‌ی انواع رسانه‌های اجتماعی است به عنوان مثال فیس‌بوک، توییتر، بلاگ، جوامع وغیره می‌باشد و $J=1$ بیان‌کننده‌ی سطح تعامل و $j=2$ سطح اثرگذاری کاربران می‌باشد. با توجه به این مقادیر کاربر در یکی از سه رده‌ی بالا، پایین یا متوسط طبقه‌بندی می‌شود. در جدول ۳، مقادیر SM_{ij} برای ۳ کاربر در رسانه‌های اجتماعی مختلف بیان شده است.

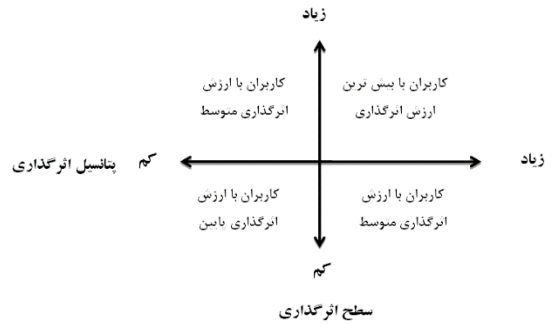
با داشتن تمامی موارد بیان شده مشتریان A,B,C دارای ارزش اثرگذاری زیر در رسانه‌های اجتماعی مربوطه می‌باشند:

جدول ۴- ارزش اثرگذاری کاربران در رسانه‌های اجتماعی

	فیس بوک	توییتر	بلاگ	جوامع
کاربرA	M	L	M	**H**
کاربر B	H	M	L	**M**
کاربر C	L	M	M	**H**

در واقع هدف از این مقاله فراهم کردن یک چارچوب و راهنمای کلی برای محاسبه‌ی ارزش تعاملی کاربران در پلت فرم رسانه‌های اجتماعی، جهت راهنمایی کسب - وکارها می‌باشد. بدین صورت می‌توان کلیه‌ی اثرات و مزایای ناملموس کاربران برای برند و سازمان که از طریق ارائه‌ی نظرات جامع و مثبت، ارتباطات و اثرگذری‌های مستقیم و غیرمستقیم ایجاد می‌شود را اندازه‌گیری نمود.

صاحبان کسب وکار می‌توانند با مورد توجه قرار دادن کاربران با ارزش اثرگذاری بیشتر بر ارزش حاصل شده و سودآوری خود بیفزایند و از این طریق مدیریت بهتر رسانه‌های اجتماعی برای کسب وکارها فراهم خواهد شد. این تقسیم‌بندی در شکل ۴ نشان داده شده است.



شکل ۴- دسته‌بندی کاربران رسانه های اجتماعی

۵. نتیجه‌گیری وپیشنهادات برای کارهای آتی

در این مقاله تلاش شد که بر اساس روش‌ها و مفاهیم موجود در کاوش رسانه‌های اجتماعی، چارچوبی برای ارزیابی ارزش اثرگذاری کاربران در این رسانه‌ها ارائه شود. برای این منظور مطالعات پیشین در این زمینه مورد بررسی قرار گرفت. هر کدام از پژوهش‌های انجام شده در این زمینه، معیارهای مجزایی را برای ارزیابی اثرگذاری در نظر گرفته‌اند. اکثر این مقالات از تحلیل‌های شبکه و معیارهای مربوط به ساختار شبکه‌ای به منظور رتبه‌بندی نودها و پیدا کردن نودهای اثرگذارتر استفاده کرده‌اند. در تعداد دیگری از مقالات خصیصه‌های دیگری مانند ویژگی-های متنی و فعالیت هم در نظر گرفته شده است.در این مقاله با ادغام مدل‌ها و معیارهای موجود یک چارچوب تئوریک پیشنهاد شد. بر اساس یافته‌های این پژوهش:

✓ ارزش اثرگذاری کاربران به صورت تابعی از هر دوی پتانسیل اثرگذاری و میزان اثرگذاری در نظر گرفته شده است.

