KNO-1102-4402

سیستم توصیه گر معنایی وب بر اساس پروفایل کاربر با استفاده از حداکثرسازی نفوذ در شبکه کاربران مشابه و یادگیری عمیق

[۲]، منا مرادی[۳]*مستوره معینی[۱]، علی برومندنیا

۱- دانشجوی دکتری ، دانشکده کامپیوتر، واحد تهران جنوب، دانشگاه آزاد اسلامی ، تهران، ایران

۲- عضو هیئت علمی ، دانشکده کامپیوتر، واحد تهران جنوب ، دانشگاه آزاد اسلامی، تهران ، ایران. *

۳- عضو هیئت علمی ، دانشکده کامپیوتر، واحد تهران مرکز ، دانشگاه آزاد اسلامی، تهران ، ایران.

[1]st_m_moini@azad.ac.ir, [2*] Ali.Broumandnia@iau.ac.ir , [3] mona.moradi@iau.ac

چکیده

سیستم‌های توصیه‌گر صفحات وب در راستای ارائه پیشنهادات مرتبط با محتوای وب‌سایت‌ها و صفحات اینترنتی به کاربران به منظور کاهش سربار اطلاعاتی مورد استفاده قرار می گیرند. این سیستم‌ها با استفاده از تحلیل نظرات و بازخوردهای کاربران ، می‌توانند الگوهای رفتاری و ترجیحات کاربران را شناسایی کرده و بر اساس آن‌ها پیشنهادات مناسب ارائه دهند. در این مقاله از ترکیب مسئله به حداکثر رساندن تأثیر اجتماعی و فیلتر مشارکتی مبتنی بر شبکه خودرمزگذار عمیق برای تخمین رتبه بندی کاربران در شبکه کاربران مشابه بر اساس نظرات کاربران استفاده شده است. روش پیشنهادی با یافتن شبکه کاربران مشابه بر اساس نظرات کاربران، اقدام به یافتن کاربران تأثیر گذار بر اساس الگوریتم ژنتیک می نماید. کاربران تأثیرگذار، گروه کاربران مشابه از نظر رتبه بندی را به عنوان ورودی شبکه خودرمزگذار عمیق تشکیل می دهند که خروجی آن تخمین رتبه بندی کاربران جدید است. در نهایت توصیه های بازدید وب بر اساس فیلتر مشارکتی حاصل از پیش بینی آیتم های مورد علاقه کاربران تأثیرگذرا برای کاربران جدید ارائه می شود. روش پیشنهادی با مقدار MAE، RMSE، Hit Rate و Accuracy به ترتیب برابر با ۰,۰۰۲۴، ۰,۰۰۲۵، ۹۵,۳۱٪ و ۹۵,۰۸ نسبت به سایر روش های پیشین نتایج بهتری به دست آورده است.

کلمات کلیدی : سیستم توصیه گر، پروفایل کاربری، شبکه های خودرمزگذار، فیلتر مشارکتی، الگوریتم ژنتیک

۱- مقدمه

سیستم‌های توصیه‌گر نوعی سیستم فیلتر اطلاعات هستند که هدف آن پیش‌بینی و ارائه توصیه‌های شخصی شده به کاربران است[۱]. این سیستم‌ها معمولاً در وب‌سایت‌های تجارت الکترونیک، پلتفرم‌های جریان، رسانه‌های اجتماعی و سایر خدمات آنلاین برای پیشنهاد محصولات، فیلم‌ها، موسیقی یا سایر محتوایی که کاربر ممکن است به آن علاقه‌مند باشد استفاده می‌شود. هدف نهایی سیستم‌های توصیه‌گر بهبود کاربر است. رضایت و تعامل، منجر به افزایش ماندگاری کاربر و در نهایت سودآوری کسب و کار می‌شود[۲]. سیستم‌های توصیه‌کننده توسط الگوریتم‌های یادگیری ماشینی هدایت می‌شوند که رفتار کاربر، اولویت‌ها و ویژگی‌های آیتم را تجزیه و تحلیل می‌کنند تا پیش‌بینی کنند که کاربر چه چیزی را دوست دارد. این الگوریتم‌ها را می‌توان بر اساس تکنیک‌های مختلفی مانند فیلتر مشارکتی، فیلتر مبتنی بر محتوا و رویکردهای ترکیبی استوار کرد. موفقیت یک سیستم توصیه‌گر به توانایی آن در درک و پیش‌بینی دقیق ترجیحات کاربر، و همچنین ظرفیت آن برای انطباق با تغییر رفتار و ترجیحات کاربر در طول زمان بستگی دارد[۳]. کاربران برای کاهش مشکل اضافه بار اطلاعات به سیستم‌های توصیه‌گر وابسته هستند و در میان طیف وسیعی از آیتم‌ها

مانند محصولات، فیلم‌ها، اخبار یا رستوران‌ها پیمایش می‌کنند تا موارد مورد علاقه‌شان را بیابند. در نتیجه، گرفتن دقیق ترجیحات کاربران از تعاملات گذشته آنها (به عنوان مثال، کلیک‌ها، بازدیدها، خواندن‌ها و خریدها) برای اثربخشی یک سیستم توصیه‌گر بسیار مهم است[۴].

سیستم‌های توصیه‌گر یک موضوع تحقیقاتی محبوب در زمینه‌های یادگیری ماشین، داده کاوی، بازیابی اطلاعات و تعامل انسان و کامپیوتر است. محققان به طور مداوم در حال توسعه و بهبود الگوریتم‌ها و متدولوژی‌ها برای ساختن سیستم‌های توصیه‌گر دقیق‌تر و موثرتر هستند. علاوه بر این، علاقه فزاینده‌ای به توسعه سیستم‌های توصیه‌کننده بر اساس روابط بین کاربران به عنوان مثال در شبکه‌های اجتماعی، اطلاعات مربوط به پروفایل کاربران و یا اعطای نظرات و رتبه دهی به اقلام مورد توصیه وجود دارد [۵, ۶].

شبکه اجتماعی یک ساختار به هم پیوسته از گروهی از عوامل است که برای تعاملات اجتماعی تشکیل شده است. امروزه شبکه‌های اجتماعی نقش مهمی در انتشار اطلاعات دارند و اهمیت عملی زیادی در بازاریابی ویروسی دارند. در مورد بازاریابی ویروسی یک خانه تجاری را در نظر بگیرید که در آن هدف جذب کاربران برای خرید یک محصول خاص است. بهترین راه برای انجام این کار این است که مجموعه‌ای از کاربران بسیار تأثیرگذار را انتخاب کنید و نمونه‌های رایگان آن‌ها را توزیع کنید. اگر محصول را دوست داشته باشند، اطلاعات را با همسایگان خود به اشتراک خواهند گذاشت. به دلیل نفوذ زیاد، بسیاری از همسایگان برای محصول تلاش می‌کنند و اطلاعات را با همسایگان خود به اشتراک می‌گذارند. این روند آبشاری ادامه خواهد یافت و در نهایت بخش بزرگی از کاربران برای محصول تلاش خواهند کرد. طبیعتا تعداد محصولات نمونه رایگان به دلیل اقتصادی محدود خواهد بود. از این رو، اگر بتوان نمونه‌های رایگان را بین کاربران بسیار تأثیرگذار توزیع کرد و مشکل در اینجا به انتخاب کاربران تأثیرگذار از شبکه ختم شود، این فرآیند مثمر ثمر خواهد بود. این مشکل به عنوان مشکل به حداکثر رساندن تأثیر اجتماعی شناخته می‌شود[۷]. حال می توان این مسئله را در مورد بازدید وب نیز تعمیم داد. بدین ترتیب صفحات وبی که توسط کاربران تأثیرگذار مورد پسند قرار گیرند، می‌توانند در میان همسایگان آنها نیز پخش شده و توسط آن ها نیزمورد پسند قرار گیرند.

در این مقاله از ترکیب مسئله به حداکثر رساندن تأثیر اجتماعی و فیلتر مشارکتی برای تخمین رتبه‌بندی کاربران در شبکه کاربران مشابه بر اساس پروفایل کاربری کاربران در شبکه‌های اجتماعی استفاده شده است. روش پیشنهادی در ابتدا بر اساس نظرات و رتبه بندی کاربران در مورد اقلام، اقدام به یافتن ماتریس تشابه کاربران می نماید. سپس با استفاده از یک شرط آستانه، کاربرانی که تشابه بیشتری با هم دارند به عنوان کاربران عضو جامعه کاربران مشابه، در ماتریس مجاورتی شباهت مقدار ۱ دریافت می‌کنند و کاربرانی که در شرط آستانه صدق نمی‌کنند، مقدار صفر دریافت خواهند کرد. بر اساس ماتریس مجاورتی شباهت، شبکه کاربران مشابه تشکیل می‌یابد. سپس در روش پیشنهادی اقدام به یافتن کاربران تأثیرگذار در شبکه کاربران مشابه با استفاده از الگوریتم ژنتیک خواهیم نمود. الگوریتم ژنتیک مورد استفاده به عنوان یک روش بهینه‌سازی به منظور یافتن کمترین تعداد کاربران با بیشترین تأثیر در شبکه کاربران مشابه مورد استفاده قرار می‌گیرد. کاربران تأثیرگذاری که توسط الگوریتم ژنتیک پیدا شده به عنوان هسته گروه‌های مشابه از همسایگان در نظر گرفته می‌شوند که رتبه‌بندی مشابهی دارند. بدین ترتیب رتبه‌بندی برای سایر کاربران و کاربران جدید در سیستم توصیه‌گر بر اساس شبکه خودرمزگذار عمیق تخمین زده خواهد شد. در واقع ورودی شبکه های عصبی خودرمزگذار عمیق پیشنهادی، شامل نظرات کاربران تأثیرگذار و سایر کاربران مشابه به عنوان ویژگی و رتبه‌بندی کاربران به عنوان متغیر هدف است. شبکه خودرمزگذار پیشنهادی با استخراج اطلاعات اساسی در هر ویژگی و ایجاد کاتالوگی بر اساس نظرات اعطا شده به هر ویژگی، توانایی پیش‌بینی و تخمین رتبه‌بندی مربوط به نظرات سایر کاربران و کاربران جدید در مجموعه داده تست را خواهد داشت. در روش پیشنهادی علاوه بر شبکه خودرمزگذار عمیق از چندین روش طبقه بندی دیگر به منظور بررسی قدرت یادگیری در روش پیشنهادی استفاده شده است.

این فرآیند بخشی از فیلتر مشارکتی است، یک تکنیک سیستم توصیه محبوب که در آن از ترجیحات گروهی از کاربران برای پیش‌بینی خودکار در مورد علایق کاربران استفاده می‌شود. پس از رتبه‌بندی کاربران جدید و تعیین شباهت آنها با کاربران موجود، فرآیند توصیه شامل شناسایی کاربرانی با شباهت بالا و توصیه مواردی است که بالاترین امتیاز را از این کاربران مشابه دریافت کرده اند. این رویکرد از خرد جمعی جامعه کاربر برای ارائه

توصیه‌های شخصی به تک تک کاربران استفاده می‌کند. اهمیت آکادمیک در بکارگیری اصول فیلتر مشترک برای افزایش دقت و مرتبط بودن توصیه‌های آیتم‌ها نهفته است و از این طریق به سیستم‌های بازیابی اطلاعات شخصی و توصیه کمک می‌کند. پس از رتبه‌بندی کاربرانی که به تازگی به وب‌سایت‌ها اختصاص داده شده‌اند، ارزیابی شباهت بین این کاربران جدید و پایگاه کاربران موجود انجام می‌شود. متعاقباً، توصیه‌های آیتم برای کاربر بر اساس شباهت بین کاربران و موارد ترجیحی آنهایی که مشابه تلقی می‌شوند تعیین می‌شوند. کاربرانی که شباهت قابل توجهی در ویژگی های خود نشان می دهند، و همچنین میزان تشابه بیشتری را به عنوان کاربران مشابه در نظر می گیرند. موارد K برتر، شامل موارد مورد علاقه با بالاترین امتیاز از این کاربران مشابه، سپس به آنها توصیه می‌شود. در اصل، سهم اصلی این مقاله را می توان به این صورت خلاصه کرد.

1. **استخراج ماتریس مجاورتی شباهت بر اساس تحلیل معنایی نظرات کاربران**
2. **یافتن کاربران تأثیر گذار به منظور حداکثر سازی نفوذ در شبکه کاربران مشابه با استفاده از الگوریتم ژنتیک**
3. **ایجاد گروه های کاملا مشابه به کاربران تأثیرگذار بر اساس رتبه بندی به عنوان ویژگی های ورودی شبکه عصبی رمزگذارخودکار عمیق در راستای تخمین رتبه بندی کاربران جدید**
4. **یافتن کاربران – آیتم های مشابه بر اساس رویکرد فیلتر مشارکتی مبتنی بر شبکه خودرمزگذار عمیق پیشنهادی و توصیه آیتم های برتر**

ساختار ادامه مقاله به شرح زیر است. در بخش دوم کارهای مرتبط بررسی خواهد شد. در بخش سوم جزئیات روش شناسی مقاله ارائه خواهد شد. در بخش چهارم نتایج آزمایشات ارائه خواهد شد. در بخش پنجم نتیجه‌گیری و کارهای آتی مقاله بیان خواهد شد.

۲- کارهای مرتبط

در سیستم‌های توصیه‌گر، یافتن کاربران تأثیرگذار بسیار اهمیت دارد زیرا این کاربران می‌توانند تأثیر بزرگی بر رفتار و نگرش دیگر کاربران داشته باشند. با انتخاب کاربرانی که بیشترین توانایی انتشار اطلاعات را دارند، سیستم توصیه‌گر می‌تواند بهبود یابد و اطمینان حاصل شود که پیشنهادات به گروه‌های بزرگتری انتشار می‌یابد. این امر می‌تواند منجر به افزایش ارتباطات و تعاملات درون شبکه‌های اجتماعی شود و در نهایت به بهبود عملکرد سیستم توصیه‌گر و ارتقاء تجربه کاربران منجر شود. از این رو در ادامه به بررسی برخی از سیستم های توصیه گر که از حداکثرسازی تأثیر استفاده نموده اند، خواهیم پرداخت.

