مقاله پژوهشی

ارائه یک سیستم توصیه‌گر وب شخصی‌سازی‌شده بر اساس خوشه‌بندی ترتیبی و خودرمزگذار عمیق

مستوره معینی، علی برومندنیا و منا مرادی

چکیده: حجم اطلاعات انتشاریافته در وب، استفاده از سیستم‌های توصیه‌گر را اجتناب‌ناپذیر کرده است. سیستم‌های توصیه‌گر وب به کاربران پیشنهادهای دقیق و مرتبط بر اساس علاقه‌ها و سلیقه‌هایشان ارائه می‌دهند. این توصیه‌ها می‌توانند به کاربران کمک کنند تا به سرعت به اطلاعات مورد نیاز خود دسترسی پیدا کنند و زمان جستجو را کاهش دهند. در این مقاله یک سیستم توصیه‌گر ترکیبی بر اساس ترکیب خوشه‌بندی ترتیبی فازی و شبکه خودرمزگذار عمیق بر اساس اطلاعات پروفایل کاربری و رتبه‌بندی وب‌سایت‌ها توسط کاربران ارائه شده است. در این سیستم توصیه‌گر، ابتدا کاربران بر اساس شباهت نظرات خود به صورت ترتیبی خوشه‌بندی می‌شوند. سپس رتبه‌بندی جدید برای کاربران با توجه به تابع عضویت فازی پیش‌بینی می‌شود. در نهایت اطلاعات موجود در پروفایل کاربری و رتبه‌بندی جدید کاربران به هر وب‌سایتی به‌منظور پیش‌بینی رتبه‌بندی وب‌سایت‌ها توسط کاربران، به عنوان ورودی شبکه خودرمزگذار عمیق ارائه‌شده مورد استفاده قرار می‌گیرد. در نهایت پس از یافتن کاربران مشابه، اقدام به ارائه توصیه بازدید و شخصی‌سازی صفحه وب کاربران جدید بر اساس وب‌سایت‌های مورد علاقه کاربران مشابه می‌نماید. روش پیشنهادی با توجه به لایه‌های آموزش عمیق و تکمیل فرایند آموزش در لایه میانی در مقایسه با سایر روش‌های طبقه‌بندی از نظر دقت آماری حدود ٤٢٪، نسبت توصیه‌های موفق به توصیه‌های مفید حدود ٤٪ و دقت تشخیص کاربران مشابه حدود ۲۰٪ بهبود داشته است.

کلیدواژه: **سیستم توصیه‌گر، پروفایل کاربری، شبکه‌های خودرمزگذار، فیلتر مشارکتی.**

۱- مقدمه

توسعه‌های اخیر در فناوری همراه با شیوع خدمات آنلاین، توانایی بیشتری برای دسترسی به حجم عظیمی از اطلاعات آنلاین را با سرعت بیشتری فراهم کرده است. کاربران می‌توانند بررسی‌ها، نظرات و امتیازها را برای انواع مختلفی از خدمات و محصولات آنلاین ارسال کنند. با این حال، توسعه‌های اخیر در محاسبات همه‌جا موجب بروز مشکل اضافه بار[1] داده آنلاین شده است. این اضافه بار داده، فرایند یافتن محتوای مرتبط و مفید در اینترنت را پیچیده می‌کند. اما اخیراً توسعه چندین روش با نیازهای محاسباتی کمتر می‌تواند کاربران را به محتوای مرتبط به راحتی و با سرعت بیشتری هدایت کند. به همین دلیل، توسعه سیستم‌های توصیه‌گر در دوران اخیر توجه زیادی را به خود جلب کرده است. به طور کلی، سیستم‌های توصیه‌گر به عنوان ابزارهای فیلترینگ اطلاعات عمل می‌کنند و به کاربران محتوای مناسب و شخصی‌سازی‌شده ارائه می‌دهند. هدف سیستم‌های توصیه‌گر در اصل کاهش تلاش و زمان مورد نیاز کاربر برای جستجوی اطلاعات مرتبط در اینترنت است [۱].

در حال حاضر، سیستم‌های توصیه‌گر در تعداد زیادی از برنامه‌ها استفاده می‌شوند؛ بنابراین ساخت سیستم‌های توصیه‌گر با کیفیت بالا و منحصربه‌فرد برای ارائه توصیه‌های شخصی‌سازی‌شده به کاربران در برنامه‌های مختلف بسیار مهم است. با وجود پیشرفت‌های مختلف در سیستم‌های توصیه‌گر، نسل فعلی سیستم‌های توصیه‌گر نیازمند بهبودهای بیشتری است تا توصیه‌های کارآمدتری را که قابل استفاده در یک دامنه گسترده‌تر از برنامه‌ها هستند، ارائه دهند. سیستم‌های توصیه‌گر کاربردهای متنوعی دارند که نیاز است تحقیقات بیشتر برای بررسی مزایای آنها صورت گیرد [۲].

از پرکاربردترین روش‌های مورد استفاده در سیستم‌های توصیه‌گر در شبکه‌های اجتماعی می‌توان به روش‌های مبتنی بر فیلتر مشارکتی اشاره کرد [۳]. در فیلتر مشارکتی، وظیفه این است که آیتم‌ها را با توجه به ترجیحات و سایر کاربران به یک کاربر فعال توصیه کنیم. در بیشتر موارد، ترجیحات به صورت نمرات ارزیابی عددی بیان می‌شوند [۴].

رویکردهای مبتنی بر فیلتر مشارکتی به دو نوع تقسیم می‌شوند: رویکردهای مبتنی بر حافظه و رویکردهای مبتنی بر مدل. رویکردهای مبتنی بر حافظه با در نظر گرفتن ترجیحات همسایگان، موارد جدیدی را توصیه می‌کنند. آنها از ماتریس ترجیحات مستقیماً برای پیش‌بینی استفاده می‌کنند. در این رویکرد، مرحله اول ساخت یک مدل است. مدل معادل یک تابع است که ماتریس ترجیحات را به عنوان ورودی می‌گیرد.

سپس توصیه‌ها بر اساس تابعی که مدل و پروفایل کاربر را به عنوان ورودی می‌گیرد، صورت می‌گیرد. در اینجا تنها می‌توان به کاربرانی که پروفایل کاربری آنها متعلق به ماتریس ترجیحات است، توصیه کرد. بنابراین برای توصیه به یک کاربر جدید، پروفایل کاربر باید به ماتریس ترجیحات اضافه شود و ماتریس شباهت باید محاسبه مجدد شود که موجب افزایش پیچیدگی محاسباتی خواهد شد [۵].

سیستم‌های مبتنی بر مدل از الگوریتم‌های داده‌کاوی و یادگیری ماشین مختلف برای توسعه یک مدل برای پیش‌بینی امتیاز کاربر برای یک مورد بدون امتیاز استفاده می‌کنند. زمانی که توصیه‌ها محاسبه می‌شوند، آنها به داده‌های کامل وابستگی ندارند، بلکه ویژگی‌ها را از داده‌ها استخراج کرده

این مقاله در تاریخ ۲۵ خرداد ماه ۱۴۰۳ دریافت و در تاریخ ۹ مهر ماه ۱۴۰۳ بازنگری شد.

مستوره معینی، دانشکده کامپیوتر، واحد تهران‌جنوب، دانشگاه آزاد اسلامی، تهران، ایران، (email: st_m_moini@azad.ac.ir).

علی برومندنیا، دانشکده کامپیوتر، واحد تهران‌جنوب، دانشگاه آزاد اسلامی، تهران، ایران، (email: Ali.Broumandnia@iau.ac.ir).

منا مرادی، دانشکده کامپیوتر، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران، (email: mona.moradi@iau.ac).

1. Overload

و مدل محاسبه می‌شود. بنابراین این روش به نام روش مبتنی بر مدل شناخته می‌شود. این تکنیک‌ها نیز برای پیش‌بینی به دو مرحله نیاز دارند: مرحله اول ساخت مدل است و مرحله دوم پیش‌بینی امتیازها با استفاده از یک تابع (f) است که مدل تعریف‌شده در مرحله اول و پروفایل کاربر را به عنوان ورودی می‌گیرد. تکنیک‌های مبتنی بر مدل نیازی به اضافه‌کردن پروفایل کاربر جدید به ماتریس خدمات عمومی قبل از انجام پیش‌بینی ندارند. ما می‌توانیم حتی به کاربرانی که در مدل حضور ندارند توصیه دهیم. سیستم‌های مبتنی بر مدل برای توصیه‌های گروهی بهتر عمل می‌کنند و می‌توانند به‌سرعت یک گروه از موارد را با استفاده از مدل پیش‌آموزش داده‌شده توصیه کنند. دقت این تکنیک به طور قابل توجهی بستگی به کارایی الگوریتم یادگیری دارد که برای ایجاد مدل استفاده می‌شود. تکنیک‌های مبتنی بر مدل قادر به حل برخی از مشکلات سنتی سیستم‌های توصیه‌گر مانند پراکندگی و قابلیت مقیاس‌پذیری با استفاده از تکنیک‌های کاهش بعد و یادگیری مدل هستند [۶].

یک تکنیک ترکیبی مجموعه‌ای از دو یا چند تکنیک است که با هم برای رسیدگی به محدودیت‌های تکنیک‌های توصیه‌کننده فردی به کار می‌روند. ترکیب تکنیک‌های مختلف می‌تواند به روش‌های مختلفی انجام شود. یک الگوریتم ترکیبی ممکن است نتایج به‌دست‌آمده از تکنیک‌های جداگانه را ترکیب کند یا می‌تواند از فیلتر مبتنی بر محتوا در یک روش مشارکتی استفاده کند یا از یک تکنیک فیلتر مشارکتی در یک روش مبتنی بر محتوا استفاده کند. این تکنیک ترکیبی از تکنیک‌های مختلف به طور کلی منجر به افزایش عملکرد و افزایش دقت در بسیاری از برنامه‌های توصیه‌کننده می‌شود [۷] و [۸].

در این مقاله یک سیستم توصیه‌گر ترکیبی بر اساس پروفایل کاربری ارائه شده که در آن از خوشه‌بندی ترتیبی فازی FOCA[1] بر اساس شباهت نظرات کاربران استفاده شده و بر این اساس به پیش‌بینی رتبه‌بندی کاربران جدید با استفاده از شبکه عصبی خودرمزگذار پرداخته است. بر اساس ترکیب این دو روش می‌توان دقت تشخیص کاربران مشابه و دقت پیش‌بینی رتبه‌بندی نظرات سایر کاربران را افزایش داد. روش ارائه‌شده یک رویکرد سلسله‌مراتبی است که در مرحله اول از اطلاعات پروفایل کاربری به استخراج کاربران مشابه در مجموعه داده آموزشی و وزن شباهت کاربران آموزشی استفاده می‌شود. از این اطلاعات به‌منظور خوشه‌بندی ترتیبی فازی کاربران استفاده خواهد شد. در نهایت وزن شباهت کاربران به همراه اطلاعات موجود در پروفایل کاربری به عنوان ویژگی برای کاربران به شبکه عصبی رمزگذاری خودکار ارائه می‌شود تا به‌جای استفاده از وزن اولیه تصادفی برای سوگیری[2]، از وزن شباهت کاربران استفاده شود. در نهایت ویژگی‌های کاربران در پروفایل کاربری و وزن شباهت آنها به عنوان ورودی شبکه عصبی رمزگذار خودکار عمیق ارائه‌شده مورد استفاده قرار می‌گیرد تا رتبه‌بندی اعطایی کاربران جدید به وب‌سایت‌ها پیش‌بینی شود. پس از تعیین رتبه‌بندی اعطایی کاربران جدید به وب‌سایت‌ها، تشابه بین کاربران جدید و کاربران آموزشی محاسبه می‌شود. در نهایت تصمیم برای ارائه توصیه آیتم به یک کاربر بر اساس شباهت بین کاربران و آیتم‌های مورد علاقه کاربران مشابه اتخاذ می‌شود. کاربرانی که تشابه ویژگی‌های آنها بالاست و وزن شباهت بیشتری دارند، می‌توانند به عنوان کاربران مشابه در نظر گرفته شوند و آیتم‌های مورد علاقه آنها که با بالاترین نمره به آنها نظر داده‌اند، به عنوان K آیتم برتر به همدیگر پیشنهاد شوند. روش اجرایی این مقاله به صورت زیر خلاصه شده است:

۱) همجوشی داده‌ها در قالب یک سیستم تصمیم یار
۲) استفاده از خوشه‌بندی ترتیبی فازی به منظور یافتن کاربران مشابه
۳) استفاده از شبکه عصبی رمزگذارخودکار عمیق به منظور پیش‌بینی رتبه‌بندی اعطایی کاربران جدید به وب‌سایت‌ها
۴) محاسبه شباهت بین کاربران بر اساس شبکه‌های عصبی خودرمزگذار عمیق[3] مبتنی بر فیلتر مشارکتی
۵) یافتن K آیتم برتر بر اساس کاربران مشابه در شبکه اجتماعی

ساختار ادامه مقاله به شرح زیر است. در بخش دوم کارهای مرتبط بررسی گردیده و در بخش سوم جزئیات روش پیشنهادی آمده است. در بخش چهارم نتایج پیاده‌سازی و در بخش پنجم نتیجه‌گیری مقاله بیان خواهد شد.