✓ در پتانسیل اثرگذاری از طریق بررسی میزان محبوبیت و شهرت کاربران در داخل رسانه‌ی اجتماعی میزان قابلیت کاربران برای اثرگذاری بر دیگران به طور غیرمستقیم لحاظ می‌شود. معیارهای رتبه‌ی کاربرمبتنی بر ویژگی‌های

ساختاری یا فعالیت، کیفیت و میزان تحلیلی و قضاوتی بودن متون تولید شده توسط کاربران، برای محاسبه‌ی این بعد ارائه گردیده است.

✓ میزان اثرگذاری از طریق معرفی پارامتر تعداد کاربران فعال قابلیت کاربران برای اثرگذاری به طور مستقیم مورد ارزیابی قرار می‌گیرد. خصیصه‌های مبتنی بر شباهت، خصیصه‌های مبتنی بر دانش و خصیصه‌های مبتنی بر ارتباط برای اندازه‌گیری شدت ارتباط بین کاربران جهت محاسبه‌ی تعداد کاربران فعالارائه گردیده است.

به این ترتیب سوالات اساسی مطرح در بخش مقدمه‌ی این مقاله پاسخ داده می‌شود. نتایج این تحقیق می‌تواند به سازمان‌ها کمک نماید از طریق شناسایی و مدیریت کاربران اثرگذار، راه‌های بهتری را برای انتشار پیام‌های بازاریابی خود و توسعه تبلیغات کلامی مثبت برای شرکت و محصولاتشان در رسانه‌های اجتماعی طراحی نمایند. همچنین نتایج این مقاله می‌تواند به عنوان پایه‌ای برای تحقیقات آتی به خصوص در زمینه ایجاد ارزش مشارکتی به خدمت گرفته شود.

منابع

1.Agarwal, N., Kumar, S., Gao, H., Zafarani, R., & Liu, H. (2011). Analyzing Behavior Of The Influentials Across Social Media.

2.Bi, B., Tian, Y., Sismanis, Y., Balmin, A. & Cho. J. (2014). Scalable topic-speci_c inuence analysis on microblogs.
In Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM '14, 513-522, New York, NY, USA.

3.Eirinaki, M., Monga, S.,Sundaram, S. (2012). Identification of influential social network users. *International Journal of Web Based Communities* 8(2),136–158

4.Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. Nat Phys, 6(11), 888-893. doi: http://www.nature.com/nphys/journal/v6/n11/abs/nphys1746.html#supplementary-information

5.Kumar, V., Bhaskaran, V., Mirchandani, R., Shah,. Milap, S. (2013). Creating a Measurable Social Media Marketing Strategy: Increasing the Value and ROI of Intangibles and Tangibles for Hokey Pokey. *Journal of Marketing Science*, 32(2), 194–212.

6.Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, Th., & Tillmanns, S. (2010). Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value. *Journal of Service Research*, 13(3), 297-310.

7.Lerman, K., & Ghosh, R. (2010). Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks. CoRR, abs/1003.2664.

8.Liang, T. P., & Turban, E. (2012). Introduction to the Special Issue Social Commerce: A Research Framework for Social Commerce, *International Journal of Electronic Commerce*, 16(2), 5–13.

9.Libai, B., Muller, E. & Peres, R. (2012). Decomposing the Value of Word-of-Mouth Seeding Programs: Acceleration vs. Expansion, Journal of Marketing Research.

10. Li, Y., Ma, S., Zhang, Y. & Huang, R. (2013). An improved mix framework for opinion leader identification in online learning communities, Knowledge-Based Systems, (43), 43–51.

11. Li, Y.-M., Lai, C.-Y., and Chen, C.-W. (2011). Discovering influencers for marketing in the blogosphere. Inf. Sci., 181(23), 5143-5157.

12. Li, Y.-M., Lin, C.-H., & Lai, C.-Y. 2010. Identifying influential reviewers for word-of-mouth marketing. Electronic Commerce Research and Applications. 9(4), 294-304.