در[۸] یک مدل انتشار نفوذ عصبی به نام DiffNet معرفی شده است که فرآیند انتشار اجتماعی بازگشتی را برای هر کاربر ثبت می‌کند تا نفوذ شبکه اجتماعی مرتبه بالاتر را در توصیه به کاربر بگنجاند. با این حال، آنها استدلال کرده اند که تمرکز صرفاً بر نفوذ شبکه های اجتماعی، منافع مشترک کاربران در شبکه مورد علاقه کاربر را نادیده می گیرد. برای پرداختن به این موضوع، در [۹]، DiffNet++ پیشنهاد شده است، یک الگوریتم پیشرفته که انتشار نفوذ عصبی و انتشار علاقه را در یک چارچوب یکپارچه ادغام می‌کند. با فرمول‌بندی توصیه‌های اجتماعی به‌عنوان یک نمودار ناهمگن با ورودی‌هایی از شبکه‌های اجتماعی و علاقه‌مندی، DiffNet++ با تجمیع مکرر توصیه های هر کاربر از توصیه قبلی‌اش، تجمع نفوذ همسایگان اجتماعی، و تجمیع علاقه همسایگان آیتم، DiffNet را بهبود می‌بخشد. در[۱۰]، به مشکل توصیه گروهی با ترکیب انتشار اطلاعات در یک شبکه اجتماعی به عنوان یک عامل کلیدی برای به حداکثر رساندن تأثیر می پردازد. بر خلاف مشکل بیشینه سازی نفوذ سنتی، که در آن هدف انتخاب کاربران برای انتشار اطلاعات است، این مطالعه بر توصیه مواردی به یک گروه معین تمرکز دارد که با ترجیحات آن‌ها همسو هستند و پتانسیل انتشار گسترده در شبکه اصلی را دارند. در، [۱۱] یک رویکرد چندهدفه را برای مشکل حداکثرسازی نفوذ در کمپین‌های بازاریابی ویروسی در سیستم های توصیه گر با هدف افزایش درآمد و در عین حال کاهش هزینه‌ها معرفی شده است. با استفاده از معیارهای شبکه اجتماعی محلی برای شناسایی افراد تأثیرگذار، این مطالعه از الگوریتم‌های بهینه‌سازی چندهدفه تکاملی و یک چارچوب واقعی بازار مبتنی بر عامل برای ارزیابی استراتژی‌های هدف‌گیری استفاده می‌کند. در [۱۲]، مدل‌ها و الگوریتم‌های شناختی را برای ارائه توصیه از طریق شبکه‌های اجتماعی آنلاین با تمرکز بر شناسایی کاربران تأثیرگذار و مدل‌سازی انتشار بدون اطلاعات قبلی ارائه می‌کند. این مطالعه با استفاده از شبکه اجتماعی YELP به عنوان یک مطالعه موردی، تولید یک نمودار تأثیر را بر اساس مسیرهای تأثیرگذار مختلف، مانند بررسی های مشترک، نشان می دهد. در [۱۳]، یک مدل به حداکثر رساندن تأثیر مبتنی بر نظر کاربران را برای بازاریابی ویروسی در سیستم های توصیه گر معرفی می‌کند که به سه سناریو مبتنی بر نظر می‌پردازد. شش معیار تأثیر، دو معیار برای هر سناریو، پیشنهاد شده‌اند و یک مدل حداکثرسازی تأثیر برای شناسایی تأثیرگذاران استفاده شده است.

در [۱۴]، مفهوم اینترنت اجتماعی اشیاء (SIoT) و ارتباط آن را در برنامه های کاربردی مدرن مانند حمل و نقل هوشمند، سیستم های توصیه گر و بازاریابی ویروسی معرفی شده است. این کار، الگوریتم به حداکثر رساندن تأثیر تبلیغات برای اینترنت اجتماعی اشیاء را پیشنهاد می‌کند، که برای انتخاب مؤثر گره‌های تأثیرگذار برای به حداکثر رساندن دسترسی و در عین حال به حداقل رساندن همپوشانی در سیستم های توصیه گر و بازاریابی طراحی شده است.

در [۱۵]، نقش کاربران تأثیرگذار در کسب و کار بررسی شده است. استفاده از اینفلوئنسرها، ایجاد محتوای با کیفیت بالا و استفاده از ابزارها برای گسترش دسترسی، استراتژی های کلیدی هستند. سرمایه اجتماعی دسترسی به منابع را فراهم می کند، شهرت ایجاد می کند و پشتیبانی ارائه می دهد. کسب‌وکارهای موفق قدرت شبکه‌های اجتماعی را در ایجاد روابط با مشتری و ترویج محصولات نشان می‌دهند. به طور کلی، حداکثر کردن نفوذ برای کارآفرینان ضروری است تا دید، اعتبار و نفوذ را در عصر دیجیتال افزایش دهند و به دستیابی موثر به هدف کمک کنند.

با توجه به بررسی های انجام شده می توان گفت، یافتن کاربران تأثیرگذار در شبکه کاربران مشابه برای ارائه توصیه در سیستم‌های توصیه‌گر مفید است زیرا کاربران تأثیرگذار می‌توانند توصیه‌های خود را به گستره‌ای از کاربران منتقل کنند و از این طریق تأثیر بزرگی در فعالیت‌ها و تصمیم‌گیری‌های دیگران داشته باشند. این امر می‌تواند به بهبود دقت و کارایی سیستم توصیه‌گر منجر شود و باعث افزایش رضایت کاربران از سیستم توصیه‌گر شود که به عنوان محرک اصلی این مقاله در نظر گرفته شده است.

۳- روش شناسی

در این مقاله یک سیستم توصیه گر وب بر اساس حداکثر سازی تأثیر در شبکه کابران مشابه بر اساس نظرات کاربران در پروفایل کاربری ارائه شده است. روش پیشنهادی به منظور حداکثر سازی تأثیر از الگوریتم ژنتیک استفاده کرده است. کاربران تأثیرگذاری که بر اساس الگوریتم ژنتیک یافت می شوند به عنوان هسته گروه های مشابه با رتبه کاملا مشابه به عنوان ورودی شبکه خودرمزگذار عمیق در نظر گرفته می شوند. شبکه خودرمزگذار عمیق رتبه تخمینی برای کاربران جدید را مهیا می کند و در نهایت با استفاده از رویکرد فیلتر مشارکتی سعی در یافتن K آیتم توصیه برتر به کاربران جدید خواهیم نمود. فلوچارت روش پیشنهادی در شکل ۱ نشان داده شده است.

۳-۱ پیش پردازش نظرات کاربران

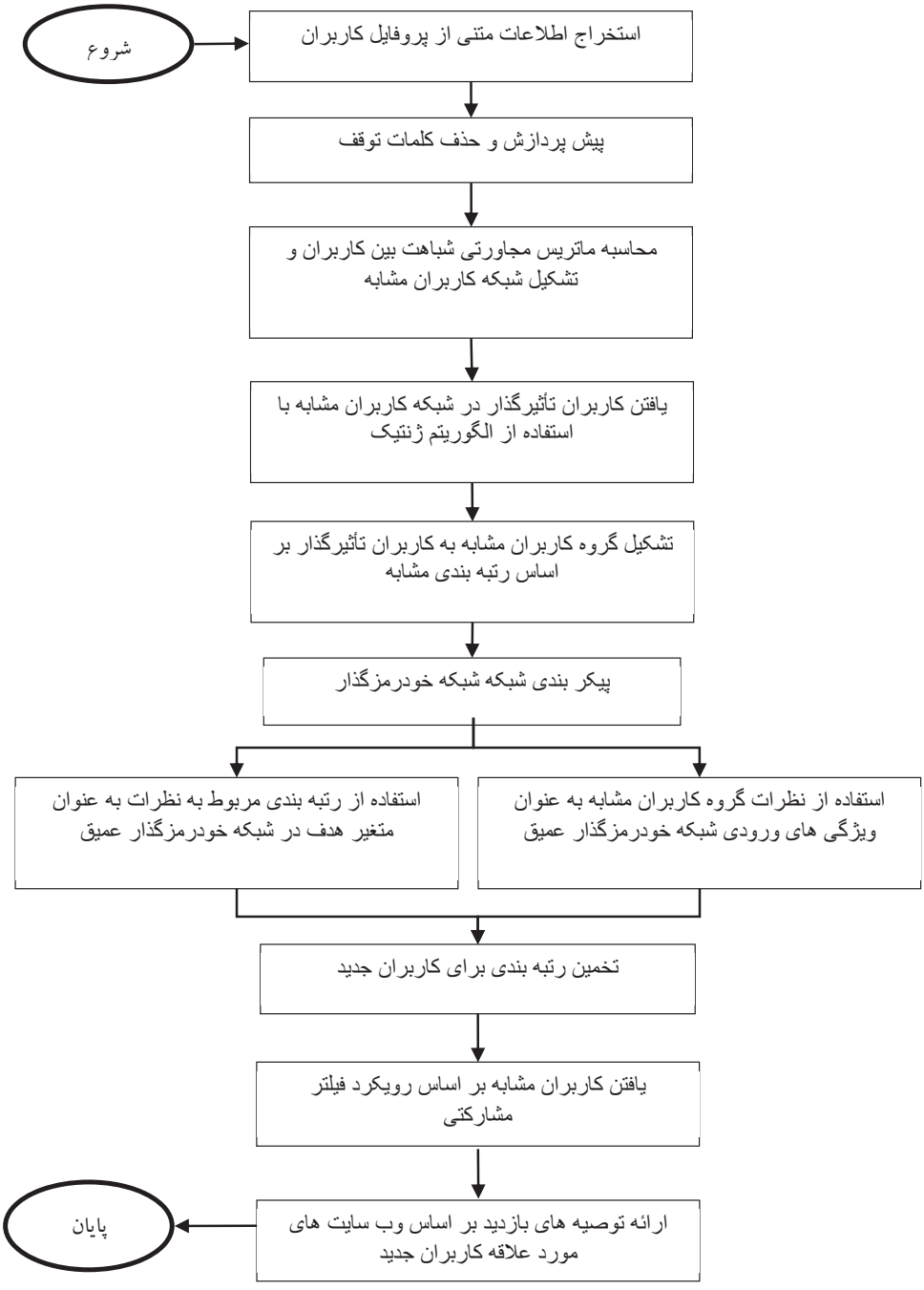
روش پیشنهادی شامل استخراج ماتریس شباهت کاربر بر اساس نظرات آنها با استفاده از معیار تشابه کسینوس است که معمولاً در بازیابی اطلاعات استفاده می شود. این اندازه گیری یک سند متنی را به عنوان بردار اصطلاحات مدل می کند و مقدار کسینوس بین بردارهای کلمات در دو متن را برای تعیین شباهت آنها محاسبه می کند. با این حال، در حالی که شباهت کسینوس به طور موثر شباهت سطح کلمه بین اسناد را اندازه گیری می کند، برای بررسی شباهت معنایی محتوای متن نیز تلاش می کند. این می تواند منجر به نتایج نامعتبر در تطابق نحوی شود و ممکن است تفاوت های معنایی مورد نظر را برطرف نکند. در زمینه سیستم های بازیابی اطلاعات، این می تواند منجر به نتایج نادرست و کاهش کارایی شود. بنابراین، رابطه معیار تشابه کسینوس باید برای فرآیندهای بعدی به دقت در نظر گرفته شود.

$$Sim(\vec{q},\vec{d}) = \frac{\vec{q}.\vec{d}}{|\vec{q}||\vec{d}|} = \frac{\sum_{k=1}^{t} w_{qk} \times w_{dk}}{\sqrt{\sum_{k=1}^{t}(w_{qk})^2}.\sqrt{\sum_{k=1}^{t}(w_{dk})^2}} \quad (۱)$$

که در آن q وd به متن‌ها اشاره دارند و t تعداد کلمات در متن و w وزن هر کلمه است. بدین ترتیب ماتریس مشابهت بین کاربران بر اساس مقادیر معیار شباهت کسینوسی تشکیل می شود. در روش پیشنهادی برای انتخاب کاربران به عنوان اعضای شبکه کاربران مشابه از یک آستانه استفاده شده است. در این آستانه در صورتی که شباهت بین دو کاربر از ۸۰٪ مقدار کل شباهت های بین کاربران بیشتر باشد، به عنوان کاربران مشابه در

[1] Social Internet of Things

نظر گرفته می شود. کاربران مشابه دارای مقدار ۱ در ماتریس مجاورتی خواهند بود و کاربران غیر مشابه با مقدار صفر در ماتریس مجاورت شباهت، تعیین می شوند. بدین ترتیب در روش پیشنهادی در گام اول ماتریس مجاورتی شباهت نظرات کاربران تعیین می شود.



شکل ۱. فلوچارت روش پیشنهادی

۲-۳ حداکثر سازی تأثیر مبتنی بر الگوریتم ژنتیک

همانطور که اشاره شد، در روش پیشنهادی از الگوریتم ژنتیک به منظور حداکثرسازی تأثیر و انتخاب کاربران تأثیرگذار در شبکه کاربران مشابه، استفاده شده است. بر اساس ماتریس مجاورت، شبکه های پیچیده به عنوان یک گراف $G = (V, E)$ در نظر گرفته می شوند، که در آن V و E نشان دهنده رئوس و یال های نمودار هستند. هر رأس در G نشان دهنده یک کاربر در یک شبکه اجتماعی است و هر یال نشان دهنده ارتباط بین یک جفت کاربر است. اندازه یک شبکه بر اساس تعداد کاربران شبکه $N = |V|$ و تعداد لینک های شبکه $M = |E|$ تعریف می شود. ساختار شبکه به صورت یک ماتریس مجاورت مربعی $A = (aij)n\times n$ نشان داده می شود، که در آن هر عنصر می تواند مقدار {۰، ۱} داشته باشد، که نشان می دهد کاربر i به کاربر j متصل است ($aij = 1$) یا خیر ($aij = 0$). برای شناسایی گره های تأثیرگذار در نمودار شبکه اجتماعی $G(V, E)$، مجموعه ای از S متشکل از k گره انتخاب می شود ($S\subset V(G)$). این گره ها بر اساس توانایی آنها در انتشار اطلاعات در نمودار تحت یک مدل انتشار به حداکثر تعداد گره در حداقل زمان انتخاب می شوند. بنابراین، کاربران K که در مجموعه S^* به عنوان کاربران seed شناخته می شوند، به عنوان کاربران تأثیرگذار بهینه قرار می گیرند و اندازه مجموعه $|S^*|$ برابر K است که بر اساس رابطه ۵ تعیین می شود.[۱۶].

$$S^* \ argmax_{S\subset V,|S|=k}\ \sigma(S) \qquad (۲)$$

در معادله ۲، $\sigma(S)$ یک تابع هدف است که گسترش نفوذ مورد انتظار را بر روی نمودار $G(V, E)$ در پایان فرآیند انتشار اطلاعات به دست می آورد و S^* نشان دهنده K گره ها با حداکثر گسترش نفوذ در گراف شبکه اجتماعی است.