۲- کارهای مرتبط

با توجه به اهمیت سیستم‌های توصیه‌گر، اخیراً کارهای متنوعی در این زمینه ارائه شده که در ادامه به برخی از آنها اشاره خواهیم کرد. در [۹] یک چارچوب توصیه خبری شخصی به نام HYPNER[4] پیشنهاد شده است. HYPNER هر دو روش فیلتر مشترک مبتنی بر فیلتر و فیلتر مبتنی بر محتوا را ترکیب می‌کند. هدف چارچوب پیشنهادی بهبود دقت توصیه‌های خبری از طریق حل مسائل مقیاس‌پذیری به دلیل مجموعه خبری بزرگ، غنی‌سازی نمایه کاربر، نمایش ویژگی‌ها و ویژگی‌های دقیق اخبار و توصیه مجموعه‌ای از اخبار متنوع است. در [۱۰] یک الگوریتم مبتنی بر یادگیری عمیق برای سیستم‌های توصیه‌گر چندمعیاره پیشنهاد شده که در آن رمزگذارهای خودکار عمیق برای بهره‌برداری از روابط غیر پیش پا افتاده، غیرخطی و پنهان بین کاربران با توجه به اولویت‌های چندمعیاره و ایجاد توصیه‌های دقیق‌تر استفاده می‌شوند. در [۱۱] تلاش شده تا یک تکنیک خوشه‌بندی جدید به نام الگوریتم مبتنی بر خوشه‌بندی مرتب (OCA)[5] با هدف کاهش تأثیر شروع سرد و مشکلات پراکندگی داده در سیستم‌های توصیه تجارت الکترونیکی بررسی شود. در [۱۲] یک سیستم توصیه‌کننده مبتنی بر شبکه عصبی کانولوشن با استفاده از ماتریس محصول بیرونی ویژگی‌ها و فیلترهای متقابل کانولوشن پیشنهاد شده است. روش پیشنهادی می‌تواند با انواع مختلف ویژگی‌ها سروکار داشته باشد و تعاملات معنی‌دار مرتبه بالاتر بین کاربران و آیتم‌ها را به تصویر بکشد و به ویژگی‌های مهم وزن بیشتری بدهد. علاوه بر این، می‌تواند مشکل اضافه برازش[6] را کاهش دهد؛ زیرا روش پیشنهادی شامل میانگین جهانی یا حداکثر ادغام به‌جای لایه‌های کاملاً متصل در ساختار است. در [۱۳] از یک رمزگذار خودکار که مدلی قدرتمند در کاهش ابعاد داده، استخراج ویژگی و بازسازی داده‌ها است، برای یادگیری و پیش‌بینی اولویت‌های دانش‌آموز در یک سیستم توصیه آموزش الکترونیکی مبتنی بر فیلترکردن مشارکتی استفاده شده است. در [۱۴] یک سیستم توصیه برای فروشگاه‌های خرده‌فروشی مد، بر اساس رویکرد چند خوشه‌بندی اقلام و نمایه‌های کاربران در فروشگاه‌های آنلاین و فیزیکی پیشنهاد شده است. راه‌حل پیشنهادی بر تکنیک‌های استخراج تکیه می‌کند که امکان

1. Fuzzy Ordered Clustering-Based Algorithm
2. Bias
3. Auto Encoder Deep Neural Network
4. Hybrid Personalized News Recommendation
5. Ordered Clustering-Based Algorithm
6. Overfitting

جدول ۱: مقایسه روش‌های پیشین.

نویسنده و سال انتشار	روش مورد استفاده	ارائه توصیه شخصی	استفاده از فیلتر مشارکتی	تحلیل معنایی پروفایل کاربر	استفاده از یادگیری عمیق
Meghanathan (۲۰۱۶)	شخصی‌سازی شبکه‌های اجتماعی با استفاده از قوانین انجمنی توسعه‌یافته	✓	×	✓	×
Li (۲۰۱۸)	شخصی‌سازی نتایج موتورهای جستجو با شبکه‌های مفهومی فازی اتوماتیک سازگار	✓	×	✓	×
de Araújo (۲۰۱۹)	مدل‌سازی رفتار کاربران شبکه‌های اجتماعی با روش‌های مبتنی بر ساختار توصیه‌گر پویا	✓	✓	✓	×
Xiao (۲۰۲۱)	آموزش آگاهانه کاربر برای توصیه	✓	×	✓	×
Mohanty (۲۰۲۱)	چارچوب صفحه وب مبتنی بر فازی با نمایه کاربر و طراحی هستی‌شناسی	✓	✓	✓	×
Vullam (۲۰۲۳)	سیستم توصیه شخصی‌سازی‌شده چند نماینده	✓	✓	×	×
Bedi (۲۰۲۲)	بررسی معنایی مسیر جستجو و ارائه یک جستجوی سفارشی‌شده	✓	×	✓	×
Naumov (۲۰۱۹)	مدل توصیه مبتنی بر یادگیری عمیق (DLRM)	✓	×	✓	✓
Doonan (۲۰۱۹)	خوشه‌بندی کاربران بر اساس معیار شباهت جدید	✓	×	×	×
Cui (۲۰۲۰)	توصیه جدید مبتنی بر ضریب همبستگی زمانی	✓	✓	✓	×
George (۲۰۲۰)	ساختار توصیه‌گر پویا مبتنی بر کاربرد و محتوا	✓	✓	✓	×
Kong (۲۰۱۹)	با استفاده از داده‌های کاربرد و محتوا و ساختار	✓	×	✓	×
Darvishy (۲۰۲۰)	ارائه توصیه‌های خبری شخصی‌شده ترکیبی	✓	✓	✓	×
Lathabai (۲۰۱۷)	استفاده از قواعد انجمنی و شاخص‌گذاری	✓	✓	×	×
Parish (۲۰۱۸)	ساختار توصیه‌گر پویا مبتنی بر محتوا	✓	×	✓	×
Tóth (۲۰۲۱)	بهبود نتایج جستجوی کاربر به صورت ضمنی	✓	✓	✓	×
Logesh (۲۰۱۹)	رویکرد ترکیبی برای پیش‌بینی توصیه‌های POI متقاعدکننده	✓	✓	✓	×

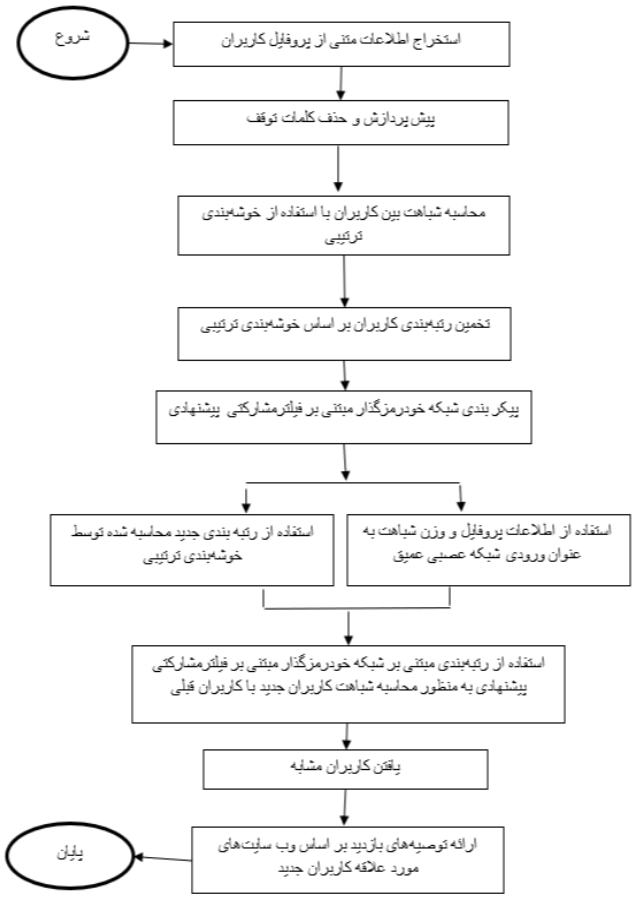
پیش‌بینی رفتار خرید مشتریان جدید را فراهم می‌کند؛ بنابراین مشکلات شروع سرد را که نمونه‌ای از سیستم‌های پیشرفته است، حل می‌کند. در [۱۵] یک سیستم توصیه گردشگری معرفی شده که ترجیحات کاربران را به منظور ارائه توصیه‌های شخصی استخراج می‌کند. برای این منظور، نظرات کاربران در شبکه‌های اجتماعی گردشگری به عنوان منبعی غنی از اطلاعات برای استخراج ترجیحات آنها استفاده می‌شود. سپس نظرات از قبل پردازش‌شده، از نظر معنایی خوشه‌بندی می‌شوند و برای شناسایی ترجیحات یک گردشگر، تحلیل احساساتی می‌شوند. به طور مشابه، همه نظرات جمع‌آوری‌شده کاربران در مورد یک جاذبه برای استخراج ویژگی‌های این نقاط مورد علاقه استفاده می‌شود. در [۱۶] این سیستم توصیه از فیلتر مشارکتی مبتنی بر مدل استفاده می‌کند و محصولات را بر اساس رتبه‌بندی و سابقه خرید قبلی کاربران قدیمی توصیه می‌کند. همچنین کاربران جدید توصیه‌هایی از محصولات جدید، محصولات پرطرفدار و محصولات موجود در فروش دریافت خواهند کرد. کاربران موجود بر اساس محصولات اخیراً مشاهده‌شده، محصولات مکمل و غیره توصیه‌هایی دریافت خواهند کرد. با توجه به این که این تحقیق بر اساس راه‌اندازی یک وب‌سایت تجارت الکترونیکی جدید انجام شده است، در ابتدا هیچ رتبه‌بندی کاربر برای محصولات مختلف وجود ندارد؛ بنابراین در این مورد، توصیه‌ها بر اساس تجزیه و تحلیل خوشه‌بندی متنی توضیحات محصول فیلتر مشارکتی مبتنی بر مدل همراه با خوشه‌بندی متنی، در بهبود دقت و هدف قراردادن انواع کاربران، ارائه شده است. در [۱۷] عملکرد الگوریتم‌های مورد استفاده در سیستم‌های توصیه‌گر بر اساس مجموعه داده‌های مختلف و معیارهای ارزیابی، مقایسه شده است. یک رویکرد یادگیری عمیق به نام GRU4REC و روش‌های ساده‌تر در مقایسه گنجانده شده‌اند. مجموعه داده‌های دنیای واقعی از سه حوزه مختلف در آزمایش‌های این تحقیق گنجانده شده است. تحقیق دیگری با این هدف که به مشتریان خود توصیه‌های شخصی بر اساس آنچه افراد مشابه با نمونه کارهای مشابه دارند ارائه دهد، پیشنهاد شده است. این سیستم از ویژگی‌های مشتری علاوه بر داده‌های نمونه کارهای مشتری استفاده می‌کند. از آنجا که تعداد محصولات پیشنهادی احتمالی کمتر از محصولات موجود در دامنه سیستم‌های توصیه‌گر است و یا اطلاعات محصولات پرتکرار ممکن است در دسترس نباشد، تصمیم گرفته شده که از شبکه‌های بیزی برای مدل‌سازی سیستم استفاده شود. همچنین یک رویکرد مبتنی بر یادگیری عمیق برای ارائه توصیه‌هایی به بالقوه (مشتریان بالقوه) ارائه شده که در آن تنها داده‌های بازاریابی خارجی در زمان پیش‌بینی در دسترس است. در [۱۸]، یک شبکه عصبی کانولوشنال (CNN)[1] با الگوریتم بهینه‌سازی ازدحام ذرات (PSO)[2] یک‌پارچه شده تا توصیه‌های بی‌نظیری ایجاد کند. روش پیشنهادی مبتنی بر نقاط تغییر تمرکز است که شامل پارامترهای غیرمنتظره و مرتبط است. " در جدول ۱ مقایسه روش های پیشین ارائه شده است.

۳- روش پیشنهادی

همان طور که اشاره شد، در این مقاله یک رویکرد توصیه صفحات وب بر اساس پروفایل کاربر در شبکه‌های اجتماعی با استفاده از ترکیب خوشه‌بندی ترتیبی فازی و شبکه‌های رمزگذاری خودکار ارائه شده است. معماری روش پیشنهادی در شکل ۱ آمده است.

روش ارائه‌شده با مجموعه داده بازبینی کاربران آغاز به کار می‌کند. این

1. Convolutional Neural Network
2. Partical Swarm Optimization



شکل ۱: معماری روش پیشنهادی.

مجموعه داده شامل ویژگی‌های کاربران و نیز رأی کاربران به آیتم‌های موجود در مجموعه داده است. در مجموعه داده Yelp [۱۹] نظرات کاربران بر اساس علاقه آنها به مکان‌هایی است که مشاهده کرده‌اند و در مجموعه داده TripAdvisor [۲۰] نظرات کاربران نسبت به محل اقامت یا هتل‌هایی است که در آن بوده‌اند. در روش پیشنهادی در مجموعه داده نظرات کاربران در مورد اقامت خود نیز اشاره شده که از آن به عنوان اطلاعات پروفایل کاربری استفاده می‌شود.

در روش پیشنهادی، در ابتدا با توجه به معیار شباهت کسینوسی، ماتریس شباهت کاربران با توجه به نظرات آنها استخراج می‌شود. تشابه کسینوسی یک معیار شناخته‌شده در بازیابی اطلاعات و مطالعات مرتبط است. این معیار یک سند متنی را به عنوان یک بردار از اصطلاحات مدل‌سازی می‌کند. با استفاده از این مدل، تشابه بین دو سند می‌تواند با محاسبه مقدار کسینوس بین بردارهای اصطلاحی دو سند، مشخص شود. اجرای این معیار می‌تواند بر روی هر دو متن (جمله، پاراگراف یا کل سند) اعمال شود. در مورد موتور جستجو، مقدار تشابه بین پرس‌وجوی[1] کاربر و سندها از بالاترین تا کمترین مرتب می‌شود. امتیاز تشابه بالاتر بین بردار اصطلاحی سند و بردار اصطلاحی پرس‌وجو به معنای وجود تطابق بیشتر بین سند و پرس‌وجو است.