13. Goldenberg,J., Hang, S., Lehmann, D., & Weon Hong, J.(2009). The Role of Hubs in the Adoption Process. *Journal of Marketing*, 73(2), 1-13

14. Greve,G. (2014). The moderating effect of customer engagement on the brand image – brand loyalty relationship,

Identifying Influential Users in Online Social Networks, *Business & Information Systems Engineering*, (3), DOI 10.1007/s12599-013-0263-7.
25. Raamakirtinan, S. & Livingston, L. M. J. (2016). Identifying Influential Users in Facebook - A Sentiment Based Approach. Indian Journal of Science & technology, 9(10).
26. Saito, K., Kimura, M., Ohara, K., & Motoda, H. (2012). Efficient discovery of influential nodes for SIS models in social networks. *Knowledge and Information Systems*, 30(3), 613-635. doi: 10.1007/s10115-011-0396-2
27. Yong, S. K. & Tran, V. L. (2013). Assessing the Ripple Effects of Online Opinion Leaders With Trust and Distrust metrics, *Expert Systems with Applications*, (40), 3500–3511.
28. Zafarani, R. & Liu, H., 2014. Behavior Analysis in Social Media, *IEEE Intelligent Systems*, 29 (4), 1-4.
29. Zhang, Y., Wang, Z., & Xia, C. (2010). Identifying Key Users for Targeted Marketing by Mining Online Social Network. Paper presented at the Proceedings of the 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops.
30. Zhao, K., Greer, G. E., Y, J., Mitra, P., Portier, K. (2014). Leader identification in an online health community for cancer survivors: a social network-based classification approach, Information System E-Business Management, DOI 10.1007/s10257-014-0260-5.
31. Zheng, X., Zhu, S., & Lin, Z. (2013). Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach, Decision Support Systems, 56(0), 211-222.
32. Zolfaghar, K. & Aghaie, A. (2011). A syntactical approach for interpersonal trust prediction in social web applications: Combining contextual and structural data, Knowledge-Based Systems.doi:10.1016/j.knosys.2010.10.07
, 148, 203 – 210.
15. Hajli, M. N. (2014). The role of social support on relationship quality and social commerce. *Technological Forecasting and Social Change*, 87(0), 17-27.
16. Heidemann, J., Klier, M., Probst, F., & Betriebswirtschaftslehr, U. L. f. (2010). Identifying Key Users in Online Social Networks:A PageRank Based Approach.
17. Ilyas, M. U., & Radha, H. (2011). Identifying Influential Nodes in Online Social Networks Using Principal Component Centrality. *Paper presented at the Communications (ICC)*, 2011 IEEE International Conference on
18. Jiliang, T., Yi, Ch. & Huan, L. 2014, Mining Social Media with Social Theories: A Survey, ACM SIGKDD Explorations Newsletters, 15(2), 20-29.
19. Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68.
20. Katsimpras, G., Vogiatzis, D., & Paliouras, G. (2015). Determining Influential Users with Supervised Random Walks. Paper presented at the Proceedings of the 24th International Conference on World Wide Web, Florence, Italy.
21. Kiss, C., & Bichler, M. (2008). Identification of influencers — measuring influence in customer networks. *Decision Support Systems*, 46(1), 233-253. doi: http://dx.doi.org/10.1016/j.dss.2008.06.007
22. Munger, T., & Zhao, J. (2015). *Identifying Influential Users in On-line Support Forums using Topical Expertise and Social Network Analysis*. Paper presented at the Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Paris, France.
23. Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B. & Pauwels,K. (2013). Social Media Metrics - A Framework and Guidelines for Managing Social Media, *Journal of Interactive Marketing*, 27, 281 – 298.
24. Probst, F., Grosswiele, L. & Pflege, R. (2013). Who will lead and who will follow:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

معرفی انجمن فناوری اطلاعات و ارتباطات ایران

اهداف

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

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

برنامه‌ها

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

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

برنامه‌ریزی فعالیتها

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

کمیته‌های تخصصی انجمن

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

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

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

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

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

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

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

Matched grouping of learners in e-learning environment using council clustering method

Malihe Kamareiy Gholamali Montazer

Abstract:

E-learning has been widely welcomed as a new method of teaching and learning in recent years. Most e-learning systems provide the same learning content for all learners, regardless of individual differences, while in web-based learning, in addition to increasing learning opportunities, learning outcomes must be improved. Also be noted. Adaptive learning systems need to group learners with similar interests to provide adaptable content to each learner, and learners' learning styles can be used to achieve this goal. Automatic grouping of learners in this environment is possible with the help of clustering methods. Due to the different results of clustering methods in different repetitions, in this study, the council clustering method for combining the results of five clustering methods of FCM, K-means, KNN, SVM and medoids-K methods for grouping learners in An e-learning environment is used. The results of experimental evaluation of the proposed clustering method based on three indicators "Davis-Boldin", "purity and aggregation" and "variance" show that this method has reduced computational cost and more accuracy and speed than other common methods in identifying groups.