۲-۳-1 جمعیت اولیه

در مسائل بهینه‌سازی و به خصوص در روش‌های فرااکتشافی، تنظیم جمعیت اولیه به گونه‌ای است که هر فرد در جمعیت اولیه یک راه‌حل نهایی را نمایش دهد. این امر امکان ارزیابی بهینگی راه‌حل‌ها بر اساس تابع تناسب فراهم می‌کند. در روش پیشنهادی، هدف یافتن k کاربر از میان سایر کاربران است که بیشترین تأثیر را در شبکه اجتماعی دارند. در الگوریتم ژنتیک، جمعیت اولیه به صورت کروموزوم‌هایی است که از ژن‌هایی در کنار هم تشکیل یک راه‌حل را می‌دهند. جمعیت اولیه در روش پیشنهادی به صورت یک بردار از مقادیر گسسته به عنوان کروموزوم تعریف می‌شود، که طول بردار برابر با K (تعداد کاربران تأثیرگذار) است، و مقدار هر درایه (ژن) از بردار نشان‌دهنده اندیس مربوط به کاربر در شبکه اجتماعی است. در شکل ۲، نموداری از کروموزوم اولیه در روش پیشنهادی نشان داده شده است.

	U_1	U_2	U_3	U_4	U_5
X	16	24	5	10	19

شکل ۲. نمونه ای از کروموزم اولیه

با توجه به شکل ۲ می توان دید، هر راه حل شامل یک بردار با ۵ درایه است که کروموزوم اولیه با تعداد محدودی ژن را تشکیل می دهد. هر ژن اندیس مربوط به یک کاربر در شبکه اجتماعی را نشان می دهند. کاربران در شبکه اجتماعی به صورت تصادفی به عنوان مقادیر ژن ها در کروموزم ها انتخاب می شوند. حال آن که با توجه به تعداد زیاد کاربران در شبکه اجتماعی، انتخاب ژن ها برای کروموزوم های اولیه و همچنین فضای جستجو برای الگوریتم ژنتیک بسیار زیاد خواهد بود. در روش پیشنهادی برای مقابله با این مسئله از یک آستانه بر اساس درجه مرکزیت کاربران در نظر گرفته شده است.

معیار مرکزیت یکی از ابزارهای مهم تجزیه و تحلیل شبکه است که برای شناسایی گره‌های قدرتمندترین و مهم‌ترین در یک شبکه اجتماعی استفاده می‌شود. این معیار به ما کمک می‌کند تا گره‌ها و افرادی که در شبکه اجتماعی بیشترین تأثیر و ارتباط را دارند را شناسایی کنیم. معیار مرکزیت بر

اساس ارتباطات مستقیم دوستی یا ارتباطات میان گره‌ها در شبکه تعریف می‌شود. این معیار میزان اهمیت یک گره را بر اساس تعداد ارتباطات مستقیم آن گره با سایر گره‌ها تعیین می‌کند. به عبارت دیگر، گره‌هایی که دارای بیشترین تعداد ارتباطات مستقیم با دیگر گره‌ها هستند، به عنوان مهم‌ترین و قدرتمندترین گره‌ها شناخته می‌شوند. با استفاده از معیار مرکزیت می‌توانیم افراد یا گره‌هایی را که در شبکه اجتماعی بیشترین تأثیر را دارند، شناسایی کنیم و از این طریق، بهبود استراتژی‌ها و تصمیم‌گیری‌های مربوط به ارتباطات و تأثیرات در شبکه اجتماعی دست یابیم.[۱۷]:

$$C(i) = \sum_{i=1}^{n} \sum_{j=1}^{n-1} \frac{a_{ij}}{n-1} \qquad (۳)$$

که در آن a_{ij} نشان دهنده ارتباط بین کاربر i و j در ماتریس مجاورت و n تعداد کل کاربران در شبکه است. برای تعیین مقدار درجه مرکزیت برای هر کاربر در یک شبکه اجتماعی، می‌توانیم تعداد ارتباطات مستقیم یا درجه هر کاربر را با سایر کاربران در نظر بگیریم. این درجه مرکزیت نشان دهنده میزان محبوبیت یا تأثیرگذاری کاربر در شبکه اجتماعی است. کاربرانی که با تعداد زیادی از دیگر کاربران ارتباط دارند، درجه مرکزیت بالایی دارند و به عنوان کاربران مهم و تأثیرگذار در شبکه شناخته می‌شوند.

در روش پیشنهادی، تنها کاربرانی که درجه مرکزیت بیشتر از مقدار میانگین درجه مرکزیت کل کاربران داشته باشند، به عنوان کاربران نامزد برای جمعیت اولیه انتخاب می‌شوند. این به این معنی است که تنها کاربرانی که میزان تأثیرگذاری بیشتری نسبت به میانگین دارند، به عنوان کاربران تأثیرگذار شناخته می‌شوند و برای شبیه‌سازی‌ها یا تحلیل‌های بعدی انتخاب می‌شوند. در رابطه ۴، آستانه مربوط به درجه مرکزیت برای انتخاب کاربران نامزد در جمعیت اولیه نشان داده شده است. این آستانه نشان می‌دهد که فقط کاربرانی که درجه مرکزیت بیشتر از این آستانه را دارند، به عنوان کاربران تأثیرگذار در نظر گرفته می‌شوند و به عنوان جمعیت اولیه برای مطالعات بعدی انتخاب می‌شوند.

$$\tau = \sum_{i=1}^{n} \frac{C_i}{n} \qquad (۴)$$

که در آن Ci مقدار درجه مرکزیت کاربران و n تعداد کل کاربران در شبکه اجتماعی است. بدین ترتیب کاربرانی با درجه مرکزیت کمتر از میانگین درجه مرکزیت کل کاربران که احتمال بسیار کمی برای تبدیل شدن به کاربرتأثیرگذار دارند، از فضای مسئله حذف می شوند.

۴-۲-۳ تابع تناسب پیشنهادی

این فرآیند شامل انتخاب کاربران تصادفی است که آستانه مشخصی را به عنوان کروموزوم های اولیه برآورده می کنند و به دنبال آن، این کروموزوم های اولیه را برای ارزیابی اثربخشی راه حل های تصادفی ارزیابی می کنند. روش پیشنهادی از مرکزیت به عنوان یک ضریب برای پارامترها در تابع تناسب استفاده می کند که بر اساس درجات ورودی و خروجی کاربران قابل محاسبه است. تابع تناسب اندام پیشنهادی بر اساس این معیار تنظیم خواهد شد. با توجه به تغییرات بالقوه در مقادیر پارامترهای مورد استفاده در تابع تناسب، عادی سازی داده ها و نگاشت آن به یک محدوده ثابت برای ارزیابی دقیق و قابل مقایسه کروموزوم های اولیه ضروری است. در روش پیشنهادی از روش نرمال‌سازی min-max برای این منظور استفاده شده است که به صورت رابطه ۵ بیان شده است[۱۸].

$$ND_i = \frac{D_i - \min(D)}{\max(D) - \min(D)} \qquad (۵)$$

که در آن ND_i به عنوان مقدار نرمال سازی شده داده ها برای کاربر i، D_i به عنوان مقدار اصلی داده ها برای کاربر i، min(D) به عنوان کمترین و max(D) به عنوان بیشترین مقدار داده ها برای کل کاربران در نظر گرفته شده است. پس از نرمال‌سازی مقادیر مربوط به معیارهای اهمیت، تابع تناسب پیشنهادی بر اساس این مقادیر در رابطه ۶ ارائه شده است.

$$max\, f = \sum_{i=1}^{n} w_i^1 C_i +$$

$$s.t.$$
$$\sum_{i=1}^{n} C_i \geq C_\mu$$
$$\sum_{i=1}^{n} w_i = 1$$
$$i > 0 \qquad (۶)$$

که درآن i نشان دهنده اندیس کاربر در شبکه اجتماعی، Ci نشان دهنده مقدار معیار مرکزیت، C_μ نشان دهنده مقدار میانگین معیار مرکزیت برای کل کاربران w_i نشان دهنده وزن معیار مرکزیت برای کاربر i هستند. محدودیت های موجود برای مسئله بهینه سازی حاضر نشان می دهند که مقدار معیار مرکزیت باید بیشتر از مقدار میانگین معیار مرکزیت برای کل کاربران باشد و مقادیر سایر معیارهای مرکزیت بیشتر از صفر باشد. همچنین مجموع وزن های مربوط به مقادیر معیارهای اهمیت باید برابر با یک باشد. پس از یافتن کاربران تأثیرگذار در شبکه کاربران مشابه، همسایگان کاربران مشابه با توجه به ماتریس مجاورتی یافت می شود. از میان همسایگان آن دسته از کاربرانی که رتبه اعطایی توسط آنان دقیقاً مشابه با کاربران تأثیرگذار است به عنوان گروه کاربران مشابه با رتبه یکسان به عنوان ورودی شبکه های خودرمزگذار عمیق در نظر گرفته می شوند. شبکه های خودرمزگار نسبت به استخراج ویژگی های متنی مهم هر گروه نسبت به رتبه اعطایی اقدام می کند. سپس برای کاربران جدید با توجه به نظرات آن ها، رتبه تخمینی توسط شبکه خودرمزگذار عمیق ارائه می شود.

۳-۳ معماری شبکه‌های عصبی خودرمزگذار عمیق پیشنهادی[۲]

در این بخش، بحث پیرامون استفاده از یک شبکه رمزگذار خودکار برای استخراج ویژگی در طبقه‌بندی مدل‌سازی پیش‌بینی‌کننده می‌چرخد. تمرکز اصلی بر آموزش شبکه برای یادگیری یک نمایش فشرده از ویژگی های ورودی و استخراج ویژگی های حیاتی برای هر دسته بر اساس یک متغیر هدف و ویژگی های گروه کاربر است. این شبکه به طور خاص برای تقریب و تکثیر ورودی‌هایی طراحی شده است که مشابه داده‌های آموزشی هستند و آن را قادر می‌سازد تا ویژگی‌های ارزشمندی را بیاموزد. این ویژگی‌های استخراج‌شده و وزن‌های مربوط به آن‌ها متعاقباً برای پیش‌بینی رتبه‌بندی کاربران جدید استفاده می‌شوند. این فرآیند اساساً شامل یادگیری شبکه رمزگذار خودکار و گرفتن ویژگی‌های مهم از داده‌های ورودی است که سپس برای مدل‌سازی پیش‌بینی‌کننده و اهداف رتبه‌بندی کاربر استفاده می‌شود.

در این بخش، روشی برای تعریف رمزگذار و رمزگشا به‌کار گرفته شده است که شامل دو لایه پنهان است. لایه اول از اندازه دو برابر تعداد ورودی (2^n) و لایه دوم از اندازه مشابه با تعداد ورودی (n) تشکیل شده است. سپس یک لایه گلوگاه با همان تعداد ورودی به عنوان مجموعه داده ورودی (n) وجود دارد. برای بهبود یادگیری مدل، از روش نرمال‌سازی دسته‌ای و فعال‌سازی leaky ReLU استفاده شده است.

در مقابل، رمزگشا نیز در یک ساختار مشابه، اما به صورت معکوس تعریف شده است. این رمزگشا نیز دو لایه پنهان دارد، لایه اول از تعداد ورودی‌های مجموعه داده (n) و لایه دوم از اندازه دو برابر تعداد ورودی‌ها (2^n) تشکیل شده است. سپس لایه خروجی همان تعداد گره‌های ستون‌های داده ورودی را دارد و از یک تابع فعال‌سازی خطی برای خروجی مقادیر عددی استفاده می‌کند. مدل با استفاده از بهینه‌ساز Adam و گرادیان تصادفی نزولی، با کمینه کردن میانگین مربعات خطا آموزش داده می‌شود. این مرحله از فرایند، خروجی را به یک ماژول توصیه می‌دهد که در واقع یک سیستم توصیه‌دهنده است. تفاوت اصلی بین شبکه‌های عصبی ارائه شده و شبکه‌های قبلی، تعداد لایه‌های میانی است که اجازه می‌دهد تا ویژگی‌های پروفایل کاربران به صورت عمیق‌تری آموزش داده شود. این موضوع باعث افزایش دقت پیش‌بینی رتبه‌بندی به کاربران جدید می‌شود. به عبارت دیگر، این شبکه‌های عصبی قادرند تا ویژگی‌های پیچیده‌تر و عمیق‌تری را از داده‌ها استخراج کرده و از آن‌ها در فرایند توصیه‌دهی به کاربران استفاده کنند.