تشابه کسینوسی برای اندازه‌گیری تشابه بین سند و پرس‌وجوی کاربر باید با معنای کلمات سازگار باشد. با این حال، تشابه کسینوسی هنوز نمی‌تواند به‌خوبی معنای معنایی متن را کنترل کند. اجرای اندازه‌گیری تشابه کسینوسی بین دو بردار اصطلاحی از نظر نحوی گاهی نتایج نامعتبری تولید می‌کند. تطبیق نحوی ممکن است نتواند با مشکل تفاوت معنایی مطلوب روبه‌رو شود. برای فرایند بعدی، مانند سیستم بازیابی اطلاعات، ممکن است نتایج نادرستی تولید کند و باعث کاهش کارایی آن شود. رابطه معیار شباهت کسینوسی به صورت زیر بیان می‌شود

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

که در آن q و d به متن‌ها اشاره دارند و t تعداد کلمات در متن و w وزن هر کلمه است. ماتریس شباهت به عنوان ورودی روش خوشه‌بندی ترتیبی ارسال می‌شود. در خوشه‌بندی ترتیبی ابتدا مقادیر مشابهت به ترتیب نزولی مرتب‌سازی می‌شوند. سپس اولین مقدار که نشان‌دهنده بیشترین شباهت بین کاربران است، به عنوان اولین خوشه انتخاب می‌شود. کاربرانی که بر اساس این مقدار با هم شباهت دارند درون اولین خوشه قرار می‌گیرند. پس از آن مقدار دوم در ماتریس مرتب‌شده شباهت به عنوان خوشه دوم در نظر گرفته می‌شود. در آخر برای هر مقدار شباهت در ماتریس همسایگی یک خوشه در نظر گرفته می‌شود. سپس رتبه‌بندی کاربران این خوشه با توجه به تابع عضویت فازی پیش‌بینی می‌شود. تابع عضویت فازی در روش پیشنهادی به صورت زیر تعریف می‌شود

$$FMEM(i,k)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{k} Rate_i - F_j\times Sim_{qd}}{n} \qquad (2)$$

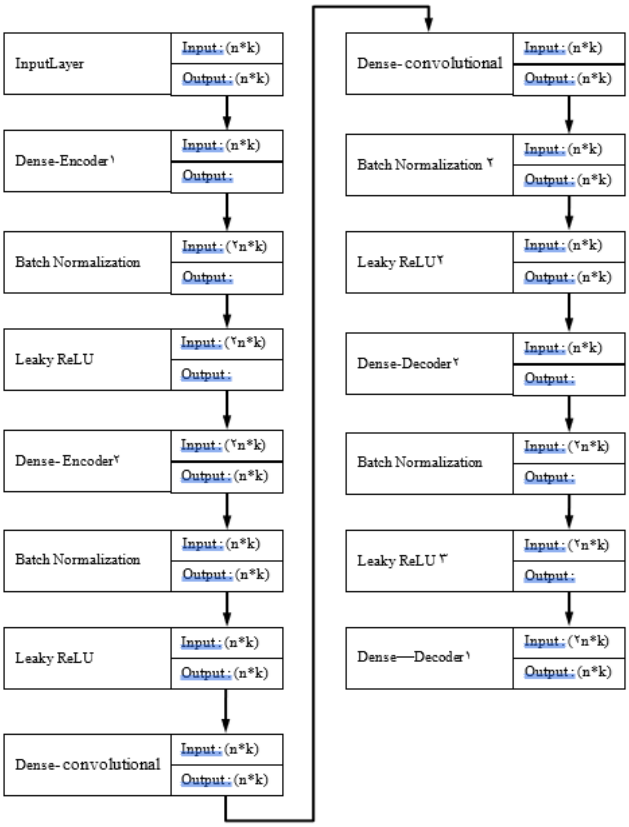
که در آن n تعداد کاربران در خوشه و k به عنوان کلاس یا رتبه است که می‌تواند مقداری بین بازه حقیقی یک تا پنج باشد. $Rate_i$ مقدار واقعی رتبه اعطاشده توسط کاربر و Sim_{qd} شباهت بین دو کاربر است. بر این اساس مقدار عضویت نظرات کاربران به هر یک از ۵ کلاس که شامل رتبه‌های ۱ تا ۵ است، مشخص می‌شود. پس از تعیین مقدار جدی رتبه‌بندی کاربران، این مقادیر به همراه نظرات کاربران به عنوان ورودی شبکه‌های عصبی خودرمزگذار پیشنهادی مورد استفاده قرار می‌گیرند. در ادامه به بررسی معماری شبکه‌های عصبی خودرمزگذار عمیق ارائه‌شده خواهیم پرداخت.

۳-۱ معماری شبکه‌های عصبی خودرمزگذار عمیق ارائه‌شده

در این بخش، یک رمزگذار خودکار برای یادگیری یک نمایش فشرده از ویژگی‌های ورودی برای یک مسئله مدل‌سازی پیش‌بینی طبقه‌بندی ایجاد می‌شود. ابتدا یک مسئله مدل‌سازی پیش‌بینی طبقه‌بندی تعریف می‌شود. برای تعریف یک کار طبقه‌بندی باینری (۲ کلاسه)، ویژگی‌های پروفایل کاربران در شبکه‌های اجتماعی با استفاده از ماتریس ورودی با ستون‌های ویژگی F و ردیف‌هایی از n کاربر در ماتریس ورودی اولیه مدل‌سازی می‌شوند. همچنین یک ماتریس $n\times n$ به عنوان وزن تشابه اولیه وارد رمزگذار خودکار می‌شود. در خودرمزگذارها وزن اولیه به صورت تصادفی برای ورودی‌ها تولید می‌شود و پس از چندین بار تکرار این وزن به‌روز شده و به سمت وزن ایده‌آل کاربران همگرا می‌شود. به شرطی که وزن اولیه بر اساس شباهت نظرات کاربران ارائه شود که می‌تواند روند همگرایی به سمت وزن‌های بهینه را تسریع بخشد.

وقتی شبکه رمزگذار خود را به عنوان یک استخراج‌کننده ویژگی استفاده می‌کنیم، لازم است برخی از داده‌های ورودی را در لایه میانی

1. Query



شکل ۲: معماری شبکه خودرمزگذار ارائه‌شده.

حذف کنیم تا ویژگی‌های اساسی را از ورودی‌ها استخراج کنیم. بنابراین مسئله به گونه‌ای تعریف می‌شود که شامل برخی از متغیرهای ورودی اضافی باشد و این امکان را به رمزگذار می‌دهد که در ادامه یک نمایش فشرده مفید را یاد بگیرد. شبکه‌های رمزگذار معمولاً به گونه‌ای محدود می‌شوند که فقط به تقریب و کپی‌کردن ورودی‌هایی که شبیه داده‌های آموزشی هستند، اجازه داده می‌شود. از آنجایی که مدل مجبور است اولویت‌بندی کند که کدام جنبه از ورودی باید کپی شود، اغلب ویژگی‌های مفید داده‌ها را یاد می‌گیرد. سپس داده‌های ورودی می‌توانند از دامنه به مدل ارائه شوند و خروجی مدل در گلوگاه می‌تواند به عنوان یک بردار ویژگی در یک مدل یادگیری نظارت‌شده، برای تجسم یا به طور کلی برای کاهش ابعاد استفاده شود. همچنین وزن هر لایه به عنوان وزن شباهت کاربران در نظر گرفته می‌شود و این وزن در هر لایه به‌روزرسانی می‌شود تا در نهایت وزن واقعی شباهت کاربران از شبکه رمزگذار استخراج شود.

رمزگذار به گونه‌ای تعریف شده که دو لایه پنهان داشته باشد: لایه اول با دو برابر تعداد ورودی $(۲n)$ و لایه دوم با همان تعداد ورودی (n) و پس از آن لایه گلوگاه با همان تعداد ورودی به عنوان مجموعه داده ورودی (n) است. برای اطمینان از یادگیری بهتر مدل، از نرمال‌سازی دسته‌ای و فعال‌سازی leaky ReLU[1] استفاده می‌شود.

رمزگشا در یک ساختار مشابه، به صورت معکوس، تعریف شده و دو لایه پنهان دارد: لایه اول با تعداد ورودی‌های مجموعه داده (n) و لایه دوم با دو برابر تعداد ورودی‌ها $(۲n)$. 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 با نزول تصادفی گرادیان[2]، با کمینه‌کردن میانگین مربعات خطا آموزش داده می‌شود. شکل ۲ معماری رمزگذار ارائه‌شده را نشان می‌دهد. مطابق با شکل، معماری شبکه رمزگشایی بر اساس لایه‌های تعریف‌شده، ارائه گردیده است. خروجی این مرحله به ماژول توصیه داده می‌شود. تفاوت بین شبکه‌های عصبی ارائه‌شده و شبکه‌های قبلی در تعداد لایه‌های میانی است که اجازه آموزش عمیق بر روی ویژگی‌های پروفایل کاربران را می‌دهد و دقت پیش‌بینی رتبه‌بندی به کاربران جدید را افزایش می‌دهد. در ادامه به توضیح ماژول توصیه ارائه‌شده خواهیم پرداخت.

۳-۲ ماژول توصیه ارائه‌شده

ماژول توصیه در روش ارائه‌شده، آخرین بخشی است که در آن توصیه‌های بازدید صفحات به کاربران بر اساس نتایج شبکه رمزگذار ارائه‌شده اتفاق می‌افتد. خروجی شبکه رمزگذار شامل ویژگی‌های مهم استخراج‌شده از پروفایل کاربران و وزن شباهت کاربران است که در بخش قبلی توضیح داده شد. در این مرحله، بر اساس وزن شباهت کاربران در خوشه‌ها و رتبه‌بندی جدید بر اساس خوشه‌بندی ترتیبی فازی، برای هر کاربر، k کاربر برتر مشابه انتخاب می‌شود. مقدار k در چند سناریوی مختلف، مقادیر مختلفی را می‌گیرد و در نهایت عملکرد سیستم بر اساس تعداد k کاربر مشابه نیز سنجیده می‌شود. این k کاربر به عنوان کاربرانی در نظر گرفته می‌شوند که از نظر روش ارائه‌شده نظرات و سلایق مشابهی می‌توانند داشته باشند. بنابراین برای هر کاربر، آیتم‌هایی که k کاربر مشابه بیشترین رتبه را به آنها اختصاص داده‌اند، می‌تواند به عنوان گزینه‌های مناسبی برای توصیه استفاده شوند. به عبارت دیگر، اگر کاربر i و j بر اساس مقادیر وزن شباهت، مشابه هم در نظر گرفته شده باشند، آیتم‌هایی که از نظر کاربر j بیشترین رتبه یا ۵ را دریافت کرده‌اند، می‌توانند به عنوان آیتم‌های مناسبی برای پیشنهاد به کاربر i در نظر گرفته شوند و همچنین آیتم‌هایی که از نظر کاربر i بیشترین رتبه یا ۵ را دریافت کرده‌اند، می‌توانند به عنوان آیتم‌های مناسبی برای پیشنهاد به کاربر j در نظر گرفته شوند.

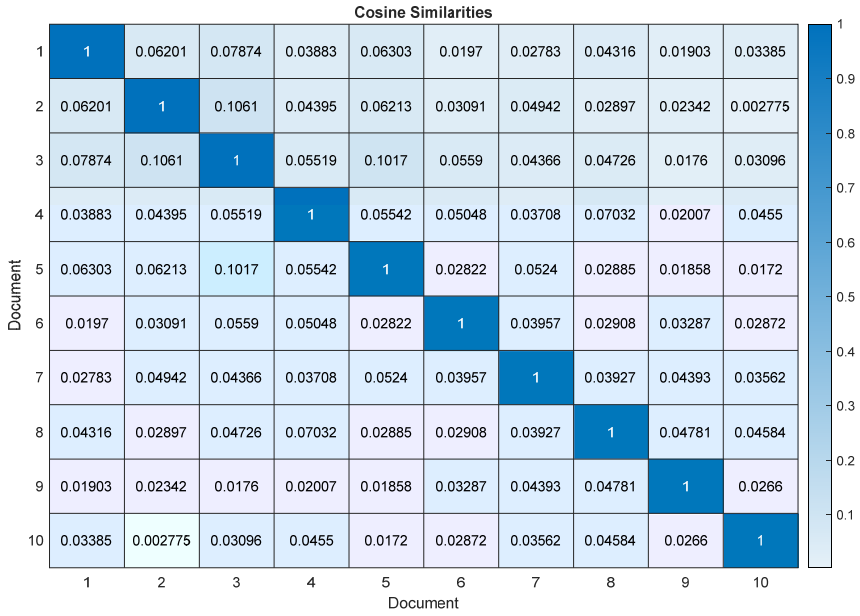
٤- پیاده‌سازی روش ارائه‌شده

همان طور که اشاره شده است، در روش ارائه‌شده به منظور پیاده‌سازی سیستم توصیه‌گر شخصی‌سازی‌شده بر مبنای پروفایل کاربر از دو مجموعه داده زمان واقعی مقیاس بزرگ Yelp و TripAdvisor استفاده شده است. این مجموعه داده‌ها شامل اطلاعاتی مستخرج از پروفایل کاربران در مورد محصولات به صورت متنی و رتبه‌بندی کاربران از محصولات موجود می‌باشد. کاربران با توجه به شناسه منحصربه‌فرد در مجموعه داده مشخص شده‌اند. روش ارائه‌شده با توجه به این داده‌ها در مجموعه داده سعی در شخصی‌سازی صفحه وب بر اساس رفتار کاربران دارد. رفتار کاربران بر اساس اطلاعات موجود در پروفایل شخصی کاربر استنتاج شده و شباهت بین کاربران بر اساس آن محاسبه می‌شود. در جدول ۲ مشخصات مربوط به مجموعه داده‌های مورد استفاده آمده است.

در این مجموعه، داده‌ها به دو قسمت داده‌های آموزشی با نسبت ۷۰٪ کل داده‌ها و داده‌های تست با ۳۰٪ کل داده‌ها تقسیم می‌شود. در مجموعه داده آموزشی Yelp در کنار رتبه‌بندی کاربران نسبت به محصولات، نظرات و کاربران و اطلاعات پروفایل آنها به صورت متنی درج شده است. اطلاعات متنی درج‌شده در پروفایل کاربر در واقع

1. Batch Normalization and Leaky ReLU Activation

2. Stochastic Gradient Descent



شکل ۳: ماتریس شباهت بین کاربران.

جدول ۲: مشخصات مربوط به مجموعه داده‌های مورد استفاده.

عنوان مجموعه داده	تعداد نمونه‌ها	تعداد فیلد	عناوین فیلدها
TripAdvisor	۲۰۴۹۲	۲	Review Rating
Yelp	۱۰۰۰۰۰	۱۰	business_id date review_id stars Text Type user_id Cool useful funny

نشان‌دهنده احساسات و عواطف کاربران نسبت به محصول هستند و در کنار مقادیر رتبه اعطاشده به محصول توسط کاربر، می‌توانند برای تحلیل رفتار کاربران مورد استفاده قرار گیرند. سیستم‌های توصیه‌گر سنتی، تحلیل اطلاعات متنی موجود در پروفایل کاربران را نادیده می‌گیرند و برای محاسبه شباهت بین کاربران از رتبه‌بندی آنها استفاده می‌کنند. حال آن که اطلاعات متنی درج‌شده در پروفایل کاربران حاوی دانش نهفته ارزشمندی است که بر اساس آن می‌توان شباهت کاربران را با دقت بالاتری سنجید. از این رو در روش ارائه‌شده تحلیل اطلاعات موجود در پروفایل کاربری ارائه خواهد شد.

اطلاعات موجود در پروفایل کاربری به صورت متن پیوسته است که قبل از اعمال تغییرات بر روی آن نیاز به پیش‌پردازش و حذف کلمات توقف[1] است.