Keywords: E-learning, learner grouping, clustering, learning style, council approach

Using web analytics in forecasting the stock price of chemical products group in the stock exchange

Amir Daee OmidMahdi Ebadati Keyvan Borna

Abstract:

Forecasting markets, including stocks, has been attractive to researchers and investors due to the high volume of transactions and liquidity. The ability to predict the price enables us to achieve higher returns by reducing risk and avoiding financial losses. News plays an important role in the process of assessing current stock prices. The development of data mining methods, computational intelligence and machine learning algorithms have led to the creation of new models in prediction. The purpose of this study is to store news agencies' news and use text mining methods and support vector machine algorithm to predict the next day's stock price. For this purpose, the news published in 17 news agencies has been stored and categorized using a thematic language in Phoenician. Then, using text mining methods, support vector machine algorithm and different kernels, the stock price forecast of the chemical products group in the stock exchange is predicted. In this study, 300,000 news items in political and economic categories and stock prices of 25 selected companies in the period from November to March 1997 in 122 trading days have been used. The results show that with the support vector machine model with linear kernel, prices can be predicted by an average of 83%. Using nonlinear kernels and the quadratic equation of the support vector machine, the prediction accuracy increases by an average of 85% and other kernels show poorer results.

Keywords:Text mining, web content exploration, web crawler, stock market forecasting, backup vector machine

Routing improvement to control congestion in software defined networks by using distributed controllers

Saeid Bakhtiyari Ardeshir azarnejad

Abstract:

Software defined networks (SDNs) are flexible for use in determining network traffic routing because they separate data plane and control plane. One of the major challenges facing SDNs is choosing the right locations to place and distribute controllers; in such a way that the delay between controllers and switches in wide area networks can be reduced. In this regard, most of the proposed methods have focused on reducing latency. But latency is just one factor in network efficiency and overall cost reduction between controllers and related switches. This article examines more factors to reduce the cost between controllers and switches, such as communication link traffic. In this regard, a cluster-based algorithm is provided for network segmentation. Using this algorithm, it can be ensured that each part of the network can reduce the maximum cost (including delays and traffic on links) between the controller and its related switches. In this paper, using Topology Zoo, extensive simulations have been performed under real network topologies. The results of the simulations show that when the probability of congestion in the network increases, the proposed algorithm has been able to control the congestion in the network by identifying the bottleneck links in the communication paths of each node with other nodes. Therefore, considering the two criteria of delay and the degree of busyness of the links, the process of placing and distributing the controllers in the clustering operation has been done with higher accuracy. By doing so, the maximum end-to-end cost between each controller and its related switches, in the topologies Chinanet of China, Uunet of the United States, DFN of Germany, and Rediris of Spain, is decreased 41.2694%, 29.2853%, 21.3805% and 46.2829% respectively.

Keywords: Software-based networks, distributed controllers, controller placement, clustering, congestion control.

Provide a model of human resource knowledge architecture in knowledge-based organizations using a mixed approach

Abdollah Saedi Reza Sepahvand Najmoddin Mosavi Mohammad Hakkak

Abstract:

Knowledge architecture of human resources is more important than any other tool or element in the way of creating, organizing, storing, distributing and applying knowledge to achieve organizational goals. The present study was conducted to present the architectural model of human resource knowledge in knowledge-based organizations using a mixed approach. This research is based on mixed research and quantitatively and qualitatively, which is descriptive-survey in terms of purpose, application and nature and method. The statistical population of the present study consists of knowledge-based organizations in Lorestan province, 30 of whose experts have been selected based on the principle of theoretical adequacy and using purposive sampling. The data collection tool in the qualitative part of the research is a semi-structured interview and in the quantitative part is a questionnaire. In the qualitative section, data and information were identified using Atlas.ti software and analysis coding method and architectural indicators of human resource knowledge. In the quantitative part of the research, the final model of the research has been developed and presented using Matlab software and interpretive structural modeling method. The research results include the indicators and components of human resource knowledge architecture and the presentation of human resource knowledge architecture model in knowledge-based organizations. Thus, in addition to developing a human resource knowledge architecture architecture model, the findings indicate the identification of the main components of human resource knowledge architecture, knowledge management infrastructure, professional characteristics, situational characteristics and achievements of human resource knowledge architecture.

Keywords: Knowledge Architecture, Human Resource Knowledge Architecture, Blended Approach, Knowledge-Based Organizations

Investigating the effect of using various marketing strategies on social networks on gaining the trust of council customers

Farzaneh Milani Jafar zenouzi

Abstract:

The aim of this study is to investigate the effect of using variety of marketing strategies in social networks to build customers' trust. Marketing through social networks has made appropriate opportunities for companies to attract more customers. Building customers' trust and attracting the customers can be mentioned as marketing challenges on these networks. So, in order to create competitive advantages, companies need to use appropriate strategies of building trust. The population of this study consists of all Iranian users of social networking sites that affected by companies advertisements. Also the sample size by using snowball sampling method is 446. The research method is descriptive survey research and data collection tool is questionnaire. To test hypotheses the partial least squares (PLS) technique and SmartPLS 3 software has been used. The results show that all four variables include: transactional, relationship, database and knowledge-based marketing strategies in social networks have a significant impact to build customers' trust. Indeed, transactional strategy has negative impact on trust so the relation between this variable and dependent variable is reverse. knowledge-based marketing strategy has the most positive impact on customers' trust.

Keywords:Marketing strategies, social networks, customer trust

Criteria for evaluating the effectiveness of social media users - a framework based on social media exploration

Rojyar Pirmohammadiani shahriyar Mohammadi

Abstract:

Nowadays, users' interactive behaviors on social media have become an important and influential resource on marketing activities in various businesses. Despite the importance of this issue, providing appropriate criteria for evaluating the influential behavior of users in recent studies has received less attention.

For this purpose, in the first step, an innovative theory framework including two main dimensions: potential of the influence and the level of the influence is presented. Then, in order to define criteria for measuring each dimension, by providing a comprehensive and combined classification including three domains, user-based analysis, relationship-based analysis and content-based analysis, exploration techniques Social media has been examined to analyze the effective behaviors of users.

In the following, according to the literature review, the criteria of "number of active users", “ranked of users based on the structural indexes and activity", “quality and the subjectiveness of content” have been defined to measure each of the aforementioned dimensions. The criteria proposed in this article are effective for creating dashboards to assess the value of users' influence in various businesses.

It also a comprehensive roadmap has been provided for businesses about the data they need to collect and the required techniques to determine each of these metrics through a cross-disciplinary and academic classification of social media exploration techniques.

Keywords: Impact value, social media exploration, user interactive behaviors

Contents

Iranian Journal of
Information Technology & Communication
No. 39-40, Vol.11, March –September-2019

- **Matched grouping of learners in e-learning environment using council clustering method** 1
  Malihe Kamareiy, Gholamali Montazer
- **Using web analytics in forecasting the stock price of chemical products group in the stock exchange** 19
  Amir Daee, OmidMahdi Ebadati, Keyvan Borna
- **Routing improvement to control congestion in software defined networks by using distributed controllers** 49
  Saeid Bakhtiyari, Ardeshir azarnejad
- **Provide a model of human resource knowledge architecture in knowledge-based organizations using a mixed approach** 73
  Abdollah Saedi , Reza Sepahvand, Najmoddin Mosavi, Mohammad Hakkak
- **Investigating the effect of using various marketing strategies on social networks on gaining the trust of council customers** 91
  Farzaneh Milani, Jafar zenouzi
- **Criteria for evaluating the effectiveness of social media users - a framework based on social media exploration** 109
  Rojyar Pirmohammadiani, shahriyar Mohammadi