[2] Auto encoder deep Neural Network

۳-۲ ماژول توصیه پیشنهادی

ماژول توصیه در روش ارائه شده، آخرین بخشی است که در آن توصیه‌های بازدید صفحات به کاربران براساس نتایج شبکه رمزگذار ارائه شده اتفاق می‌افتد. خروجی شبکه رمزگذار شامل رتبه های پیش بینی شده برای کاربران جدید است. در این مرحله، بر اساس رتبه پیش بینی شده بر ای کاربران ابتدا مقدار میانگین خطای مطلق[3](MAE)، خطای میانگین مربعات[4](RMSE)، برای کاربران محاسبه می شود. سپس توصیه صفحات وب به آن دسته از کاربرانی که رتبه آن ها به درستی پیش بینی شده است، با توجه به صفحات وبی که مقدار رتبه بندی را از کاربران تأثیرگذار مشابه دریافت کرده اند، ارائه می شود. در واقع در این مرحله رویکرد فیلتر مشارکتی به منظور تخمین رتبه یک کاربر نسبت صفحات وبی که کاربر تأثیر گذار در گروه کاربران مشابه بررسی کرده است ولی کاربر هدف آن صفحات وب را بررسی نکرده است، مورد استفاده قرار می گیرد. از این رو رابطه فیلتر مشارکتی در روش پیشنهادی به صورت زیر تعریف خواهد شد:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^{h}(r_{u,i}-\bar{r}_u)* w_{a,u}}{\sum_{h=1}^{H}|w_{a,h}|} \qquad (7)$$

که در رابطه ۷:

- **a کاربر هدف، u کاربر تأثیرگذار و h کاربران در گروه کاربران مشابه است.**
- **$w_{a,u}$ میزان شباهت کاربر هدف با کاربر تأثیر گذار داده شده u است.**
- **$r_{u,i}$ رتبه اعطا شده توسط کاربر تأثیر گذار برای یک آیتم i است**
- **$\bar{r}u$ میانگین همه رتبه بندی‌های کاربر تاثیرگذار (u) است.**
- **$\bar{r}a$ میانگین همه رتبه بندی‌های کاربر هدف (a) است.**
- **$w_{a,u}$ میزان شباهت بین کاربر هدف و کاربر تأثیرگذار است.**
- **H تعداد کاربران در گروه کاربران مشابه است.**

با توجه به رتبه تخمینی برای کاربر هدف بر اساس رتبه اعطا شده وسط کاربر تأثیرگذار و میزان شباهت به گروه کاربران مشابه، می توان K آیتم برتر که بیشتر مقدار رتبه تخمین را دارند به عنوان توصیه بازدید صفحات وب به کاربر هدف ارائه شود.

۴. نتایج آزمایشات

روش ارائه شده برای پیاده سازی یک سیستم توصیه شخصی مبتنی بر داده های پروفایل کاربر از مجموعه داده بلادرنگ مقیاس بزرگ Yelp است. این مجموعه داده ها حاوی اطلاعات استخراج شده از نمایه های کاربر در مورد محصولات به صورت متنی و رتبه بندی کاربران از محصولات است. کاربران با شناسه های منحصر به فرد در مجموعه داده شناسایی می شوند. روش ارائه شده از این داده ها برای شخصی سازی صفحات وب بر اساس رفتار کاربر استفاده می کند. رفتار کاربر از اطلاعات پروفایل کاربر استنباط می شود و شباهت های بین کاربران بر این اساس محاسبه می شود. جدول ۱ جزئیات مجموعه داده های مورد استفاده را نشان می دهد.

جدول ۱. مشخصات مربوط به مجموعه داده های مورد استفاده

عنوان مجموعه داده	تعداد نمونه ها	تعداد فیلد	عناوین فیلدها
Yelp	10000	10	{business_id, Date, review_id Stars, Text, Type, user_id, Cool, useful, funny}

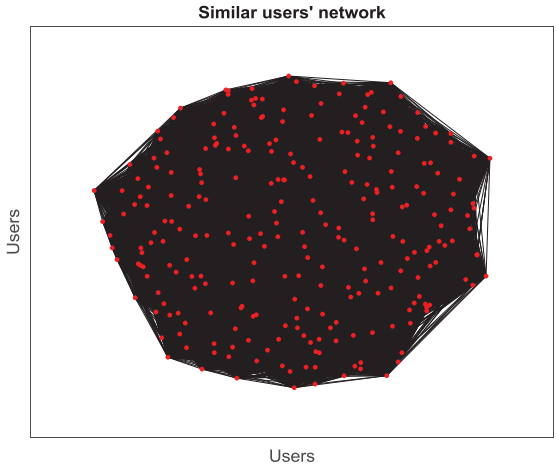
[3] Mean Absolute Error
[4] Root Mean Square Error

در این مجموعه داده ها، داده ها به دو بخش تقسیم می شوند: ۷۰ درصد برای داده های آموزشی و ۳۰ درصد برای داده های آزمایشی. در مجموعه داده های آموزشی Yelp، در کنار رتبه بندی کاربران از محصولات، نظرات کاربران، پروفایل ها و اطلاعات متنی گنجانده شده است. اطلاعات متنی در نمایه‌های کاربر، احساسات و عواطف کاربران را نسبت به محصولات منعکس می‌کند و در ارتباط با رتبه‌بندی محصول، می‌تواند برای تجزیه و تحلیل رفتار کاربر استفاده شود. سیستم‌های توصیه سنتی به اطلاعات متنی در نمایه‌های کاربر توجه نمی‌کنند و برای محاسبات شباهت صرفاً به رتبه‌بندی کاربران متکی هستند. با این حال، اطلاعات متنی در نمایه‌های کاربر حاوی دانش ضمنی ارزشمندی است که می‌تواند منجر به ارزیابی دقیق‌تر شباهت کاربر شود. بنابراین، روش پیشنهادی بر تجزیه و تحلیل اطلاعات در پروفایل کاربر متمرکز است. اطلاعات متنی در نمایه های کاربر به صورت پیوسته است و قبل از استخراج شباهت بر اساس معیار تشابه کسینوس از نظرات کاربران، نیاز به پیش پردازش از جمله حذف کلمات توقف دارد.

مقدار شباهت برای هر جفت کاربر، مقداری در محدوده [۰، ۱] است که نشان دهنده سطح شباهت بین دو کاربر بر اساس اطلاعات پروفایل کاربری آنها است. مقدار بالاتر نشان دهنده شباهت بیشتر بین دو کاربر از نظر اطلاعات پروفایل کاربری آنها است. بنابراین در روش پیشنهادی آستانه ای برای یافتن کاربران مشابه تعریف می شود که به صورت رابطه زیر بیان می شود:

$$\alpha = \frac{sim(i,j)}{\frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n} sim(i,j)}{n}} \geq 1 \qquad (۸)$$

در رابطه ۸، i و j اندیس کاربران و n تعداد کل کاربران در مجموعه داده آموزشی است. پس از محاسبه شباهت بین نظرات کاربران، در روش ارائه شده ماتریس مجاورتی شبکه کاربران مشابه با توجه به آستانه اشاره شده، ایجاد می شود. در ماتریس مجاورتی شباهت کاربران، کاربرانی که نظرات مشابهی دارند، مقدار ۱ و کاربران غیر مشابه یا با مشابهت نظرات کمتر مقدار صفر دریافت می کنند. از این رو می توان شبکه کاربران مشابه را در شکل ۳ مشاهده کرد .



شکل ۳. شبکه کاربران مشابه

همانطور که در شکل ۳ می توان دید، شبکه کاربران مشابه با توجه تشابه نظرات کابران و آستانه مشخص شده تشکیل شده است. حال در این مرحله از روش پیشنهادی به یافتن کاربران تأثیرگذار با استفاده از الگوریتم ژنتیک خواهیم پرداخت.

یافتن کاربران تأثیرگذار در شبکه کاربران مشابه یک مسئله مهم در حوزه تحلیل شبکه‌های اجتماعی و سامانه‌های توصیه‌گر است. در این رویکرد، ابتدا با استفاده از معیار شباهت بین کاربران بر اساس اطلاعات پروفایل آن‌ها، یک شبکه کاربران مشابه ساخته می‌شود. سپس با تعریف یک معیار تأثیرگذاری برای کاربران در این شبکه، کاربرانی که بیشترین تأثیر را دارند شناسایی می‌شوند. برای یافتن کاربران تأثیرگذار، معیارهایی مانند مرکزیت درجه گره، میانگین نزدیک ترین مسیر بین کاربران و سایر معیارهای تاپولوژیکی دیگر در نظر گرفته می‌شود. این معیارها بر اساس ارتباطات موجود در شبکه کاربران مشابه، تعیین می‌کنند که کدام کاربران بیشترین تأثیر را بر رفتار و انتخاب‌های دیگر کاربران دارند. با شناسایی کاربران

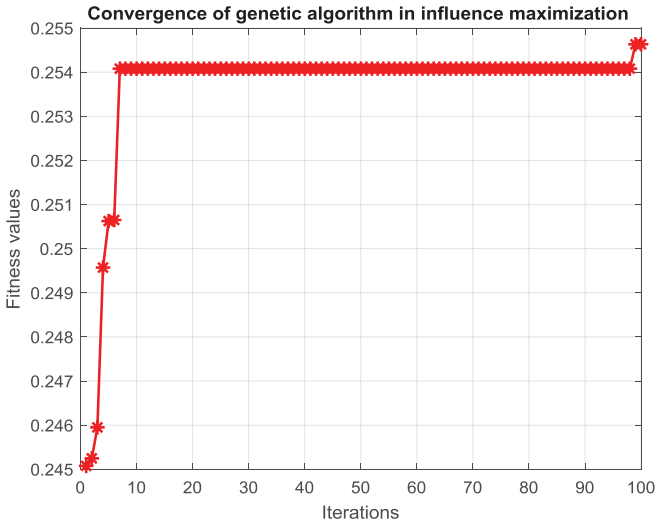
تأثیرگذار، امکان ارائه توصیه‌های بهتر و دقیق‌تر در سامانه‌های توصیه‌گر و همچنین برنامه‌های بازاریابی و تبلیغاتی فراهم می‌شود. این امر می‌تواند منجر به افزایش رضایت کاربران، افزایش نرخ برخورد توصیه های مفید شود.

در روش پیشنهادی از معیار مرکزیت درجه به منظور ضریب تابع تناسب برای یافتن کاربران تأثیرگذار استفاده شده است. این معیار از معروف ترین معیارها در روابط بین کاربران است و می تواند به سایر معیارهای مبتنی بردرجه گره ها تعمیم یابد. در جدول ۲ تعداد متفاوتی از کاربران تأثیرگذار با توجه به معیار مرکزیت درجه و حداکثر درجه تأثیر بر اساس الگوریتم ژنتیک نشان داده شده است.

جدول ۲ کاربران تأثیر گذار بر اساس الگوریتم ژنتیک

تعداد کاربران تأثیر گذار	اندیس کاربران تأثیرگذار	مقدار تابع تناسب	حداکثر درجه تأثیر
۵	261 207 169 253 152	254.6329	11085
۶	169 253 38 125 261 207	253.2676	13159
۷	253 63 169 69 207 261 38	252.8250	13186
۸	75 261 63 38 207 169 253 69	251.9010	17041
۹	69 207 38 261 63 169 155 253 125	250.7974	18416
۱۰	169 28 261 125 129 69 245 253 126 207	249.6514	21785

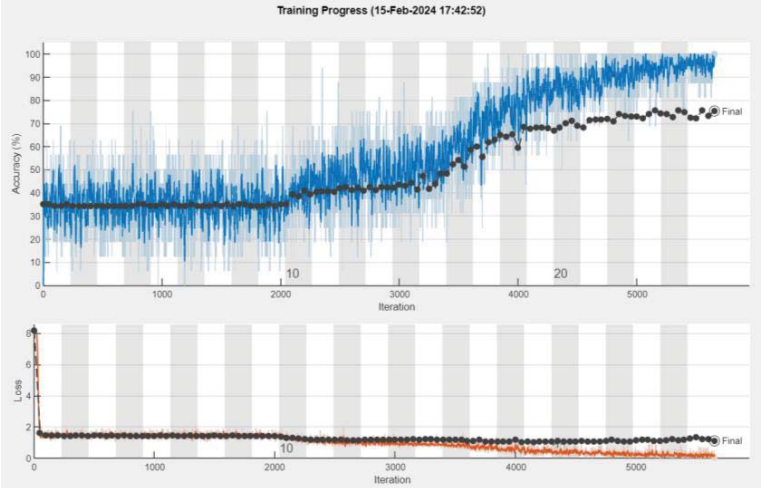
با توجه به جدول ۲ می توان دید با افزایش تعداد کاربران تأثیر گذار، نرخ تأثیر نیز طبیعتاً افزایش می یابد، این در حالی است که مقدار تابع تناسب که ترکیبی از تعداد کاربران و گسترش نفوذ این کاربران است، در حال کاهش است. این پدیده نشان می دهد که افزایش تأثیر با افزایش تعداد کاربران تأثیرگذار، رفتار برعکسی دارد. از این رو کمترین تعداد کاربران تأثیرگذار که بیشترین مقدار تابع تناسب را داشته باشد می تواند به عنوان مرجعی برای کاربران تأثیرگذار انتخاب شد. در شکل ۴ همگرائی الگوریتم ژنتیک به سمت نقطه بهنیه بر اساس کمترین تعداد کاربران تأثیرگذار با بیشترین مقدار تابع تناسب نشان داده شده است.



شکل ۴. همگرائی الگوریتم ژنتیک در حداکثر سازی تأثیر در شبکه کاربران مشابه

با توجه به شکل ۴ و جدول ۲ می توان دید کاربران تأثیر گذار به عنوان کاربرانی با بیشترین مقدار ارتباطات در شبکه که در این مقاله به عنوان کاربرانی با بیشترین شباهت بین سایر کاربران بر اساس الگوریتم ژنتیک پیدا شده اند. حال هر یک از کاربران با توجه به کاربران مجاور خود که رتبه بندی یکسانی در مورد صفحات دارند، به عنوان گروه کاربری مشابه در نظر گرفته می شود. نظرات کاربران در گروه کاربری مشابه به عنوان ویژگی ها ورودی به شبکه خودرمزگذار عمیق و مقدار رتبه اعطا شده به صفحات وب به عنوان متغیر هدف که بایستی در مورد کاربران جدید تخمین زده شود، در نظر گرفته می شود.