ماتریس کاربر- کاربر یک ماتریس متقارن است که بر اساس مشابهت نظرات اعطاشده توسط کاربران تنظیم شده است. در این ماتریس هر درایه می‌تواند مقدار بین ۰ تا m دریافت کند که نشان‌دهنده اندیس آیتمی است که دو کاربر در مورد آن مشابهت دارند. همچنین می‌توان دید در برخی درایه‌ها بیش از یک عدد وجود دارد که نشان می‌دهد دو کاربر به بیش از یک آیتم نظر مشابهی اعطا نموده‌اند. چنین کاربرانی که در بیش از یک مورد با هم شباهت دارند، گزینه‌های مناسبی برای بررسی پروفایل کاربری آنها در شبکه‌های اجتماعی به منظور ایجاد سیستم‌های توصیه‌گر صفحات وب هستند که در روش ارائه‌شده بر روی آن تمرکز شده است. بنابراین در روش ارائه‌شده کاربرانی که در مورد یک یا بیشتر آیتم با هم نظرات مشابهی دارند به عنوان کاربران مشابه در نظر گرفته می‌شوند. آیتم‌هایی که این کاربران در مورد آنها با هم تشابه دارند به شکل مقداری نرمال‌شده به عنوان وزن اولیه در روش ارائه‌شده در نظر گرفته می‌شود. این وزن به عنوان یک ویژگی با عنوان وزن شباهت در روش ارائه‌شده به همراه سایر ویژگی‌های موجود در پروفایل کاربری به شبکه عصبی ارائه‌شده وارد می‌شود. پیش‌پردازش مجموعه داده، کاربران با رتبه‌بندی کمتر در مکان‌ها و نتایج با ۳۹۱۰۴ مکان، ۲۰۱۶۶ کاربر منحصربه‌فرد و در کل ۵۸۶۲۷۴ رتبه‌بندی را حذف می‌کند. TripAdvisor یک وب‌سایت توصیه سفر است که از بررسی‌ها و بازخوردهای مکان‌ها تشکیل شده است. مشابه پیش‌پردازش مجموعه داده Yelp، مجموعه داده‌های TripAdvisor نیز از قبل پردازش شده و کاربران با رتبه‌بندی کمتر حذف می‌شوند. مجموعه داده فیلترشده تریپ‌ادوایزر شامل ۹۱۴۹ مکان، ۱۳۴۱۰ کاربر منحصربه‌فرد و در کل ۱۵۲۷۲۱ رتبه‌بندی است.

پس از پیش‌پردازش، شباهت بین نظرات کاربران بر اساس معیار شباهت کسینوسی استخراج می‌شود. در شکل ۳ ماتریس شباهت کسینوسی برای ۱۰ کاربر از مجموعه داده آموزشی به عنوان نمونه نشان داده شده است. با توجه به شکل می‌توان دید ماتریس مشابهت برای کاربران در سیستم توصیه‌گر ارائه‌شده بر اساس اطلاعات موجود در پروفایل کاربری محاسبه شده است. همان طور که دیده می‌شود مقدار شباهت برای هر زوج کاربر مقداری در محدوده [۰,۱] است که میزان شباهت بین دو کاربر از نظر اطلاعات پروفایل کاربری را نشان می‌دهد. هرچه این مقدار بیشتر باشد نشان‌دهنده شباهت بیشتر دو کاربر از نظر اطلاعات پروفایل کاربری است. از این رو در روش ارائه‌شده برای یافتن کاربران مشابه یک آستانه تعریف می‌شود که به صورت رابطه زیر است

$$\alpha = \frac{sim(i,j)}{\sum_{i=1}^{n}\sum_{j\neq i}^{n} sim(i,j) \Big/ n} \geq 1 \qquad (۳)$$

در (۳) i و j اندیس کاربران و n تعداد کل کاربران در مجموعه داده

1. Stopword Remove

جدول ۳: خوشه‌بندی ترتیبی کاربران بر اساس شباهت نظرات.

رتبه پیش‌بینی‌شده	کاربران مشابه	مقدار شباهت	شماره خوشه
۵	۶۸۰ و ۲۱۶	۰/۴۶۴۰	۱
۲	۹۹۸ و ۵۵۴	۰/۴۴۸۰	۲
۴	۴۵۷ و ۱۸	۰/۴۳۷۰	۳
۵	۹۳۵ و ۲۴۵	۰/۴۲۴۰	۴
۵	۵۴۲ و ۱۳۳	۰/۳۸۵۰	۵

آموزشی است. پس از محاسبه شباهت بین نظرات کاربران، در روش ارائه‌شده خوشه‌بندی مبتنی ترتیب فازی انجام می‌شود که در آن کاربران به هر خوشه با توجه به مقدار شباهت اختصاص می‌یابند. پس از اختصاص کاربران به خوشه‌ها، مقدار رتبه‌بندی جدید کاربران با توجه به تابع عضویت فازی محاسبه می‌شود. در جدول ۳ نمونه‌ای از خوشه‌بندی ترتیبی کاربران بر اساس شباهت نظرات آنها نشان داده شده است.

با توجه به جدول می‌توان دید کاربران بر اساس شباهت نظرات در خوشه‌ها قرار می‌گیرند. با توجه به این که هر کابر فقط در یک خوشه می‌تواند جا گیرد، ممکن است در برخی از خوشه‌ها تنها یک کاربر وجود داشته باشد. بر اساس روش ارائه‌شده کاربران اختصاص‌یافته به یک خوشه، بالاترین احتمال رفتارهای مشابه را دارند که بر اساس آن توصیه‌های بازدید وب پیشنهاد شده است. در واقع کاربرانی با شباهت بیشتر از نظر اطلاعات پروفایل کاربری، می‌توانند گزینه مناسبی برای ورودی روش فیلتر مشارکتی بر اساس یادگیری عمیق ارائه شده باشد. در روش ارائه‌شده به منظور شناسایی دقیق کاربران مشابه از یادگیری عمیق استفاده شده است. شبکه‌های خودرمزگذار عمیق ارائه‌شده با توجه به کاربردهای متنوع، گزینه مناسبی برای یافتن کاربران مشابه از نظر اطلاعات پروفایل کاربری هستند. در این نوع از شبکه‌ها اطلاعات پروفایل کاربری و ماتریس شباهت بین کاربران به عنوان ویژگی‌های ورودی شبکه در نظر گرفته می‌شود. همچنین رتبه‌بندی کاربران به محصولات به عنوان متغیر هدف در نظر گرفته می‌شود. شبکه‌های عصبی ارائه‌شده با آموزش عمیق سعی می‌کنند ویژگی‌های هر کاربر را یاد گرفته و کاربران مشابه از نظر رتبه‌های اعطاشده به وب‌سایت‌ها توسط کاربران را پیدا کنند. در نهایت به منظور اعتبارسنجی آموزش خود از شبکه تشابه بین کاربران به عنوان معیار محک و ارزیابی استفاده می‌کنند. در صورتی که دقت شبکه‌های عصبی ارائه‌شده در حد مطلوبی باشد، نتایج آموزش به صورت وزن ویژگی‌ها و کلمات موجود در پروفایل کاربری استخراج شده و به عنوان الگویی برای یافتن کاربران مشابه جدید مورد استفاده قرار می‌گیرند. از این رو هرچه آموزش شبکه‌های عصبی عمیق دقیق‌تر باشد، کاربران مشابه با دقت بیشتری یافته می‌شود. یافتن کاربران مشابه می‌تواند موجب ارائه توصیه‌های دقیق‌تری به کاربران و علی‌الخصوص کاربران جدید باشد. در شکل ۴ و ۵ روند آموزش شبکه عصبی ارائه‌شده بر روی مجموعه داده‌های آموزشی TripAdvisor و Yelp نشان داده شده است. همچنین در جداول ۴ و ۵ میزان دقت شبکه‌های عصبی ارائه‌شده در یافتن کاربران مشابه در طی مراحل تکرار در مجموعه داده آموزشی TripAdvisor و Yelp نشان داده شده است.

با توجه به شکل ۴ و ۵ و جداول ۴ و ۵ می‌توان دید شبکه‌های عصبی ارائه شده در ابتدا در تنظیم وزن‌ها و آموزش ویژگی‌های مربوط به کاربران مقدار خطای بالایی داشته که با افزایش تعداد تکرارها این خطا به مقدار نزدیک به صفر رسیده است.

در این مرحله نتایج به‌دست‌آمده از شبکه عصبی عمیق برای کاربران در مجموعه داده تست بر اساس رتبه‌های اعطاشده به محصولات توسط کاربران تست سنجیده می‌شود. بدین منظور کاربران تستی که به وب‌سایت حداکثر رتبه را اعطا کرده باشند و بر اساس پیش‌بین شبکه عصبی عمیق نیز رتبه اعطایی این کاربران به وب‌سایت حداکثر باشد، به عنوان نمونه‌های مثبت صحیحی [1] (TP) در نظر گرفته می‌شوند. کاربران تستی که به وب‌سایت حداکثر رتبه را اعطا نکرده باشند و بر اساس پیش‌بین شبکه عصبی عمیق نیز رتبه اعطایی این کاربران به وب‌سایت حداکثر نباشد، به عنوان نمونه‌های منفی صحیحی [2] (TN) در نظر گرفته می‌شوند. کاربران تستی که به وب‌سایت حداکثر رتبه را اعطا کرده باشند ولی بر اساس پیش‌بین شبکه عصبی عمیق، رتبه اعطایی این کاربران به وب‌سایت حداکثر نباشد، به عنوان نمونه‌های مثبت کاذب [3] (FP) در نظر گرفته می‌شوند. در نهایت کاربران تستی که به وب‌سایت حداکثر رتبه را اعطا نکرده باشند ولی بر اساس پیش‌بین شبکه عصبی عمیق، رتبه اعطایی این کاربران به وب‌سایت حداکثر باشد، به عنوان نمونه‌های منفی کاذب [4] (FN) در نظر گرفته می‌شوند. این چهار پارامتر با هم به عنوان پارامترهای ماتریس آشفتگی[5] تعیین می‌شوند. در شکل ۶ ماتریس آشفتگی مربوط به شبکه‌های خودرمزگذار عمیق ارائه‌شده برای کاربران تست در مجموعه داده TripAdvisor آمده است.

همان طور که در شکل ۶ نشان داده شده است، ماتریس آشفتگی مربوط به روش شبکه‌های خودرمزگذار عمیق پیشنهادی ارائه شده است. در ادامه ارزیابی روش پیشنهادی ارائه خواهد شد.

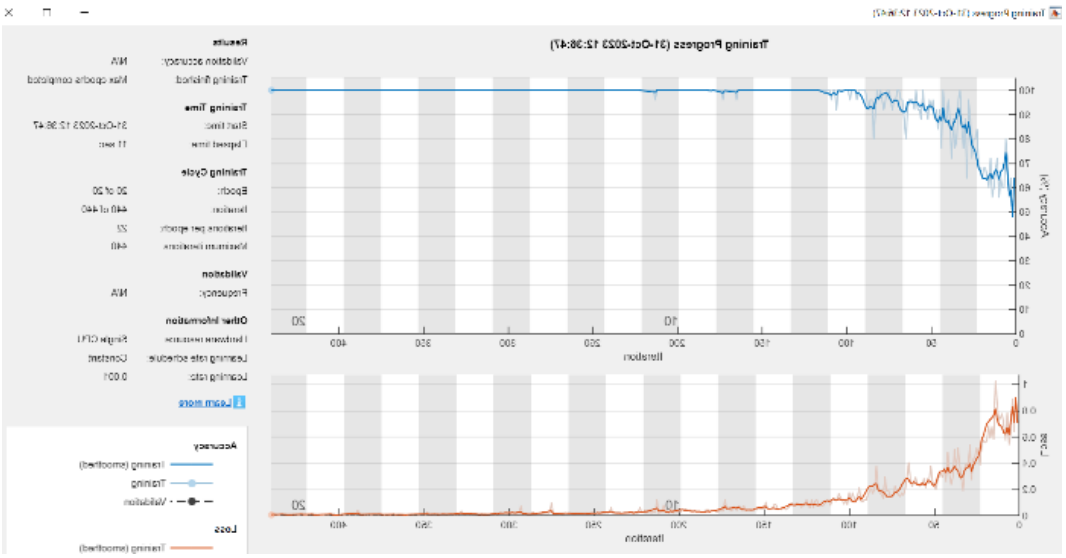
٤-١ ارزیابی روش ارائه‌شده

کیفیت عملکرد سیستم توصیه‌گر می‌تواند توسط معیارهای زیادی مورد سنجش قرار گیرد. نوع معیار مورد استفاده به روش فیلتر و ارائه به‌کاررفته بستگی دارد. دقت عملکرد سیستم توصیه‌گر بر اساس دقت توصیه‌های ارائه‌شده به کاربران تعیین می‌شود. در این تحقیق با توجه به ترکیب در روش فیلتر مشارکتی و یادگیری عمیق، از دو معیار برای سنجش دقت عملکرد روش ارائه‌شده ترکیبی استفاده شده است. معیار اول معیار دقت آماری برای سنجش میزان دقت رتبه‌بندی پیش‌بینی‌شده برای کاربران در روش فیلتر مشارکتی و دقت در ارائه توصیه به کاربران مشابه در میان کاربران جدید است. معیار دوم ارزیابی عملکرد شبکه‌های عصبی عمیق به منظور پیش‌بینی رتبه ارائه‌شده به محصول بر اساس اطلاعات موجود در پروفایل کاربر و مقایسه آن با سایر روش‌های طبقه‌بندی است. در ادامه ابتدا ارزیابی روش پیشنهادی بر اساس دقت آماری ارائه خواهد شد و بعد از آن دقت روش‌های طبقه‌بندی با هم مقایسه می‌گردد.

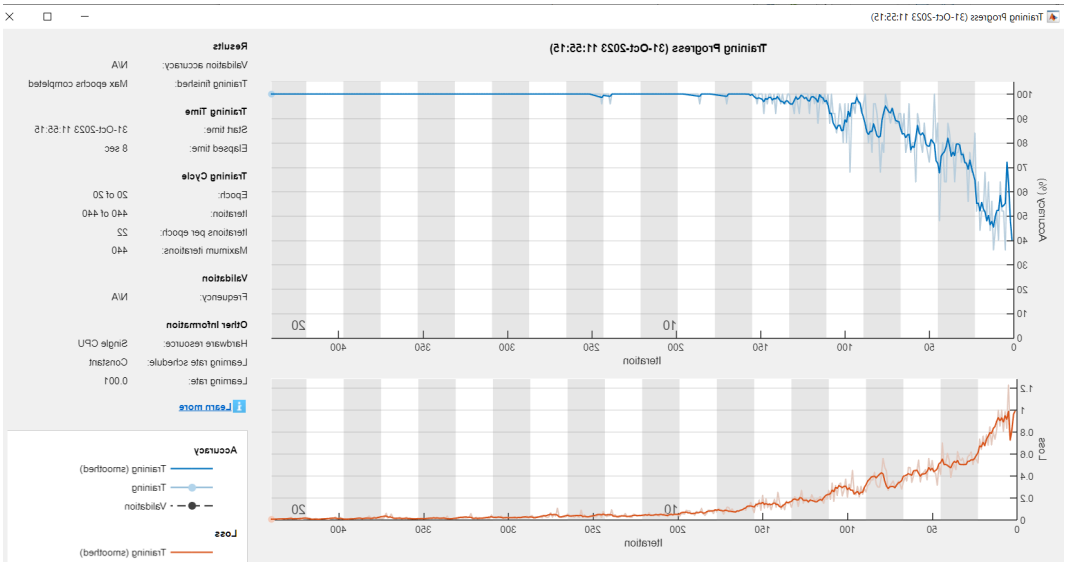
٤-١-١ ارزیابی عملکرد روش ارائه‌شده بر اساس دقت آماری

همان طور که اشاره شد، معیارهای دقت آماری، دقت از روش فیلتر مشارکتی را با مقایسه به طور مستقیم رتبه‌های پیش‌بینی‌شده برای کاربران را با امتیاز واقعی‌شده آنها مورد بررسی قرار می‌دهد. میانگین خطای مطلق [6] (MAE) و خطای میانگین مربعات [7] ($RMSE$) معمولاً

1. True Positive
2. True Negative
3. False Positive
4. False Negative
5. Confusion Matrix
6. Mean Absolute Error
7. Root Mean Square Error



شکل ۴: روند آموزش شبکه عصبی ارائه‌شده بر روی مجموعه داده آموزشی TripAdvisor.



شکل ۵: روند آموزش شبکه عصبی ارائه‌شده بر روی مجموعه داده آموزشی Yelp.

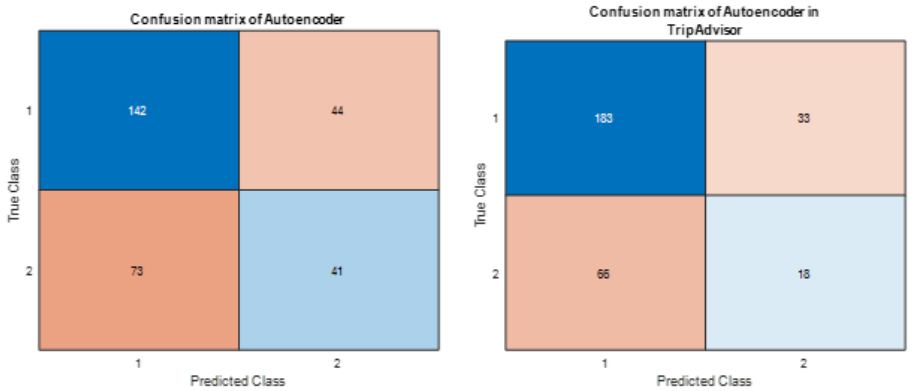
جدول ۴: میزان دقت شبکه عصبی ارائه‌شده در یافتن کاربران مشابه در طی مراحل تکرار در مجموعه داده آموزشی TripAdvisor.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
۱	۱	۰۰:۰۰:۰۴	٪۶۴٫۰۰	۰٫۷۰۴۷	۰٫۰۰۱۰
۳	۵۰	۰۰:۰۰:۰۵	٪۹۲٫۰۰	۰٫۲۲۵۷	۰٫۰۰۱۰
۵	۱۰۰	۰۰:۰۰:۰۶	٪۱۰۰٫۰۰	۰٫۱۷۶۳	۰٫۰۰۱۰
۷	۱۵۰	۰۰:۰۰:۰۶	٪۱۰۰٫۰۰	۰٫۰۴۹۵	۰٫۰۰۱۰
۱۰	۲۰۰	۰۰:۰۰:۰۷	٪۱۰۰٫۰۰	۰٫۰۳۳۸	۰٫۰۰۱۰
۱۲	۲۵۰	۰۰:۰۰:۰۸	٪۱۰۰٫۰۰	۰٫۰۱۴۷	۰٫۰۰۱۰
۱۴	۳۰۰	۰۰:۰۰:۰۹	٪۱۰۰٫۰۰	۰٫۰۲۷۳	۰٫۰۰۱۰
۱۶	۳۵۰	۰۰:۰۰:۱۰	٪۱۰۰٫۰۰	۰٫۰۰۸۷	۰٫۰۰۱۰
۱۹	۴۰۰	۰۰:۰۰:۱۰	٪۱۰۰٫۰۰	۰٫۰۰۷۳	۰٫۰۰۱۰
۲۰	۴۴۰	۰۰:۰۰:۱۱	٪۱۰۰٫۰۰	۰٫۰۰۵۹	۰٫۰۰۱۰

جدول ۵: میزان دقت شبکه عصبی ارائه‌شده در یافتن کاربران مشابه در طی مراحل تکرار در مجموعه داده آموزشی YELP.

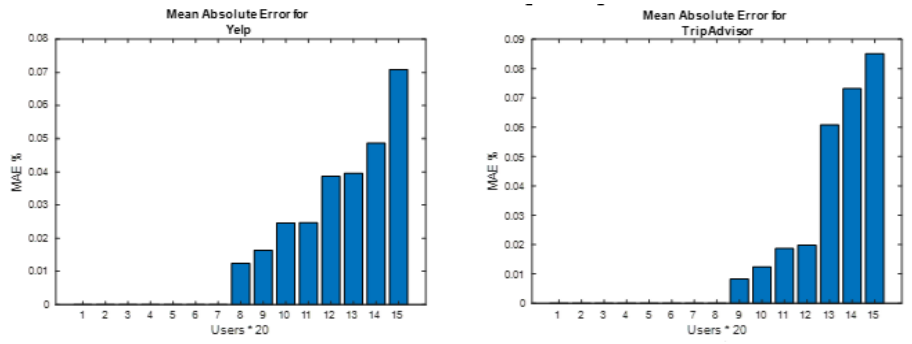
Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
۱	۱	۰۰:۰۰:۰۳	٪۴۰٫۰۰	۰٫۹۹۱۸	۰٫۰۰۱۰
۳	۵۰	۰۰:۰۰:۰۳	٪۷۲٫۰۰	۰٫۵۳۹۴	۰٫۰۰۱۰
۵	۱۰۰	۰۰:۰۰:۰۴	٪۹۶٫۰۰	۰٫۲۰۶۰	۰٫۰۰۱۰
۷	۱۵۰	۰۰:۰۰:۰۴	٪۱۰۰٫۰۰	۰٫۲۲۱۴	۰٫۰۰۱۰
۱۰	۲۰۰	۰۰:۰۰:۰۵	٪۱۰۰٫۰۰	۰٫۰۵۲۴	۰٫۰۰۱۰
۱۲	۲۵۰	۰۰:۰۰:۰۶	٪۱۰۰٫۰۰	۰٫۰۳۷۶	۰٫۰۰۱۰
۱۴	۳۰۰	۰۰:۰۰:۰۶	٪۱۰۰٫۰۰	۰٫۰۲۲۰	۰٫۰۰۱۰
۱۶	۳۵۰	۰۰:۰۰:۰۷	٪۱۰۰٫۰۰	۰٫۰۱۱۵	۰٫۰۰۱۰
۱۹	۴۰۰	۰۰:۰۰:۰۷	٪۱۰۰٫۰۰	۰٫۰۰۸۱	۰٫۰۰۱۰
۲۰	۴۴۰	۰۰:۰۰:۰۸	٪۱۰۰٫۰۰	۰٫۰۰۶۲	۰٫۰۰۱۰

به عنوان معیارهای دقت آماری استفاده می‌شود. *MAE* از محبوب‌ترین و معمول‌ترین روش‌هایی است که در بین بیشتر روش‌ها مورد استفاده قرار گرفته است. طبیعتاً هرچه مقدار این معیار در ازای کاربران، کمتر باشد، دقت عملکرد روش فیلتر مشارکتی بیشتر خواهد بود. به عبارت دیگر هرچه مقدار این معیار کمتر باشد، رتبه‌بندی پیش‌بینی‌شده برای کاربران با امتیاز واقعی آنها نزدیک‌تر است. این معیارها به صورت (۴) و (۵) محاسبه می‌شوند



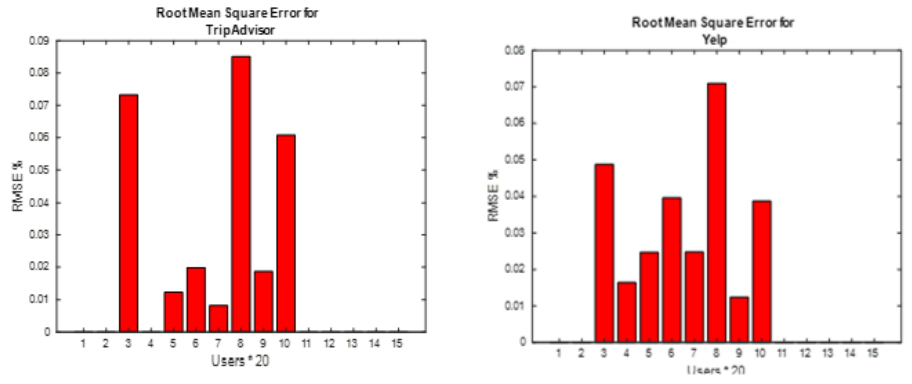
(ب) (الف)

شکل ۶: ماتریس آشفتگی مربوط به شبکه‌های خودرمزگذار عمیق در مجموعه داده‌های (الف) TripAdvisor و (ب) Yelp.



(ب) (الف)

شکل ۷: مقدار خطای مطلق در ازای کاربران تست.



(ب) (الف)

شکل ۸: مقدار خطای میانگین مربعات در ازای کاربران تست.

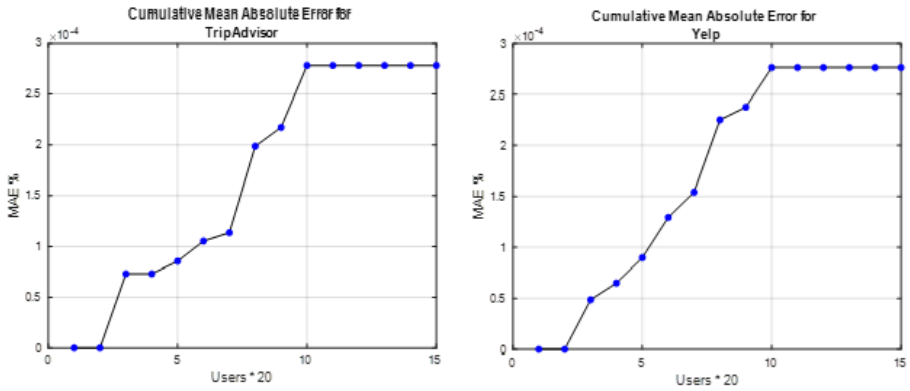
$$MAE = \frac{1}{N}\sum_{u,i}\left|p_{u,i} - r_{u,i}\right| \quad (۴)$$

$$RMSE = \sqrt{\frac{\sum_{u,i}(p_{u,i} - r_{u,i})^2}{N}} \quad (۵)$$

که $P_{u,i}$ به عنوان رتبه پیش‌بینی‌شده برای کاربر u روی محصول i، $r_{u,i}$ رتبه واقعی کاربر u روی محصول i و N کل تعداد رتبه‌بندی در مجموعه گزینه‌ها تعریف شده است. شکل ۷ مقدار خطای مطلق در ازای کاربران تست در مجموعه داده‌های TripAdvisor و Yelp را نشان می‌دهد. همچنین در شکل ۷ مقدار خطای میانگین مربعات در ازای کاربران تست در مجموعه داده TripAdvisor و Yelp آمده است.

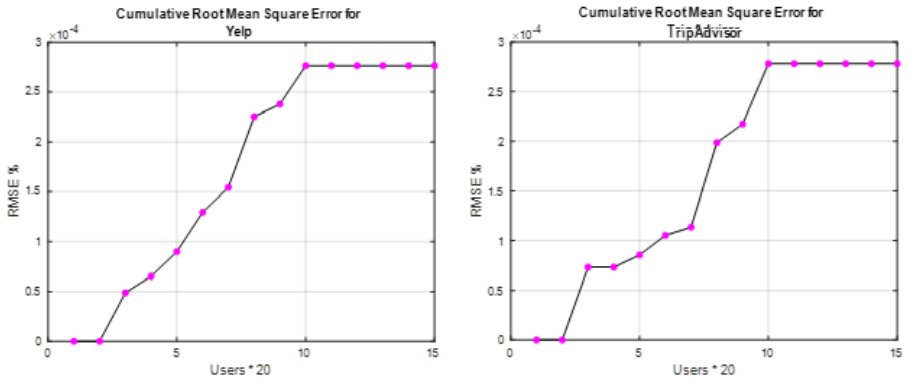
همان طور که در شکل ۷ و ۸ نشان داده شده است، مقادیر خطای مطلق و خطای میانگین مربعات برای هر یک از کاربران در مجموعه داده‌های تست TripAdvisor و Yelp محاسبه شده است. حال مقادیر تجمعی خطاهای مطلق و میانگین مربعات به ازای همه کاربران تست در مجموعه داده‌های TripAdvisor و Yelp محاسبه خواهد شد. شکل ۹ نمودار تجمعی مقدار خطای مطلق در ازای همه کاربران تست در مجموع داده TripAdvisor را نشان می‌دهد. همچنین در شکل ۱۰ نمودار تجمعی خطای میانگین مربعات در ازای همه کاربران تست در مجموع داده TripAdvisor نشان داده شده است.