در روش ارائه شده به منظور تخمین دقیق کاربران رتبه بندی کاربران جدید از شبکه خودرمزگذار عمیق استفاده شده است. شبکه‌های خود رمزگذار عمیق ارائه شده با توجه به کاربردهای متنوع، گزینه مناسبی برای یافتن ویژگی های مفید از میان نظرات در ارتباط با کلاس های متناظر با هر مقدار رتبه است. در واقع با توجه به این که مقادیر رتبه بندی از ۱ تا ۵ به صورت گسسته توزیع شده است، شبکه خودرمزگذار عمیق سعی دارد وزن تعلق هر نظر به هر یک از کلاس ها را آموزش ببیند. بدین ترتیب کاربران جدید را می توان بر اساس نظرات آن ها در یکی از کلاس ها دسته بندی کرد و مقدار رتبه بندی کاربر جدید را تخمین زد. در شکل ۵ روند آموزش شبکه عصبی ارائه شده بر روی مجموعه داده های آموزشی Yelp نشان داده شده است.



شکل ۵. روند آموزش شبکه عصبی ارائه شده بر روی مجموعه داده Yelp

با توجه به شکل ۵ می‌توان دید شبکه‌های عصبی ارائه شده در ابتدا در تنظیم وزن‌ها و آموزش ویژگی های مربوط به کاربران مقدار خطای بالایی داشته که با افزایش تعداد تکرارها این خطا به مقدار کمتری میل کرده است.

۴-۲ ارزیابی عملکرد روش پیشنهادی

کیفیت عملکرد یک سیستم توصیه را می توان با استفاده از معیارهای مختلف ارزیابی کرد. نوع متریک مورد استفاده بستگی به روش فیلتر و رویکرد به کار گرفته شده دارد. دقت یک سیستم توصیه بر اساس دقت توصیه های ارائه شده به کاربران تعیین می شود. در این تحقیق با توجه به ترکیب روش های مورد استفاده، از دو معیار برای ارزیابی دقت روش پیشنهادی استفاده شده است. متریک اول، معیار دقت آماری برای ارزیابی دقت رتبه‌بندی پیش‌بینی‌شده برای کاربران در روش فیلتر مشارکتی و دقت در ارائه توصیه‌ها به کاربران مشابه در میان کاربران جدید است. معیار دوم عملکرد شبکه های عصبی عمیق را در پیش بینی رتبه بندی محصولات بر اساس اطلاعات پروفایل کاربر و مقایسه آن با سایر روش های طبقه بندی ارزیابی می کند. در ادامه، ابتدا روش پیشنهادی را بر اساس دقت آماری ارزیابی کرده و سپس دقت روش های طبقه بندی را با هم مقایسه می کنیم.

۴-۲-۱ ارزیابی عملکرد روش ارائه شده بر اساس دقت آماری

همانطور که گفته شد، معیارهای دقت آماری مستقیماً دقت روش فیلتر مشارکتی را با مقایسه رتبه‌بندی پیش‌بینی‌شده برای کاربران با رتبه‌بندی واقعی آنها ارزیابی می‌کنند. میانگین خطای مطلق[۵](MAE)، خطای میانگین مربعات[۶](RMSE) معمولاً به عنوان معیارهای دقت آماری در حوزه فیلترمشارکتی استفاده می شود. MAE به صورت میانگین اختلاف مطلق بین رتبه پیش‌بینی شده و امتیاز واقعی برای تمام کاربران محاسبه می‌شود. به عبارت دیگر، MAE برابر با میانگین از مقادیر مطلق اختلاف بین رتبه پیش‌بینی شده و امتیاز واقعی است. از سوی دیگر، RMSE نیز به صورت میانگین اختلاف مربعاتی بین رتبه پیش‌بینی شده و امتیاز واقعی برای تمام کاربران محاسبه می‌شود و سپس جذر دوم از این مقدار

[5] Mean Absolute Error
[6] Root Mean Square Error

محاسبه می‌شود. این دو معیار به ما اندازه‌گیری دقیقی از عملکرد روش فیلتر مشارکتی در پیش‌بینی امتیاز کاربران ارائه می‌دهند. قطعاً هرچه ارزش این معیارها برای کاربران کمتر باشد، دقت روش فیلتر مشارکتی بالاتر است. به عبارت دیگر، هرچه رتبه پیش‌بینی‌شده به رتبه واقعی نزدیک‌تر باشد، ارزش این معیار کمتر می‌شود. رابطه این معیارها به صورت زیر محاسبه می‌شود:

$$MAE = \frac{1}{N}\sum_{u,i}|p_{u,i} - r_{u,i}| \quad (۹)$$

$$RMSE = \sqrt{\frac{\sum_{u,i}(p_{u,i}-r_{u,i})^2}{N}} \quad (۱۰)$$

که در آن $P_{u,i}$ به عنوان رتبه پیش‌بینی شده برای کاربر u روی محصول i و $r_{u,i}$ رتبه واقعی کاربر u روی محصول i و N کل تعداد رتبه‌بندی در مجموعه گزینه‌ها تعریف شده است.

از سوی دیگر می توان گفت کاهش (MAE) و (RMSE) موجب افزایش نرخ برخورد نیز خواهد شد. نرخ برخورد یکی از معیارهای ارزیابی عملکرد سیستم‌های توصیه‌گر است که نشان می‌دهد که چه مقدار از آیتم‌های مورد علاقه کاربران توسط سیستم توصیه‌گر شناسایی شده‌اند. به عبارت دیگر، نرخ برخورد نسبت تعداد آیتم‌های مورد علاقه کاربران که به درستی توسط سیستم توصیه‌گر شناسایی شده‌اند به تعداد کل آیتم‌های مورد علاقه کاربران است. برای محاسبه نرخ برخورد، ابتدا باید تمام آیتم‌های مورد علاقه کاربران را در مجموعه داده مشخص کرده و سپس آیتم‌هایی که توسط سیستم توصیه‌گر پیشنهاد شده‌اند و در آیتم‌های مورد علاقه کاربران وجود دارند را شناسایی کرد. سپس با محاسبه نسبت تعداد آیتم‌هایی که به درستی توسط سیستم توصیه‌گر شناسایی شده‌اند به تعداد کل آیتم‌های مورد علاقه کاربران، می‌توان نرخ برخورد را محاسبه کرد. این معیار همراه با سایر معیارهای ارزیابی می‌تواند به ما کمک کند تا عملکرد کلی سیستم توصیه‌گر را بهتر درک کنیم و بهبودهای لازم را اعمال کنیم. در جدول ۵ مقادیر مربوط به MAE و RMSE و نرخ برخورد برای سناریوهای مختلفی از تعداد کاربران تأثیر گذار نشان داده شده است.

جدول ۵- MAE و RMSE و نرخ برخورد با توجه به تعداد کاربران تأثیرگذار

Seed Size	MAE	RMSE	Hit Rate
5	0.0028	0.0029	93.96
6	0.0025	0.0026	94.19
7	0.0026	0.0027	94.87
8	0.0024	0.0025	95.31
9	0.0025	0.0026	95.27
10	0.0024	0.0025	94.23

همانطور که در جدول ۵ قابل مشاهده است، افزایش تعداد کاربران تأثیرگذار در شبکه کاربران مشابه بر اساس نظرات، تأثیر قابل توجهی برمقدار MAE و RMSE و نرخ برخورد ندارد. این نشان می دهد روش پیشنهادی توانایی تخمین بالقوه برای کاربران جدید را در شرایط مختلفی دارد و آموزش بر روی نظرات کاربران با استفاده از شبکه خودرمزگذار عمیق در حد مناسبی انجام گرفته است.

۴-۲-۲ ارزیابی عملکرد روش ارائه شده بر اساس دقت توصیه‌های ارائه شده

پس از محاسبه دقت آماری بر اساس خطای مطلق و میانگین مربعات برای کاربران تست، در این مرحله نتایج به دست آمده از شبکه عصبی عمیق برای کاربران در مجموعه داده تست بر اساس رتبه‌های تخمینی برای کاربران تست سنجیده می‌شود. پارامترهای ماتریس آشفتگی[۷] یعنی True Positive (TP)، True Negative (TN)، False Positive (FP) و False Negative (FN)، برای ارزیابی عملکرد یک شبکه عصبی عمیق در مرحله تست استفاده می‌شوند. TP نشان دهنده تعداد کاربرانی است که به درستی توسط شبکه عصبی عمیق تشخیص داده شده‌اند و رتبه اعطایی به آن‌ها درست است. TN نشان دهنده تعداد کاربرانی است که به درستی توسط شبکه عصبی عمیق تشخیص داده نشده‌اند و رتبه اعطایی به آن‌ها نادرست است. FP نشان دهنده تعداد کاربرانی است که به اشتباه توسط شبکه عصبی عمیق تشخیص داده شده‌اند و رتبه اعطایی به آن‌ها نادرست است. FN نشان دهنده تعداد کاربرانی است که به اشتباه توسط شبکه عصبی عمیق تشخیص داده نشده‌اند و رتبه اعطایی به آن‌ها درست است. این پارامترها با هم برای ارزیابی کارایی شبکه عصبی عمیق و تصمیم‌گیری در مورد بهبود آن استفاده می‌شوند. در جدول ۶ ماتریس

[7] Confusion Matrix

آشفتگی مربوط به شبکه‌های خود رمزگذار عمیق در مقایسه با سایر روش های طبقه بندی مورد استفاده در این مقاله، برای کاربران تست نشان داده شده است.

جدول ۶. پارمترهای ماتریس آشفتگی مربوط به روش های طبقه بندی

Seed Size	Classifier	TP	FP	TN	FN	TPR	TNR
5	Autoencoder	2699	165	61	75	94.5	44.85
	Random forest	2649	222	4	125	92.52	13.79
	Logistic Regression	2689	220	6	85	92.43	6.59
	Decision tree	2576	170	56	198	93.8	22.04
6	Autoencoder	2690	131	84	95	95.35	53.07
	Random forest	2654	222	4	120	92.54	16.66
	Logistic Regression	2619	218	8	155	92.31	4.9
	Decision tree	2572	168	60	240	92.86	21.27
7	Autoencoder	2729	150	76	45	94.79	62.81
	Random forest	2649	224	2	125	92.46	7.4
	Logistic Regression	2555	214	12	219	92.27	5.19
	Decision tree	2568	163	63	206	94.01	23.42
8	Autoencoder	2630	137	89	144	95.04	38.19
	Random forest	2610	260	6	124	92.59	20
	Logistic Regression	2304	194	32	470	92.23	6.37
	Decision tree	2615	170	56	159	93.89	26.04
9	Autoencoder	2677	138	88	97	95.08	47.56
	Random forest	2653	220	6	121	92.6	22.22
	Logistic Regression	2583	215	11	191	92.31	5.44
	Decision tree	2588	161	65	186	94.14	25.89
10	Autoencoder	2680	144	82	94	94.9	46.59
	Random forest	2644	223	3	130	92.48	9.09
	Logistic Regression	2558	207	19	216	92.51	8.08
	Decision tree	2560	162	64	214	94.04	23.02

با توجه به جدول ۶ می توان دید شبکه های خودرمزگذار عمیق پیشنهادی بر اساس آموزش عمیق بر روی نظرات کاربران با توجه به تعداد کاربران تأثیرگذار در سناریوهای مختلف، مقدار مثبت صحیح(TP)، منفی صحیح(TN) و نرخ مثبت صحیح (TPR) و نرخ منفی صحیح (TNR) بالاتری نسبت به سایر روش های طبقه بندی توانسته به دست آورد. از این رو میتوان به استفاده مناسب از یادگیری عمیق در راستای تخمین نظرات کاربران جدید در روش پیشنهادی اشاره کرد.

۴-۲-۳ دقت روش‌های طبقه‌بندی

در ارزیابی روش‌های طبقه‌بندی، معیارهای مختلفی برای تجزیه و تحلیل عملکرد مدل‌های طبقه‌بندی و پیش‌بینی رتبه‌بندی نمونه‌های جدید وجود دارد. این معیارها از طریق مقایسه رتبه‌بندی پیش‌بینی شده با رتبه‌بندی واقعی نمونه‌ها برای ارزیابی عملکرد مدل استفاده می‌شوند. معیارهای ارزیابی مختلفی برای تحلیل و بررسی عملکرد روش‌های طبقه‌بندی و پیش‌بینی رتبه‌بندی وجود دارند، از جمله دقت، حساسیت، صحت، معیار f و غیره. بر اساس پارامترهای ماتریس آشفتگی می‌توان معیارهای ارزیابی زیر را تعریف کرد.