همان طور که در شکل ۹ و ۱۰ دیده می‌شود، نمودار تجمعی مقادیر خطای مطلق و خطای میانگین مربعات برای کاربران تست در هر دو مجموعه داده با شیب ملایمی افزایش می‌یابد که نشان‌دهنده دقت بالای پیش‌بینی رتبه اعطاشده به کاربر جدید بر اساس شبکه‌های خودرمزگذار عمیق ارائه شده است. در نهایت مقدار خطای مطلق تجمیعی در مجموعه داده‌های TripAdvisor و Yelp به ترتیب به حداکثر مقدار ۰/۰۰۴۲۲ و ۰/۰۰۲۲۸ و مقدار خطای میانگین مربعات تجمیعی به ترتیب به حداکثر مقدار ۰/۰۰۲۷۲ و ۰/۰۰۲۶۷ می‌رسد.



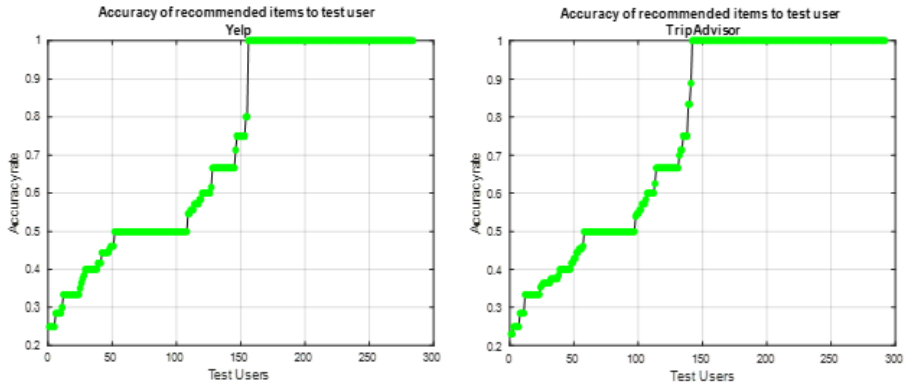
(ب) (الف)

شکل ۹: نمودار تجمعی خطای میانگین مربعات.



(ب) (الف)

شکل ۱۰: نمودار تجمعی مقدار خطای میانگین مربعات.



(ب) (الف)

شکل ۱۱: دقت توصیه‌های ارائه‌شده.

۴-۱-۲ ارزیابی عملکرد روش ارائه‌شده بر اساس دقت توصیه‌های ارائه‌شده

پس از محاسبه دقت آماری بر اساس خطای مطلق و میانگین مربعات برای کاربران تست در مجموعه داده‌های TripAdvisor و Yelp، حال نوبت به محاسبه دقت توصیه‌های ارائه‌شده به کاربران تست در این مجموعه داده‌ها می‌رسد. به منظور محاسبه دقت توصیه‌های ارائه‌شده، توصیه‌های مفید از میان تمام توصیه‌های ارائه‌شده در مجموعه داده شناسایی می‌شود. توصیه‌های مفید شامل آن دسته از وب‌سایت‌هایی است که به کاربر I توصیه شده و کاربر i در مجموعه داده اصلی به آن وب‌سایت‌ها حداکثر مقدار رتبه‌بندی را اعطا نموده است. با توجه به این که مقدار رتبه‌بندی کاربران تست در مجموعه داده اصلی مشخص است، به‌راحتی می‌توان توصیه‌های مفید را مشخص نمود. پس از تشخیص توصیه‌های مفید، نسبت توصیه‌های مفید در میان کل توصیه‌های ارائه‌شده مشخص گردیده است. در واقع دقت توصیه‌های ارائه‌شده برابر با نسبت توصیه‌های مفید ارائه‌شده نسبت به کل توصیه‌های ارائه‌شده به کاربران تست می‌باشد. در شکل ۱۱ دقت توصیه‌های ارائه‌شده در مجموعه داده‌های TripAdvisor و Yelp نشان داده شده است.

با توجه به شکل ۱۱ می‌توان دید دقت توصیه‌های ارائه‌شده در مجموعه داده TripAdvisor و Yelp به صورت صعودی افزایش می‌یابد که این مورد نشان‌دهنده پیش‌بینی دقیق رتبه اعطاشده به کاربران جدید و یافتن کاربران مشابه بر اساس اطلاعات پروفایل کاربری است که به توصیه‌های بازدید وب دقیق منجر شده است. میانگین دقت ارائه توصیه بازدید وب در مجموعه داده‌های TripAdvisor و Yelp به ترتیب برابر با ۷۴٫۵۳٪ و ۷۲٫۲۲٪ است.

٤-۱-۳ ارزیابی دقت روش‌های طبقه‌بندی

ارزیابی روش ارائه‌شده برای تجزیه و تحلیل و بررسی عملکرد روش ارائه‌شده بر روی نمونه‌های جدید که روند آموزش بر روی آنها اتفاق نیفتاده است، به منظور پیش‌بینی رتبه‌بندی این نمونه‌ها صورت می‌گیرد. به منظور بررسی عملکرد و ارزیابی روش‌های طبقه‌بندی که داده‌ها در کلاس‌ها مختلفی توزیع شده‌اند، معیارهای سنجش مختلفی وجود دارد که بیشتر این معیارها از تقابل رتبه‌بندی پیش‌بینی‌شده برای نمونه‌ها در مقابل رتبه‌بندی واقعی آنها که در ماتریس آشفتگی مشخص می‌شود، به دست می‌آیند. برای محاسبه معیارهای ارزیابی در ماتریس آشفتگی چهار پارمتر اصلی مثبت صحیح (TP)، مثبت کاذب (FP)، منفی صحیح (TN) و منفی کاذب (FN) باید از روش طبقه‌بندی به دست آید که این پارامترهای قبلاً معرفی شده‌اند. بر اساس این پارامترها می‌توان معیارهای ارزیابی زیر را تعریف کرد

$$Accurace = \frac{TP + TN}{TP + TN + FP + FN} \quad (۶)$$

$$Recall = \frac{TP}{TP + FN} \quad (۷)$$

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

$$F - measure = \frac{۲ \times Precision \times Recall}{Precision + Recall} \quad (۹)$$

معیارهای ارزیابی فوق از معروف‌ترین معیارهایی هستند که با اتکا بر آنها عملکرد یک روش طبقه‌بندی مورد سنجش قرار گرفته است. این معیارها استاندارد هستند و بر اساس آنها می‌توان روش ارائه‌شده را با سایر روش‌ها در این زمینه مورد مقایسه قرار داد. در روش ارائه‌شده علاوه بر شبکه‌های خودرمزگذار عمیق ارائه‌شده از سه طبقه‌بند معروف K نزدیک‌ترین همسایه، ماشین پشتیبان بردار و طبقه‌بند بیزین به منظور پیش‌بینی رتبه‌بندی کاربران تست استفاده گردیده و با شبکه‌های خودرمزگذار عمیق ارائه‌شده مقایسه می‌شود. در روش پیشنهادی مقدار K در الگوریتم K نزدیک‌ترین همسایه به صورت پیش‌فرض ۵ در نظر گفته شده است. همچنین برای الگوریتم بیزین از رابطه قضیه بیز که در (۱۰) نشان داده شده است، استفاده نموده‌ایم. قضیه بیز را می‌توان به صورت زیر نشان داد

$$P(Q|R) = \frac{p(R|Q)p(Q)}{P(R)} \quad (۱۰)$$

در (۱۰)، احتمال Q با توجه به R محاسبه می‌شود. $P(Q)$ احتمال Q به طور مستقل، $P(R)$ احتمال R به طور مستقل و $P(R|Q)$ احتمال R به شرط تحقق Q است. در شکل ۱۲ ماتریس آشفتگی مربوط به سایر روش‌های طبقه‌بندی نشان داده شده است.

در شکل ۱۳ نمودار معیار دقت در روش‌های طبقه‌بندی مختلف در مجموعه داده TripAdvisor و Yelp نمایش داده شده است. همان طور که در شکل می‌توان دید، در روش ارائه‌شده بر اساس اطلاعات موجود در پروفایل کاربری و رتبه‌بندی کاربران نسبت به صفحات وب و ترکیب آن با شبکه عصبی عمیق و سایر طبقه‌بندها، روش شبکه عصبی عمیق نسبت به سایر روش‌های طبقه‌بندی مقدار بهتری برای معیار دقت را به دست آورده است. با توجه به شکل ۱۳ می‌توان دریافت تعداد کاربران تست ۳۰۰ کاربر می‌باشد که مقادیر دقت پیش‌بینی از هر ۱۰ 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.

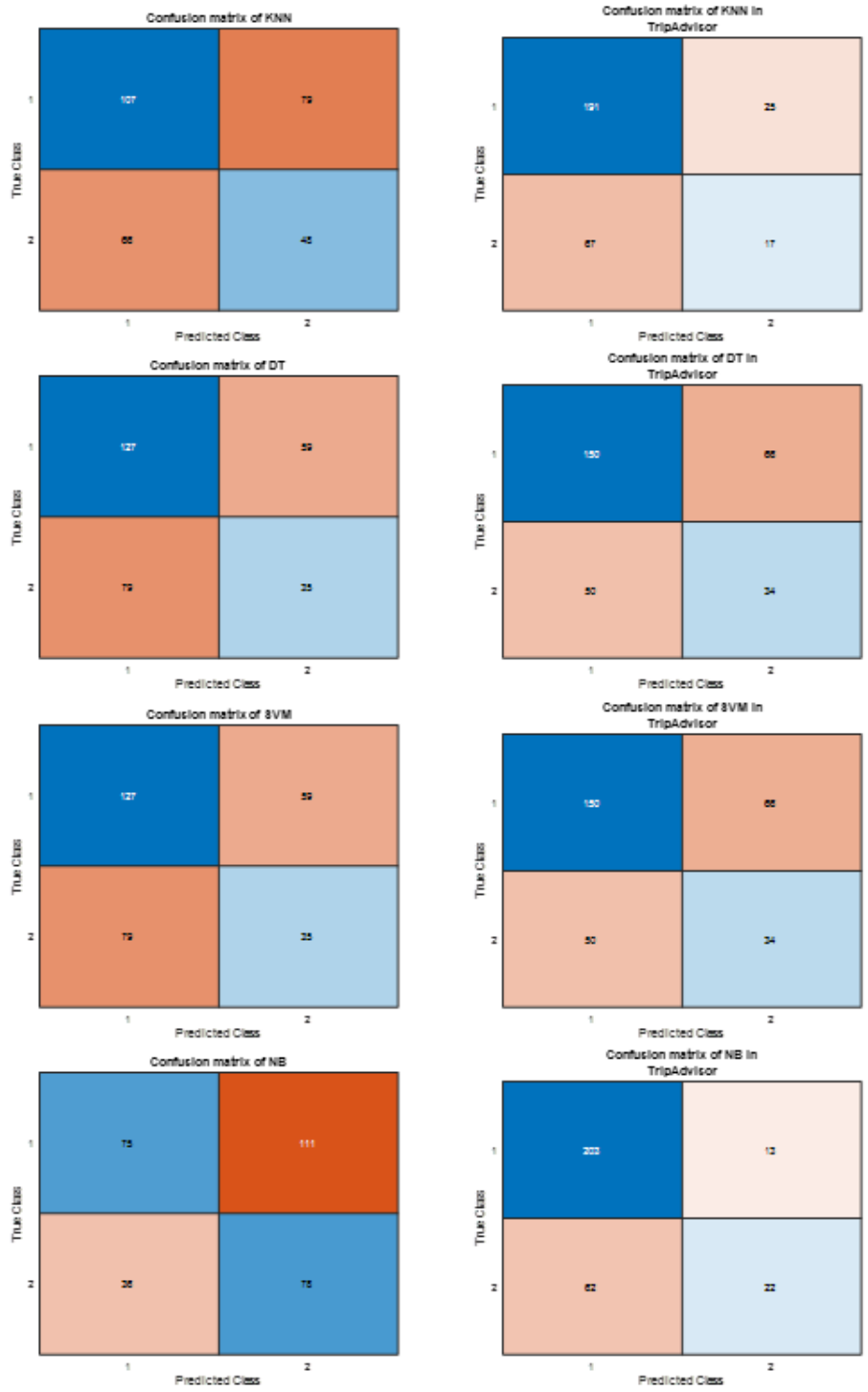
معیار دیگری که در روش پیشنهادی مورد استفاده قرار گرفته است، معیار حساسیت است. معیار حساسیت به عنوان نسبت رتبه‌های حداکثری پیش‌بینی‌شده، تعریف می‌شود که به طور صحیح در میان تمام رتبه‌های حداکثری در مجموعه داده تست پیش‌بینی شده است. در شکل ۱۴ نمودار معیار حساسیت در روش‌های طبقه‌بندی مختلف در مجموعه داده TripAdvisor و Yelp نمایش داده شده است. همان طور که در شکل می‌توان دید، در روش ارائه‌شده بر اساس اطلاعات موجود در پروفایل کاربری و رتبه‌بندی کاربران نسبت به صفحات وب و ترکیب آن با شبکه عصبی عمیق و سایر طبقه‌بندها، روش شبکه عصبی عمیق نسبت به سایر روش‌های طبقه‌بندی مقدار بهتری را برای معیار حساسیت به دست آورده است.

سومین معیار در روش ارائه‌شده، معیار صحت است. معیار صحت نشان‌دهنده درستی رتبه‌های حداکثری پیش‌بینی‌شده در میان رتبه‌های حداکثری کاربران تست کشف‌شده توسط روش‌های یادگیری عمیق است. در شکل ۱۵ نمودار معیار صحت در روش‌های طبقه‌بندی مختلف در مجموعه داده TripAdvisor و Yelp نمایش داده شده است. همان طور که در شکل می‌توان دید، در روش ارائه‌شده بر اساس اطلاعات موجود در پروفایل کاربری و رتبه‌بندی کاربران نسبت به صفحات وب و ترکیب آن با شبکه عصبی عمیق و سایر طبقه‌بندها، روش شبکه عصبی عمیق نسبت به سایر روش‌های طبقه‌بندی مقدار بهتری برای معیار صحت را به دست آورده است.

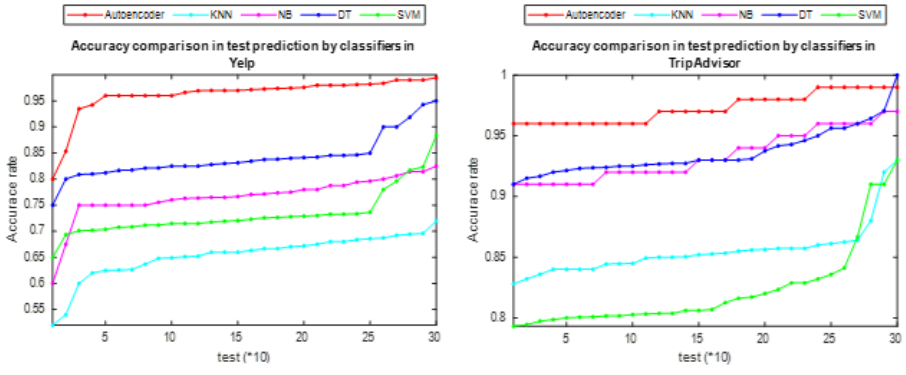
آخرین معیار استفاده‌شده، معیار F است که حاصل ترکیب صحت و حساسیت است. این معیار نشان‌دهنده عملکرد کلی مدل ارائه‌شده به ازای تشخیص رتبه‌بندی کاربران در مجموعه داده‌های استاندارد است. در شکل ۱۶ نمودار معیار F در روش‌های طبقه‌بندی مختلف در مجموعه داده TripAdvisor و Yelp نمایش داده شده است. همان طور که در شکل می‌توان دید، در روش ارائه‌شده بر اساس اطلاعات موجود در پروفایل کاربری و رتبه‌بندی کاربران نسبت به صفحات وب و ترکیب آن با شبکه عصبی عمیق و سایر طبقه‌بندها، روش شبکه عصبی عمیق نسبت به سایر روش‌های طبقه‌بندی مقدار بهتری برای معیار F را به دست آورده است.

با توجه به نتایج به‌دست‌آمده می‌توان گفت روش ارائه‌شده با استفاده از اطلاعات موجود در پروفایل کاربری و رتبه‌بندی کاربران نسبت به وب‌سایت‌ها ورودی بسیار مناسبی برای روش طبقه‌بندی شبکه عصبی عمیق مهیا کند. روش شبکه عصبی عمیق با توجه به مناسب‌بودن داده‌های ورودی، قدرت آموزش و پیش‌بینی بالایی بر روی این مجموعه ویژگی‌ها به دست آورده‌اند. در میان روش‌های طبقه‌بندی، روش شبکه عصبی عمیق مناسب‌ترین نتایج را برای طبقه‌بندی نمونه‌های آموزشی و پیش‌بینی نمونه‌های رتبه کاربران در مجموعه داده آزمایشی به دست آورده است. در شکل ۱۷ نمودار میله‌ای مقایسه تفکیکی معیارهای ارزیابی برای روش‌های طبقه بندی مختلف در مجموعه داده TripAdvisor و Yelp نشان داده شده است. در شکل ۱۸ نمودار میله‌ای مقایسه تجمیعی معیارهای ارزیابی برای روش‌های طبقه‌بندی مختلف در مجموعه داده TripAdvisor و Yelp نشان داده شده است.

با توجه به شکل ۱۸ می‌توان دید که شبکه‌های خودرمزگذار عمیق ارائه‌شده با توجه به یادگیری عمیق بر روی اطلاعات موجود در پروفایل کاربری در مجموعه داده TripAdvisor و Yelp، توانسته‌اند با بالاترین دقت رتبه‌بندی محصولات برای کاربران جدید را پیش‌بینی نماید. روش

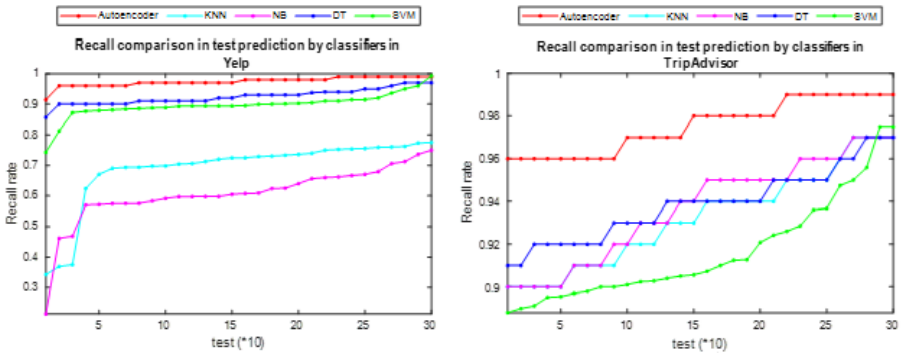


شکل ۱۲: ماتریس آشفتگی سایر روش‌های طبقه‌بندی.



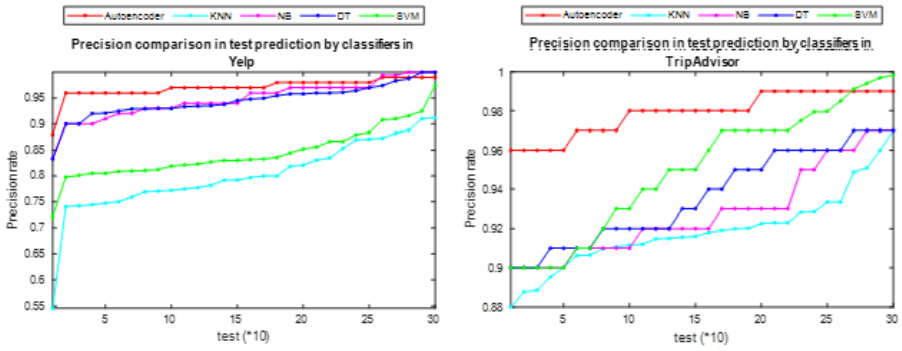
(ب) (الف)

شکل ۱۳: نمودار معیار دقت در روش‌های طبقه‌بندی.



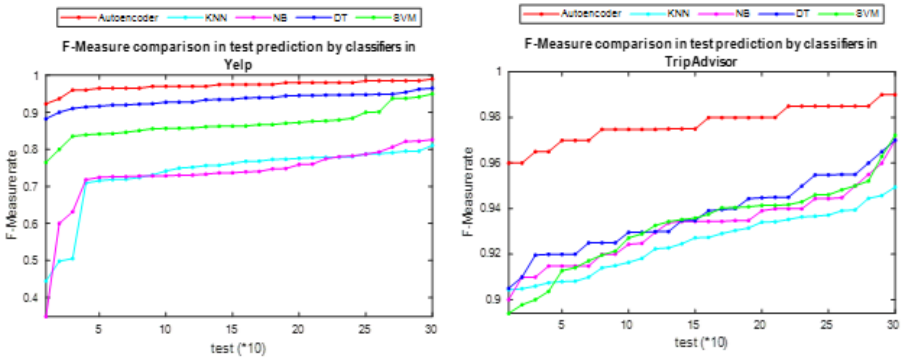
(ب) (الف)

شکل ۱۴: نمودار معیار حساسیت در روش‌های طبقه‌بندی.



(ب) (الف)

شکل ۱۵: نمودار معیار صحت در روش‌های طبقه‌بندی مختلف در مجموعه داده Yelp.



(ب) (الف)

شکل ۱۶: نمودار معیار F در روش‌های طبقه‌بندی.

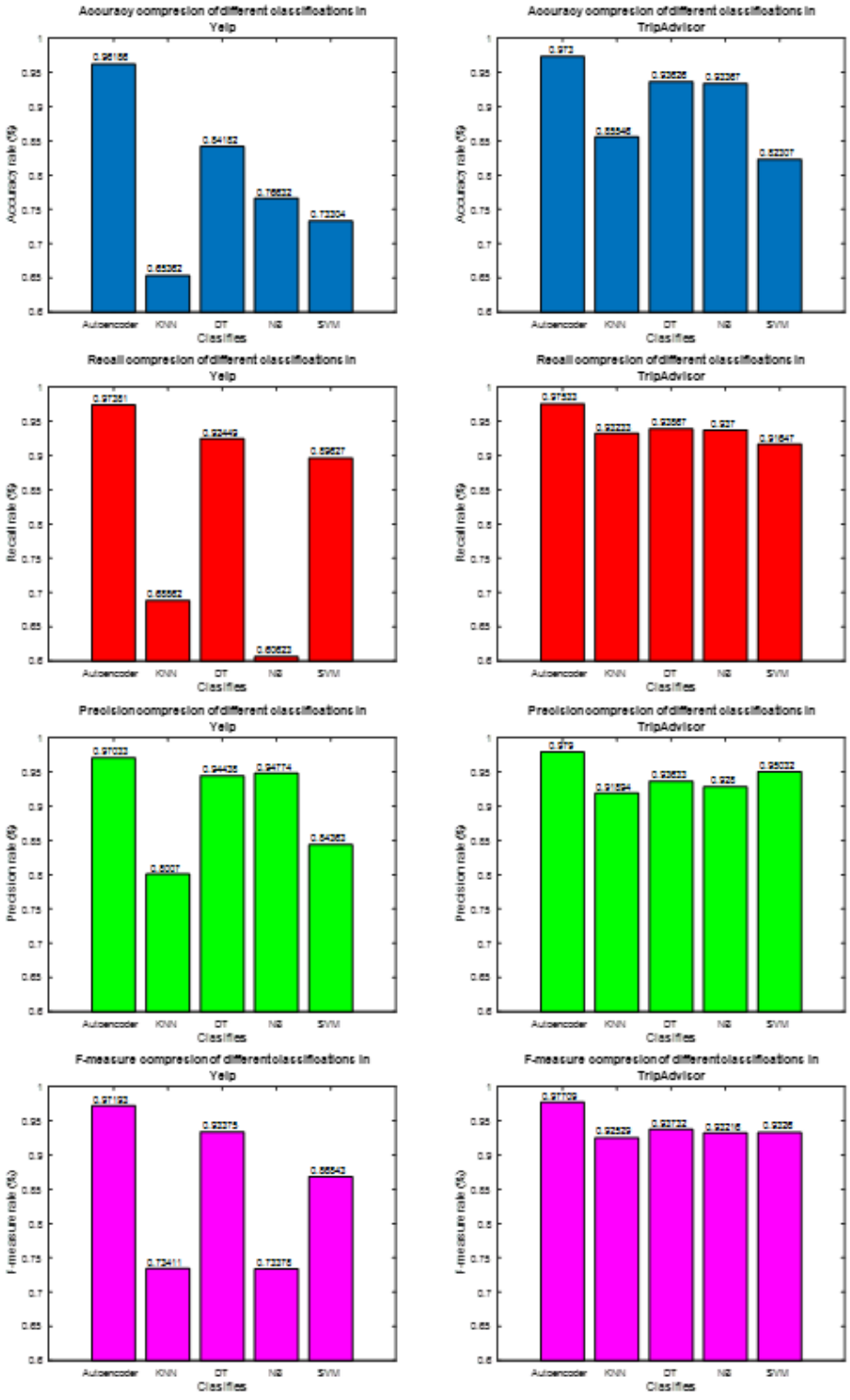
ارائه‌شده با توجه به لایه‌های آموزش عمیق و تکمیل فرایند آموزش در لایه میانی توانسته است از سایر روش‌های طبقه‌بندی از نظر معیارهای ارزیابی بهتر عمل کند.

۴-۲ مقایسه روش پیشنهادی با روش‌های پیشین

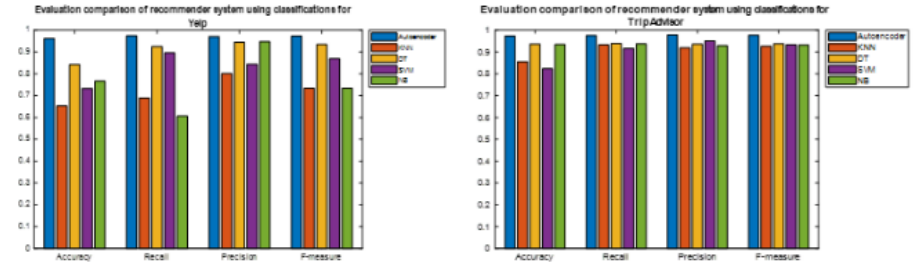
پس از پیاده‌سازی و ارزیابی به منظور اعتبارسنجی، روش پیشنهادی با روش‌های پیشین شامل [۲۱] مورد مقایسه قرار گرفته است. آنچه مسلم است این است که مقایسه بر روی نتایج روش‌های پیشین بر روی مجموعه داده‌های یکسان شامل مجموعه داده Yelp و Trip Advisor در شرایط یکسانی صورت می‌گیرد. از این رو نتایج به‌دست‌آمده در روش پیشنهادی در شرایط یکسان و مجموعه داده یکسان با روش‌های دیگری مقایسه می‌شود تا میزان بهبود حاصل از ایده پیشنهادی در این مقاله مشخص شود. با توجه به این که در روش پیشنهادی از ترکیب یادگیری عمیق و فیلتر مشارکی استفاده شده است، می‌توان برای مقایسه از معیارهایی مانند میانگین خطای مطلق، نرخ برخورد و دقت یافتن کاربران مشابه استفاده کرد. در شکل ۱۹ مقایسه روش پیشنهادی با روش‌های پیشین از نظر معیار میانگین خطای مطلق آمده است.

همان طور که در شکل ۱۹ قابل مشاهده است، روش پیشنهادی میانگین خطای مطلق کمتری نسبت به روش‌های پیشین به دست آورده است. دلیل این امر آموزش عمیق ویژگی‌های پروفایل کاربر و یافتن کاربران مشابه با دقت بالایی توسط شبکه عصبی خودرمزگذار پیشنهادی است. از این رو انتظار می‌رود با توجه به مقدار کم خطای مطلق، دقت توصیه‌های ارائه‌شده که به عنوان نرخ برخورد نیز در مقالات یاد می‌شود، در روش پیشنهادی نسبت به سایر روش‌های موجود، مقدار بهتری داشته باشد. در شکل ۲۰ مقایسه روش پیشنهادی با روش‌های پیشین از نظر نرخ برخورد یا دقت توصیه‌های ارائه‌شده به کاربران آمده است.

با توجه به شکل ۲۰ می‌توان دید روش پیشنهادی با تکیه بر آموزش عمیق در شبکه‌های خودرمزگذارپیشنهادی توانسته است، ویژگی‌های کاربران را به خوبی آموزش دهد که همین امر موجب کاهش خطای مطلق در ارائه توصیه و افزایش نرخ برخورد یا دقت توصیه‌های ارائه‌شده

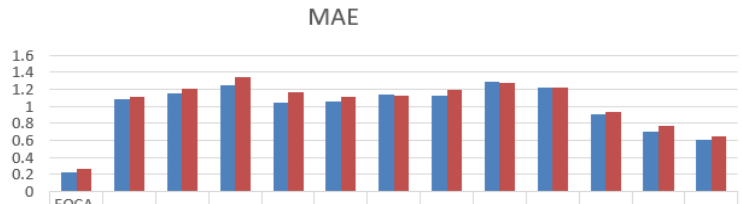


شکل ۱۷: نمودار میله‌ای مقایسه تفکیکی معیارهای ارزیابی.