$$Accurace = \frac{\mathrm{TP} + \mathrm{TN}}{TP+TN+FP+FN} \quad (۱۱)$$

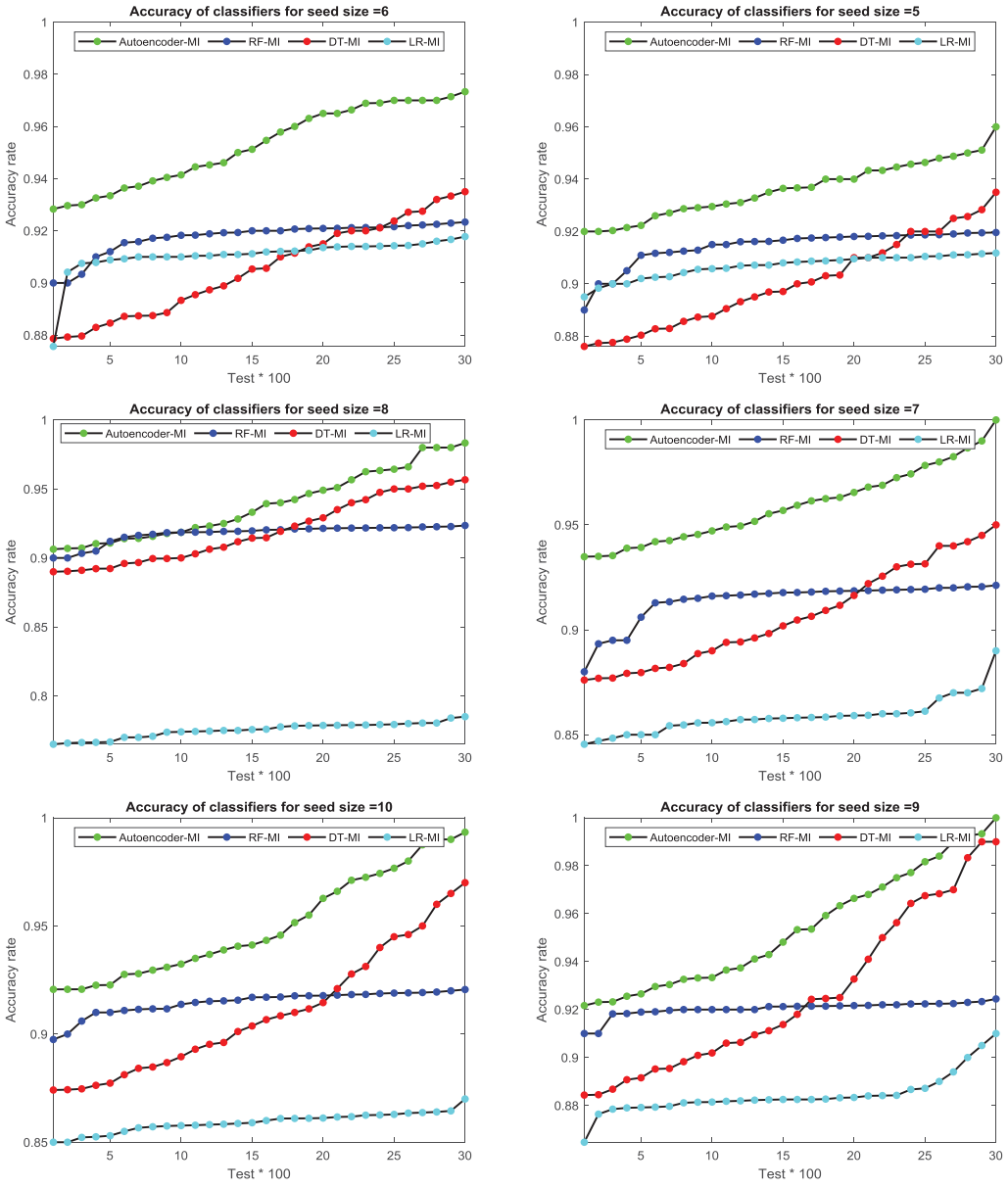
$$Recall = \frac{\mathrm{TP}}{TP+FN} \quad (۱۲)$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (۱۳) F -$$

$$measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (۱۴)$$

معیارهای ارزیابی فوق، به عنوان معیارهایی که به طور خاص برای ارزیابی عملکرد یک روش طبقه‌بندی رتبه‌بندی ارائه می‌شوند، اهمیت بیشتری دارند. این معیارها به عنوان استانداردهایی مورد قبول قرار می‌گیرند و بر اساس آن‌ها می‌توان عملکرد روش ارائه شده را با سایر روش‌ها در این زمینه

مقایسه کرد. در روش ارائه شده، علاوه بر شبکه‌های خود رمزگذار عمیق، از سه روش معروف جنگل تصادفی (RF)، رگرسیون لجستیک (LR) و درخت تصمیم (DT) برای پیش‌بینی رتبه‌بندی کاربران تست استفاده می‌شود و عملکرد آن‌ها با شبکه‌های خود رمزگذار عمیق مقایسه می‌شود. این مقایسه می‌تواند به تحلیل و بررسی عملکرد روش ارائه شده کمک کند و نقاط قوت و ضعف آن را مشخص کند. از معیارهای ارزیابی فوق می‌توان برای انتخاب بهترین روش طبقه‌بندی رتبه‌بندی استفاده کرد و بهبود عملکرد مدل‌ها کمک کرد. در شکل ۶ نمودار دقت روش های طبقه بندی برای سناریوهای مختلف با تعداد کاربران تأثیرگذار متفاوت نشان داده شده است.

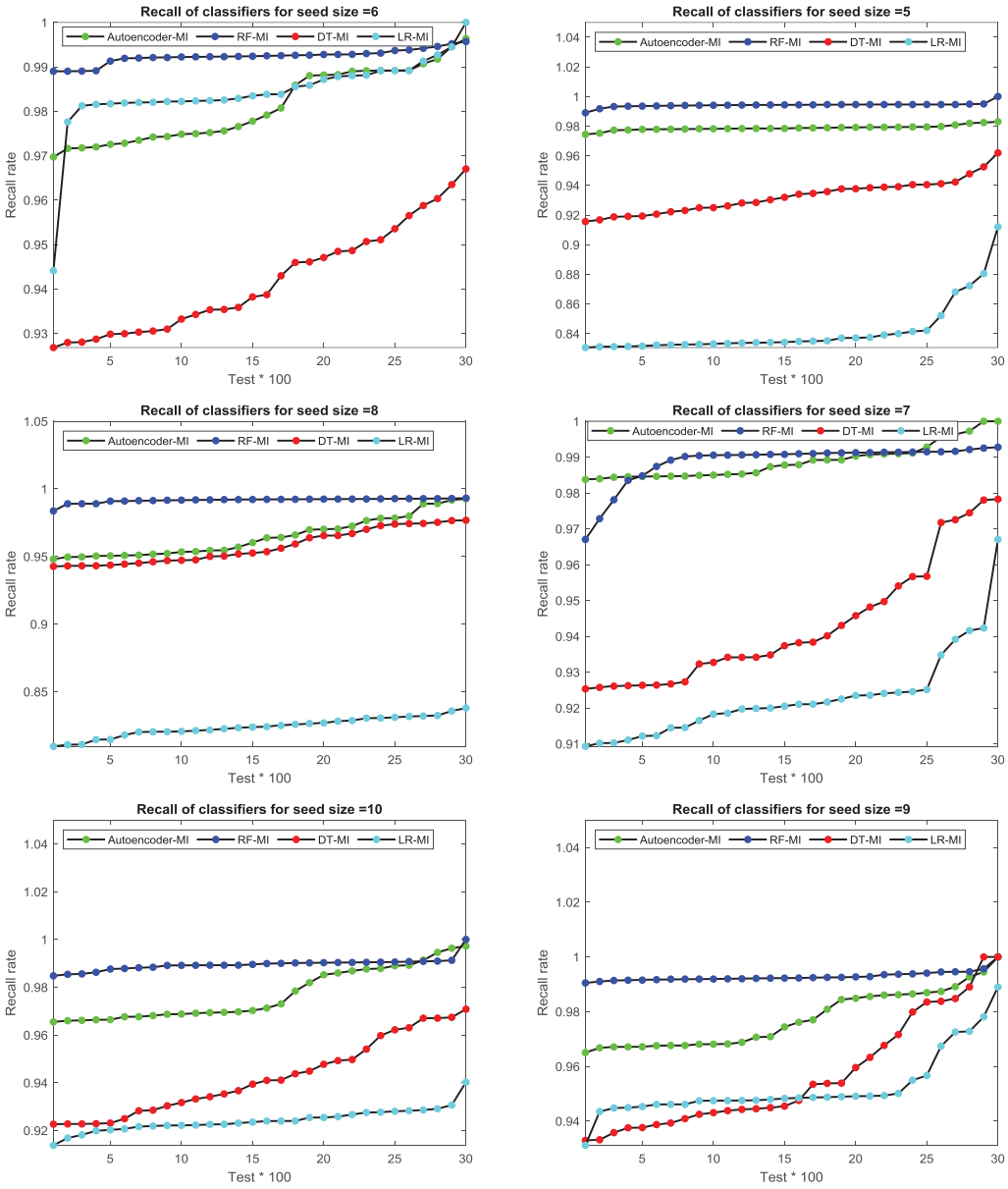


شکل ۶. نمودار معیار دقت در روش‌های طبقه‌بندی

همان‌طور که در شکل ۶ می‌توان دید، روش پیشنهادی با بهره گیری از شبکه های خودرمزگذار عمیق در سناریوهای مختلفی با توجه به تعداد کاربران تأثیرگذار برای تخمین رتبه بندی کاربران جدید، توانسته است مقدار دقت بالاتری نسبت به سایر روش های طبقه بندی به دست آورد . با توجه به شکل ۶ می‌توان دریافت تعداد کاربران تست ۳۰۰۰ کاربر می‌باشد که مقادیر دقت پیش‌بینی از هر ۱۰۰ تست یک بار در شکل ظاهر شده

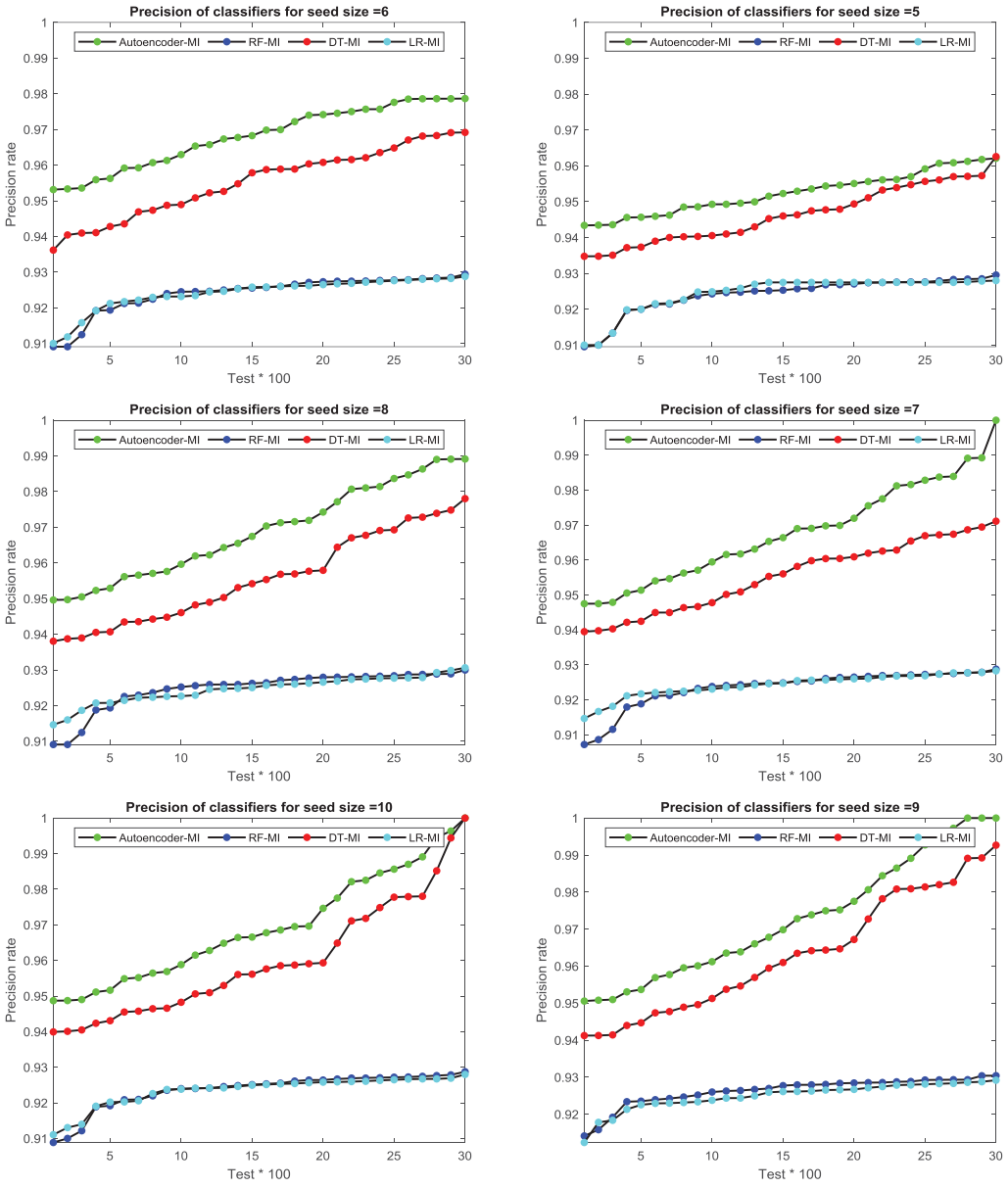
است. این کار برای جلوگیری از ازدحام نقاط بر روی نمودار دقت و سایر نمودارها صورت گرفته است تا وضوح تصاویر مربوط به دقت و سایر معیارهای ارزیابی افزایش یابد.

حساسیت، همچنین به عنوان یادآوری یا نرخ مثبت واقعی شناخته می شود، معیار دیگری است که برای ارزیابی عملکرد یک شبکه عصبی عمیق در وظایف طبقه بندی استفاده می شود. حساسیت نسبت حداکثر رتبه‌های پیش‌بینی‌شده درست را در بین همه حداکثر رتبه‌ها در مجموعه داده آزمایشی اندازه‌گیری می‌کند. این نشان می دهد که مدل تا چه حد قادر است موارد مثبت را به درستی شناسایی کند. حساسیت به عنوان تعداد پیش بینی های مثبت واقعی تقسیم بر مجموع پیش بینی های مثبت درست و منفی کاذب محاسبه می شود. در شکل ۷، نمودارمعیار حساسیت برای روش های طبقه بندی مختلف برای سناریوهای مختلف با تعداد متفاوتی از کاربران تأثیرگذار نشان داده شده است.



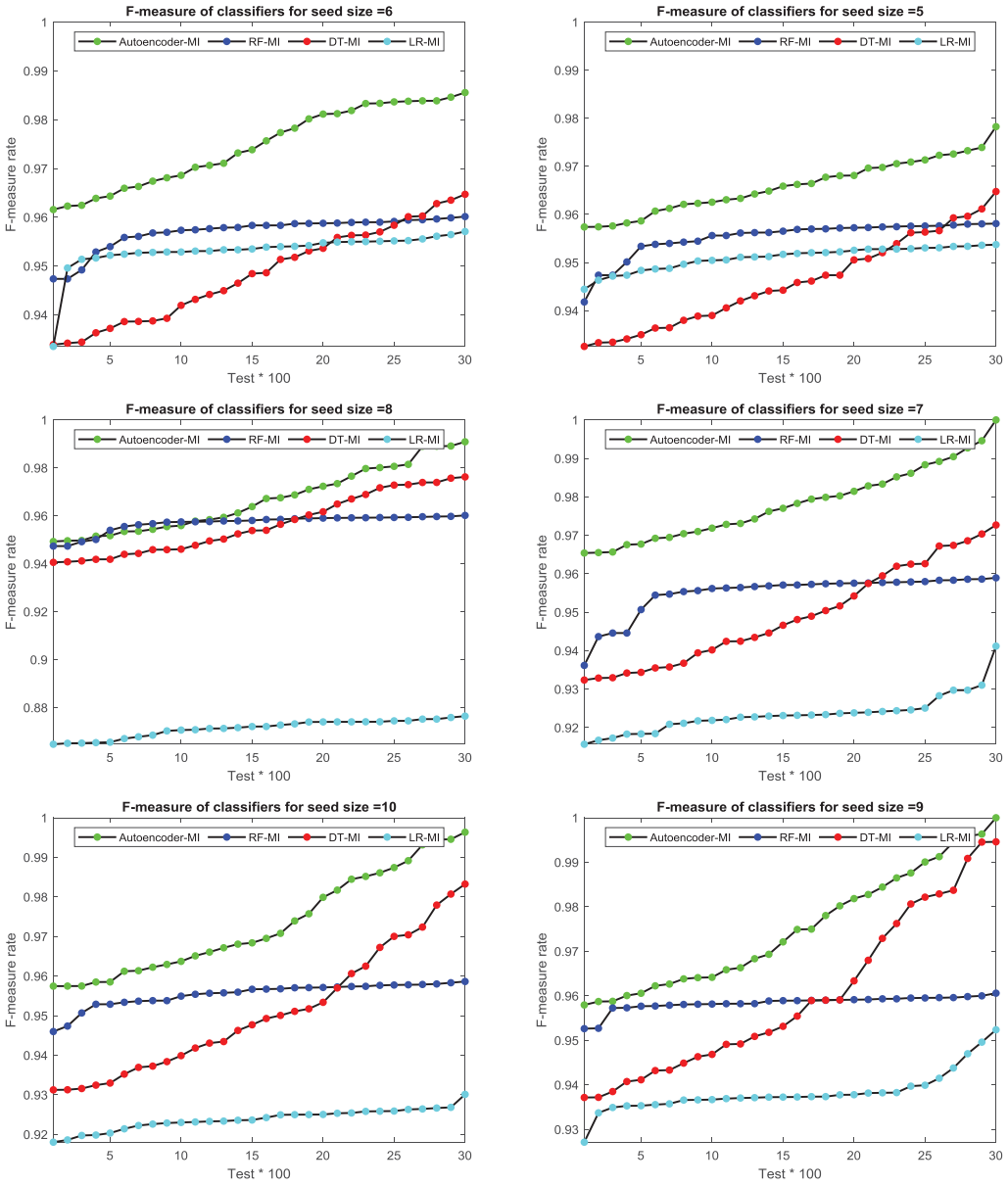
شکل ۷. نمودار معیار حساسیت در روش‌های طبقه‌بندی

همان‌طور که در شکل ۷ می‌توان دید، روش طبقه بندی جنگل تصادفی مقادیر حساسیت بهتری نسبت به شبکه های خودرمزگذار عمیق به دست آورده است. با توجه به این که در معیار حساسیت بیشتر بر روی نظرات در کلاس مثبت که شامل حداکثر رتبه اعطا شده توسط کاربران ، تمرکز شده است، روش طبقه بندی جنگل تصادفی توانسته است عملکرد خوبی داشته باشد. حال آنکه در معیار دقت که عملکرد روش های طبقه بندی را درقبال همه کلاس ها نشان میدهد، شبکه های خودرمزگذار عمیق عملکرد بهتری نسبت به جنگل تصادفی دارد.سومین معیار ارائه شده در ارزیابی صحت است. این معیار، صحت حداکثر رتبه‌های پیش‌بینی‌شده را در بین همه حداکثر رتبه‌های کشف شده در بین حداکثر رتبه‌های کاربران تست با روش‌های طبقه‌بندی اندازه‌گیری می‌کند. در شکل ۸ نمودار معیار صحت برای روش های طبقه بندی مختلف برای سناریوهای مختلف با تعداد متفاوتی از کاربران تأثیرگذار نشان داده شده است.



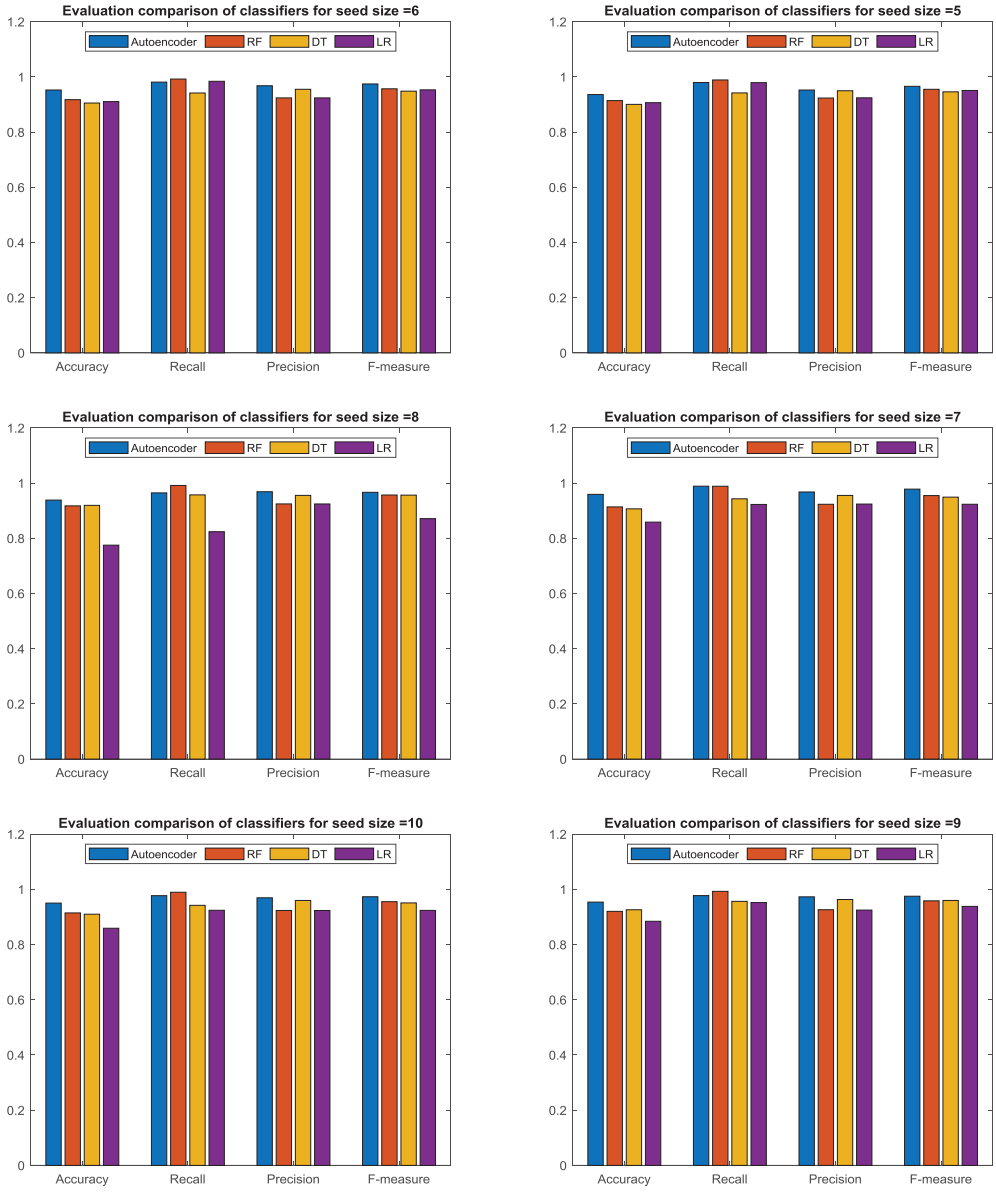
شکل ۸. نمودار معیار صحت در روش‌های طبقه‌بندی

همان‌طور که در شکل ۸ می‌توان دید، روش پیشنهادی با بهره گیری از شبکه های خودرمزگذار عمیق در سناریوهای مختلفی با توجه به تعداد کاربران تأثیرگذار برای تخمین رتبه بندی کاربران جدید، توانسته است مقدار صحت بالاتری نسبت به سایر روش های طبقه بندی به دست آورد. آخرین معیاری که در روش ارائه شده از آن بهره برده ایم، معیار F است که حاصل ترکیب صحت و حساسیت است. این معیار نشان‌دهنده عملکرد کلی مدل ارائه شده به ازای تشخیص رتبه بندی کاربران در مجموعه داده‌های استاندارد است. در شکل ۸ نمودار معیار F برای روش های طبقه بندی مختلف برای سناریوهای مختلف با تعداد متفاوتی از کاربران تأثیرگذار نشان داده شده است.



شکل ۹. نمودار معیارF در روش‌های طبقه‌بندی

همان‌طور که در شکل ۹ می‌توان دید، روش پیشنهادی با بهره گیری از شبکه های خودرمزگذار عمیق در سناریوهای مختلفی با توجه به تعداد کاربران تأثیرگذار برای تخمین رتبه بندی کاربران جدید، توانسته است مقدار معیار F بالاتری نسبت به سایر روش های طبقه بندی به دست آورد.



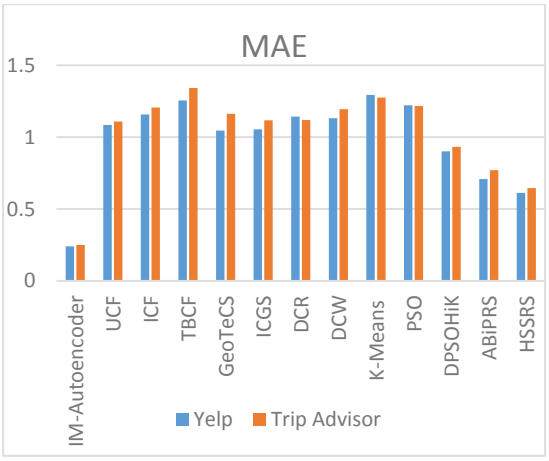
شکل ۱۰. نمودار میله ای مقایسه تجمیعی معیارهای ارزیابی

با توجه به نتایج به دست آمده می‌توان گفت روش ارائه شده با استفاده ازنظرات کاربری، شباهت بین کاربران بر اساس نظرات و گروه کاربران مشابه به کاربران تأثیرگذار با رتبه بندی مشخص، ورودی بسیار مناسبی برای شبکه خودرمزگذار عمیق مهیا کند. شبکه خودرمزگذار عمیق با توجه به مناسب بودن داده‌های ورودی، قدرت آموزش و پیش بینی بالایی بر روی این مجموعه ویژگی‌ها به دست آورده اند. در میان روش‌های طبقه‌بندی، شبکه خودرمزگذار عمیق مناسب ترین نتایج را برای طبقه‌بندی نمونه‌های آموزشی و پیش بینی نمونه‌های رتبه کاربران در مجموعه داده آزمایشی به دست آورده است. در شکل ۱۰ نمودار میله ای مقایسه تفکیکی معیارهای ارزیابی برای روش های طبقه بندی مختلف در سناریوهای مختلف با تعداد کاربران تأثیرگذار متفاوت نشان داده شده است.

بر اساس شکل ۱۰، می توان مشاهده کرد که شبکه های رمزگذار خودکار عمیق ارائه شده، با استفاده از یادگیری عمیق، توانسته اند رتبه بندی محصولات را برای کاربران جدید با بالاترین دقت پیش بینی کنند. روش پیشنهادی با استفاده از لایه‌های آموزشی عمیق و تکمیل فرآیند آموزش در لایه میانی، از نظر معیارهای ارزیابی از سایر روش‌های طبقه‌بندی بهتر عمل کرده است. شبکه‌های رمزگذار خودکار عمیق به دلیل توانایی آنها در ثبت الگوها و روابط پیچیده در داده‌ها، عملکرد برتری در پیش‌بینی رتبه‌بندی محصولات برای کاربران جدید نشان داده‌اند. با استفاده از تکنیک‌های یادگیری عمیق، شبکه‌ها می‌توانند به طور موثری ویژگی‌های زیربنایی داده‌ها را یاد بگیرند و نشان دهند که منجر به پیش‌بینی‌های دقیق‌تر می‌شود. علاوه بر این، فرآیند آموزش در لایه میانی امکان تنظیم دقیق و بهینه سازی مدل را فراهم می کند و قابلیت های پیش بینی آن را بیشتر می کند. به طور کلی، روش پیشنهادی اثربخشی شبکه‌های رمزگذار خودکار عمیق را در رتبه‌بندی وظایف پیش‌بینی نشان می‌دهد و توانایی آن‌ها را در عملکرد بهتر از سایر روش‌های طبقه‌بندی از نظر معیارهای ارزیابی نشان می‌دهد. استفاده از یادگیری عمیق و رویکرد آموزشی منحصربه‌فرد به موفقیت مدل در پیش‌بینی دقیق رتبه‌بندی محصول برای کاربران جدید کمک می‌کند.

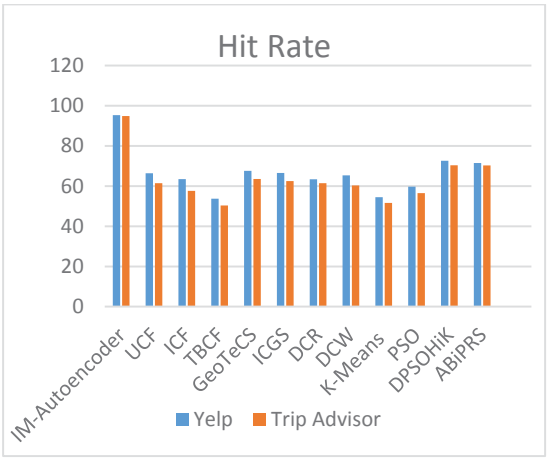
۴-۴ مقایسه روش پیشنهادی با روش های پیشین

پس از اجرا و ارزیابی روش پیشنهادی، به منظور اعتبارسنجی روش پیشنهادی در این مقاله، با روش‌های قبلی از جمله[۱۹] در این زمینه مقایسه می‌شود. بدیهی است که مقایسه نتایج روش های قبلی بر روی همان مجموعه داده در شرایط یکسان انجام خواهد شد. بنابراین، نتایج به‌دست‌آمده از روش پیشنهادی در شرایط یکسان و مجموعه داده‌های مشابه با روش‌های دیگر مقایسه می‌شود تا سطح بهبود حاصل از ایده پیشنهادی در این تحقیق مشخص شود. از آنجایی که روش پیشنهادی یادگیری عمیق و فیلتر مشارکتی را ترکیب می‌کند، معیارهایی مانند میانگین خطای مطلق، نرخ ضربه و دقت تخمین رتبه‌بندی کاربر را می‌توان برای مقایسه مورد استفاده قرار داد. در شکل ۱۱، مقایسه روش پیشنهادی با روش های قبلی از نظر میانگین متریک خطای مطلق نشان داده شده است. با مقایسه روش پیشنهادی با روش‌های قبلی با استفاده از این معیارها، محققان می‌توانند اثربخشی و عملکرد رویکرد پیشنهادی را در مقایسه با تکنیک‌های موجود ارزیابی کنند. استفاده از یادگیری عمیق و فیلتر مشارکتی در روش پیشنهادی امکان ارزیابی جامع عملکرد آن را فراهم می‌کند و هرگونه پیشرفتی را که از طریق ادغام این تکنیک‌ها به دست می‌آید برجسته می‌کند. این مقایسه به ارزیابی تأثیر و مزایای بالقوه روش پیشنهادی در زمینه مشکل خاصی که در مطالعه به آن پرداخته شده است کمک می کند.



شکل ۱۱. مقایسه روش پیشنهادی با روش های پیشین از نظر میانگین خطای مطلق

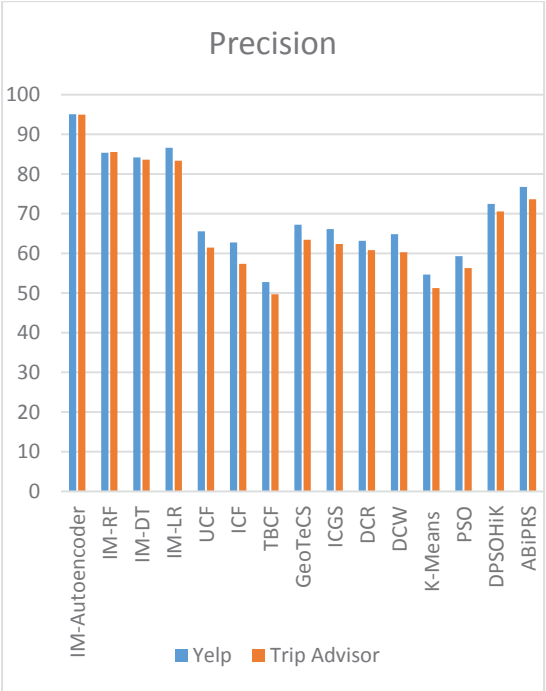
همانطور که در شکل ۱۱ قابل مشاهده است، روش پیشنهادی میانگین خطای مطلق کمتری نسبت به روش های پیشین به دست آورده است. دلیل این امر آموزش عمیق ویژگی های پروفایل کاربر و یافتن کاربران مشابه با دقت بالایی توسط شبکه عصبی خودرمزگذار پیشنهادی است. از این رو انتظار می رود با توجه به مقدار کم خطای مطلق، دقت توصیه های ارائه شده که به عنوان نرخ برخورد نیز در مقالات یاد می شود، در روش پیشنهادی نسبت به سایر روش های موجود، مقدار بهتری داشته باشد. در شکل ۱۲ به مقایسه روش پیشنهادی با روش های پیشین از نظر نرخ برخورد یا دقت توصیه های ارائه شده به کاربران، خواهیم پرداخت.



شکل ۱۲. مقایسه روش پیشنهادی با روش های پیشین از نظر نرخ برخورد

با توجه به شکل ۱۲ می توان دید روش پیشنهادی با تکیه بر آموزش عمیق در شبکه های خود رمزگذار پیشنهادی توانسته است، ویژگی های کاربران را به خوبی آموزش ببیند که همین امر موجب کاهش خطای مطلق در ارائه توصیه و افزایش نرخ برخورد یا دقت توصیه های ارائه شده، گردیده است. آنچه در شکل ۱۲ قابل مشاهده است این است که روش پیشنهادی نرخ برخورد بالاتری نسبت به روش های پیشین بر روی مجموعه داده های یکسان به دست آورده است.

با توجه به استفاده از یاد گیری عمیق و سایر روش های طبقه بندی در زمینه تشخیص کاربران مشابه بر اساس اطلاعات پروفایل کاربران، روش پیشنهادی را می توان از لحاظ دقت تشخیص کاربران مشابه با روش های پیشین مقایسه کرد. در روش پیشنهادی علاوه بر شبکه عصبی خودرمزگذار، از سایر روش های طبقه بندی جنگل تصادفی، درخت تصمیمی و رگرسیون لجستیک نیز به منظور تخمین رتبه بندی کاربران جدید استفاده شده است. مقایسه روش پیشنهادی با روش های طبقه بندی دیگر و روش های پیشین در نشریات از نظر دقت یافتن کاربران مشابه در شکل ۱۳ نشان داده شده است.



شکل ۱۳ مقایسه روش پیشنهادی با روش های پیشین از نظر دقت تشخیص کاربران مشابه

همانطور که در شکل ۱۳ نشان داده شده است، روش پیشنهادی با توجه به استفاده از یادگیری عمیق بر روی نظرات کاربران که از پروفایل آن ها استخراج می شود توانسته است دقت بالاتری در راستی تخمین نظرات کاربران نسبت به سایر روش های طبقه بندی و روش های پیشین به دست آورد.

۵. نتیجه گیری و کارهای آتی

سیستم‌های توصیه‌گر صفحات وب در ارائه پیشنهادات مرتبط با محتوای وب‌سایت‌ها و صفحات اینترنتی به کاربران، به منظور کاهش سربار اطلاعاتی و افزایش تجربه کاربری مورد استفاده قرار می‌گیرند. این سیستم‌ها از تحلیل نظرات و بازخوردهای کاربران به صفحات وب استفاده می‌کنند تا الگوهای رفتاری و ترجیحات کاربران را شناسایی کرده و پیشنهادات مناسبی ارائه دهند. شناسایی کاربران مشابه و ایجاد شبکه کاربران مشابه در این سیستم‌ها اهمیت بالایی دارد. با تشکیل این شبکه‌ها، می‌توان الگوهای رفتاری و ترجیحات کاربران را بهتر شناسایی کرد و پیشنهادات دقیق‌تری ارائه داد. علاوه بر این، کاربرانی که تشابه زیادی با سایر کاربران دارند، می‌توانند به عنوان کاربران بانفوذ در این شبکه کاربران مشابه، نقش مهمی در ارائه توصیه‌های وب داشته باشند. این کاربران می‌توانند با ارائه نظرات و بازخوردهای خود، به بهبود عملکرد سیستم توصیه‌گر و ارائه پیشنهادات دقیق‌تر کمک کنند. به طور کلی، این سیستم‌ها با استفاده از تحلیل داده‌های کاربران و ایجاد شبکه‌های کاربری، بهبود تجربه کاربری، افزایش تعامل کاربران با محتواهای وب و کاهش سربار اطلاعاتی کمک می‌کنند. این رویکرد بهبود عملکرد سیستم‌های توصیه‌گر و ارتقاء تجربه کاربری از اهمیت بالایی برخوردار است و می‌تواند به بهبود کارایی وب‌سایت‌ها و جذب کاربران جدید کمک کند.در این مقاله از ترکیب مسئله به حداکثر رساندن تأثیر اجتماعی و فیلتر مشاکتی مبتنی بر شبکه خودرمزگذار عمیق برای تخمین رتبه بندی کاربران در شبکه کاربران مشابه بر اساس نظرات کاربران استفاده شده است. شبکه خودرمزگذار عمیق با استفاده از نظرات کاربران، شباهت بین کاربران بر اساس نظرات و گروه کاربران مشابه، به کاربران تأثیرگذار با رتبه بندی مشخص، و روشی بسیار مناسبی برای ساختاردهی شبکه است. این شبکه با توجه به مناسب بودن داده‌های ورودی، قدرت آموزش و پیش‌بینی بالایی بر روی این مجموعه ویژگی‌ها به دست آورده است. روش پیشنهادی با مقدار MAE، RMSE، Hit Rate و Accuracy به ترتیب برابر با ۰,۰۰۲۴، ۰,۰۰۲۵، ۹۵,۳۱٪ و ۹۵,۰۸ نسبت به سایر روش های پیشین نتایج بهتری به دست آورده است. در میان روش‌های طبقه‌بندی، شبکه خودرمزگذار عمیق مناسب‌ترین نتایج را برای طبقه‌بندی نمونه‌های آموزشی و پیش‌بینی نمونه‌های رتبه کاربران در مجموعه داده

آزمایشی به دست آورده است. این نتایج نشان می‌دهد که استفاده از شبکه‌های عمیق می‌تواند بهبود قابل توجهی در عملکرد و دقت مدل‌های توصیه‌گر و طبقه‌بندی داده‌ها داشته باشد و باعث افزایش تجربه کاربری و کارایی سیستم‌های توصیه‌گر شود.

در راستای ارائه پیشنهادات آتی برای مقاله می توان به استفاده از یادگیری تقویتی عمیق در سیستم های توصیه گر وب بر اساس پروفایل و نظرات کاربران اشاره کرد. یادگیری تقویتی عمیق یک روش یادگیری ماشین است که به وسیله آن مدل های پیچیده و عمیقی از داده ها را برای انجام وظایف خاصی مانند توصیه وب به کاربران ایجاد می کند. در سیستم های توصیه گر وب، این روش می تواند برای بهبود دقت و کیفیت توصیه های ارائه شده به کاربران استفاده شود. یکی از روش های استفاده از یادگیری تقویتی عمیق در سیستم های توصیه گر وب، استفاده از پروفایل و نظرات کاربران است. با استفاده از اطلاعات پروفایل کاربران مانند سلیقه ها، علایق و تاریخچه خرید، می توان مدل های یادگیری تقویتی عمیق را آموزش داد تا بهترین محتواها و محصولات را برای هر کاربر توصیه کند. همچنین، با تحلیل نظرات و بازخوردهای کاربران به محتواها و محصولات، می توان الگوریتم های یادگیری تقویتی عمیق را بهبود داد و بهبودات لازم را اعمال کرد تا توصیه های ارائه شده به کاربران دقیق تر و موثر تر باشد. به طور کلی، استفاده از یادگیری تقویتی عمیق در سیستم های توصیه گر وب می تواند بهبود قابل توجهی در افزایش دقت توصیه داشته باشد.

۶- مراجع

۱. Gao, C., et al., *A survey of graph neural networks for recommender systems: Challenges, methods, and directions.* ACM Transactions on Recommender Systems, 2023. **1**(1): p. 1-51.

۲. Casillo, M., et al., *Context-aware recommender systems and cultural heritage: a survey.* Journal of Ambient Intelligence and Humanized Computing, 2023. **14**(4): p. 3109-3127.

۳. Breit, A., et al., *Combining machine learning and semantic web: A systematic mapping study.* ACM Computing Surveys, 2023.

۴. Wu, S., et al., *Graph neural networks in recommender systems: a survey.* ACM Computing Surveys, 2022. **55**(5): p. 1-37.

۵. Pham, P., et al., *An approach to semantic-aware heterogeneous network embedding for recommender systems.* IEEE Transactions on Cybernetics, 2023.

۶. Sadesh, S., et al, *Automatic Clustering of User Behaviour Profiles for Web Recommendation System.* Intelligent Automation & Soft Computing, 2023. **35**(۳).

۷. Banerjee, S., M. Jenamani, and D.K. Pratihar, *A survey on influence maximization in a social network.* Knowledge and Information Systems, 2020. **62**: p. 3417-3455.

۸. Wu, L., et al. *A neural influence diffusion model for social recommendation.* in *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*. 2019.

۹. Wu, L, et al., *Diffnet++: A neural influence and interest diffusion network for social recommendation.* IEEE Transactions on Knowledge and Data Engineering, 2020. **34**(10): p. 4753-4766.

۱۰. Sun, Y., B. Cautis, and S. Maniu. *Social Influence-Maximizing Group Recommendation*. in *Proceedings of the International AAAI Conference on Web and Social Media*. 2023.

۱۱. Robles, J.F., M. Chica, and O. Cordon, *Evolutionary multiobjective optimization to target social network influentials in viral marketing.* Expert Systems with Applications, 2020. **147**: p. 113183.

۱۲. Castiglione, A., et al., *Cognitive analysis in social networks for viral marketing.* IEEE Transactions on Industrial Informatics, 2020. **17**(9): p. 6162-6169.

۱۳. Jendoubi, S. and A. Martin, *Evidential positive opinion influence measures for viral marketing.* Knowledge and Information Systems, 2020. **62**(3): p. 1037-1062.

۱۴. Molaei, R., K. Rahsepar Fard, and A. Bouyer, *An Improved Influence Maximization Method for Online Advertising in Social Internet of Things.* Big Data, 2023.

۱۵. Venunath, M., et al., *Influence Maximization and Online Social Capital for Entrepreneurs: Leveraging Social Networks to Achieve Business Goals*, in *Social Capital in the Age of Online Networking: Genesis, Manifestations, and Implications*. ۲۳۲۰, IGI Global. p. 128-145.

۱۶. Aghaee, Z., et al., *A survey on meta-heuristic algorithms for the influence maximization problem in the social networks.* Computing, 2021. **103**(11): p. 2437-2477.

۱۷. Singh, R.R., *Centrality measures: a tool to identify key actors in social networks*, in *Principles of Social Networking*. 2022, Springer. p. 1-27.

۱۸. Borkin, D., et al., *Impact of data normalization on classification model accuracy.* Research Papers Faculty of Materials Science and Technology Slovak University of Technology, 2019. **27**(45): p. 79-84.

۱۹. Logesh, R., et al., *Efficient user profiling based intelligent travel recommender system for individual and group of users.* Mobile Networks and Applications, 2019. **24**: p. 1018-1033.