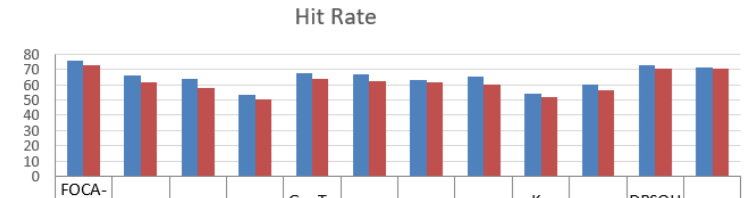
(ب) (الف)

شکل ۱۸: نمودار میله‌ای مقایسه تجمیعی معیارهای ارزیابی.



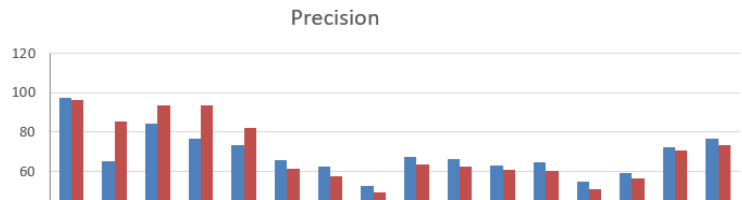
	FOCA-Autoencoder	UCF	ICF	TBCF	GeoTeCS	ICGS	DCR	DCW	K-Means	PSO	DPSOHiK	ABiPRS	HSSRS
Yelp	0.224	1.0845	1.1574	1.2548	1.0456	1.0545	1.1436	1.1321	1.2938	1.2213	0.9015	0.7082	0.6124
Trip Advisor	0.272	1.1082	1.2064	1.3426	1.1622	1.1173	1.1191	1.1949	1.2754	1.2167	0.9324	0.7704	0.6459

شکل ۱۹: مقایسه روش پیشنهادی با روش‌های پیشین از نظر میانگین خطای مطلق.



	FOCA-Autoencoder	UCF	ICF	TBCF	GeoTeCS	ICGS	DCR	DCW	K-Means	PSO	DPSOHiK	ABiPRS
Yelp	75.46	66.37	63.46	53.75	67.63	66.58	63.42	65.35	54.47	59.72	72.65	71.48
Trip Advisor	72.76	61.48	57.62	50.39	63.57	62.48	61.45	60.36	51.65	56.48	70.34	70.26

شکل ۲۰: مقایسه روش پیشنهادی با روش‌های پیشین از نظر نرخ برخورد.



	FOCA-Autoencoder	FOCA-KNN	FOCA-DT	FOCA-NBCF	FOCA-SVMCF	UCF	ICF	TBCF	GeoTeCS	ICGS	DCR	DCW	K-Means	PSO	DPSOHiK	ABiPRS
Yelp	97.3	65.36	84.18	76.63	73.3	65.54	62.73	52.76	67.22	66.14	63.15	64.82	54.67	59.28	72.46	76.74
Trip Advisor	96.18	85.55	93.63	93.37	82.31	61.45	57.37	49.67	63.41	62.35	60.82	60.27	51.26	56.31	70.57	73.64

شکل ۲۱: مقایسه روش پیشنهادی با روش‌های پیشین از نظر دقت تشخیص کاربران مشابه.

گردیده است. آنچه در شکل ۲۰ قابل مشاهده است این است که روش پیشنهادی نرخ برخورد بالاتری نسبت به روش‌های پیشین بر روی مجموعه داده‌های یکسان به دست آورده است.

با توجه به استفاده از یادگیری عمیق و سایر روش‌های طبقه‌بندی در زمینه تشخیص کاربران مشابه بر اساس اطلاعات پروفایل کاربران، روش پیشنهادی را می‌توان از لحاظ دقت تشخیص کاربران مشابه با روش‌های پیشین مقایسه کرد. در روش پیشنهادی علاوه بر شبکه عصبی خودرمزگذار، از سایر روش‌های طبقه‌بندی مانند k نزدیک‌ترین همسایه (KNN)، بیزین ساده (NB) و ماشین پشتیبان بردار (SVM) نیز به منظور یافتن کاربران مشابه بر اساس اطلاعات پروفایل کاربری استفاده شده است. مقایسه روش پیشنهادی با روش‌های طبقه‌بندی دیگر و روش‌های پیشین در نشریات از نظر دقت یافتن کاربران مشابه در شکل ۲۱ نشان داده شده است. همان طور که در شکل آمده است، روش پیشنهادی با توجه به استفاده از یادگیری عمیق بر روی ویژگی‌های کاربران که از پروفایل آنها استخراج می‌شود، توانسته است دقت بالاتری در درستی تشخیص کاربران مشابه نسبت به این روش‌های طبقه‌بندی و روش‌های پیشین به دست آورد.

۵- نتیجه‌گیری

سیستم توصیه‌گر بر اساس الگوریتم‌های پیچیده و مدل‌های یادگیری ماشین، با تحلیل رفتار کاربران و استفاده از اطلاعات موجود درباره آنها، توصیه‌هایی ارائه می‌دهد. این توصیه‌ها ممکن است بر اساس سابقه مشاهده و علاقه‌های کاربران، محتوای مرتبط و جذاب برای آنها پیشنهاد شود. در صورتی که کاربر در حال جستجوی یک محصول خاص است، محصولات مشابه و مرتبط با آن به او پیشنهاد می‌شود. علاوه بر این بر اساس رفتار کاربرانی که سلیقه و علاقه‌های مشابهی دارند، به کاربران پیشنهاد می‌شود تا با دیدن و استفاده از تجربه‌های آنها، بهترین تجربه را به دست آورند. همچنین ممکن است بر اساس زمان و فعالیت کاربران، توصیه‌هایی در زمان‌های مناسب و بهینه به آنها ارائه شود. با استفاده از این سیستم‌ها، کاربران قادرند به‌راحتی به اطلاعات مورد نیاز خود دسترسی پیدا کنند و از تجربه کاربری بهتری برخوردار شوند. همچنین وب‌سایت‌ها نیز با ارائه توصیه‌های دقیق می‌توانند بازدیدکنندگان بیشتری جذب کنند و سطح رضایت آنها را افزایش دهند. از این رو در این تحقیق یک سیستم توصیه‌گر ترکیبی بر اساس ترکیب خوشه‌بندی ترتیبی فازی و

شبکه خودرمزگذار عمیق بر اساس اطلاعات پروفایل کاربری و رتبه‌بندی وب‌سایت‌ها توسط کاربران ارائه شده است. در این سیستم توصیه‌گر، ابتدا کاربران بر اساس شباهت نظرات خود به صورت ترتیبی خوشه‌بندی می‌شوند. سپس رتبه‌بندی جدید برای کاربران با توجه به تابع عضویت فازی پیش‌بینی می‌شود. در نهایت اطلاعات موجود در پروفایل کاربری و رتبه‌بندی جدید کاربران به هر وب‌سایتی به‌منظور پیش‌بینی رتبه‌بندی وب‌سایت‌ها توسط کاربران، به عنوان ورودی شبکه خودرمزگذار عمیق ارائه‌شده مورد استفاده قرار می‌گیرد. شبکه خودرمزگذار عمیق به پیش‌بینی رتبه‌بندی کاربران جدید نسبت به وب‌سایت‌ها اقدام نموده و از این رتبه‌بندی برای یافتن کاربران مشابه بر اساس روش فیلتر مشارکتی اقدام می‌کند. در نهایت پس از یافتن کاربران مشابه، اقدام به ارائه توصیه بازدید و شخصی‌سازی صفحه وب کاربران جدید بر اساس وب‌سایت‌های مورد علاقه کاربران مشابه می‌نماید. روش ارائه‌شده با استفاده از شبکه‌های عصبی عمیق و با توجه به یادگیری عمیق بر روی اطلاعات موجود در پروفایل کاربری در مجموعه داده TripAdvisor و Yelp، توانسته است با بالاترین دقت رتبه‌بندی محصولات برای کاربران جدید را پیش‌بینی نماید. روش ارائه‌شده با توجه به لایه‌های آموزش عمیق و تکمیل فرایند آموزش در لایه میانی توانسته است از سایر روش‌های طبقه‌بندی از نظر معیارهای ارزیابی بهتر عمل کند. نتایج شبیه‌سازی نشان می‌دهد روش ارائه‌شده از نقطه‌نظر دقت آماری و نسبت توصیه‌های موفق به توصیه‌های مفید، عملکرد مناسبی دارد.

مراجع

[1] H. Ko, S. Lee, Y. Park, and A. Choi, "A survey of recommendation systems: recommendation models, techniques, and application fields," *Electronics*, vol. 11, no. 1, Article ID: 141, Jan.-1 2022.

[2] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *Journal of Big Data*, vol. 9, Article ID: 59, 2022.

[3] S. Ephina. Thendral and C. Valliyammai, "Understanding personalization of recommender system: A domain perspective," *International Journal of Applied Engineering Research*, vol. 13, no. 15, pp. 2422-12428, 2018.

[4] A. Papagrigoriou, C. Panagiotakis, E. Kosmas, and P. Fragopoulou, "Collaborative filtering recommender systems taxonomy," *Knowledge and Information Systems*, vol. 64, pp. 35-74, 2022.

[5] S. Eliyas and P. Ranjana. "Recommendation systems: Content-based filtering vs collaborative filtering" in *Proc. 22nd. Int. Conf. on Advance Computing and Innovative Technologies in Engineering*, pp. 1360-1365, Greater Noida, India, 28-29 Apr. 2022.

[6] M. Yu, T. Quan, Q. Peng, X. Yu, and L. Liu, "A model-based collaborate filtering algorithm based on stacked AutoEncoder," *Neural Computing and Applications*, vol. 34, no. 4, pp. 2503-2511, Feb. 2022.

[7] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*, vol. 16, no. 3, pp. 261-273, Nov. 2015.

[8] G. Behera and N. Nain. "Trade-off between memory and model-based collaborative filtering recommender system," in *Proc. Int. Conf. on Paradigms of Communication, Computing and Data Sciences*, pp. 137-146, Kurukshetra, India, 7-9 May 2022.

[9] A. Darvishy, H. Ibrahim, F. Sidi, and A. Mustapha, "HYPNER: A hybrid approach for personalized news recommendation," *IEEE Access*, vol. 8, pp. 46877-46894, 2020.

[10] Q. Shambour, "A deep learning based algorithm for multi-criteria recommender systems," *Knowledge-Based Systems*, vol. 211, Article ID: 106545. Jan. 2021.

[11] Y. Gulzar. A. A. Alwan, R. M. Abdullah, A. Z. Abualkishik, and M. Oumrani, "OCA: Ordered clustering-based algorithm for e-commerce recommendation system," *Sustainability*, vol. 15, no. 4, Article ID: 2947, Feb.-2 2023.

[12] S. Lee and D. Kim, "Deep learning based recommender system using cross convolutional filters," *Information Sciences*, vol. 592, pp. 112-122, May 2022.

[13] L. El Youbi El Idrissi, I. Akharraz, and A. Ahaitouf, "Personalized e-learning recommender System Based on Autoencoders" *Applied System Innovation*, vol. 6, no. 6, Article ID: 102, Dec. 2023.

[14] P. Bellini, L. Alessandro Ipsaro Palesi, P. Nesi, and G. Pantaleo, "Multi clustering recommendation system for fashion retail," *Multimedia Tools and Applications*, vol. 82, pp. 9989-10016, 2023.

[15] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Expert Systems with Applications*, vol. 167, Article ID: 114324, Apr. 2021.

[16] H. Khatter, S. Arif, U. Singh, S. Mathur, and S. Jain, "Product recommendation system for e-commerce using collaborative filtering and textual clustering," in *Proc. 3rd. Int. Conf. on Inventive Research in Computing Applications*, pp. 612-618. Coimbatore, India, 2-4 Sept. 2021.

[17] T. K. Dang, Q. P. Nguyen, and V. S. Nguyen, "A study of deep learning-based approaches for session-based recommendation systems," *SN Computer Science*, vol. 1, Article ID: 216, 2020.

[18] R. J. Ziarani and R. Ravanmehr, "Deep neural network approach for a serendipity-oriented recommendation system," *Expert Systems with Applications*, vol. 185, Article ID: 115660, Dec. 2021.

[19] Yelp Inc., *Yelp Open Dataset*, Accessed 2024. https://business.yelp.com/data/resources/open-dataset/

[20] M. H. Alam, W. -J. Ryu, and S. Lee, *TripAdvisor Dataset*. 2016, Accessed 2024. https://brightdata.com/products/datasets/tripadvisor
R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, and X. Li, "Efficient user profiling based intelligent travel recommender system for individual and group of users," *Mobile Networks and Applications*, vol. 24, pp. 1018-1033, 2019.

مستوره معینی تحصیلات کارشناسی خود را در سال ۱۳۸۹ از دانشگاه آزاد اسلامی واحد ماهشهردر حوزه IT، مدرک کارشناسی ارشد خود را در سال ۱۳۹۵ از دانشگاه آزاد اسلامی واحد علوم تحقیقات، در رشته مهندسی کامپیوتر گرایش نرم افزار، و مدرک دکترای خود را در رشته مهندسی کامپیوتر گرایش نرم افزار در سال ۱۴۰۳از دانشگاه آزاد اسلامی واحد تهران جنوب دریافت کرد. در این سال‌ها او چندین کتاب و مقاله در حوزه کامپیوتر منتشر کرده است. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: هوش مصنوعی، داده کاوی و سیستم‌های توصیه‌گر.

علی برومندنیا متولد خطه اصفهان، ایران است. او مدرک کارشناسی خود را در سال ۱۳۷۱ از دانشگاه صنعتی اصفهان، مدرک کارشناسی ارشد خود را در سال ۱۳۷۴ از دانشگاه علم و صنعت ایران، هر دو در رشته مهندسی معماری کامپیوتر، و مدرک دکترای خود را در رشته هوش مصنوعی و مهندسی کامپیوتر از دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران در سال ۱۳۸۵ دریافت کرد. از سال ۱۳۷۱ تا ۱۳۷۴، او بر روی کنترل هوشمند حمل و نقل با پردازش تصویر کار کرد و سیستم تشخیص خودکار پلاک خودرو را برای شرکت کنترل ترافیک تهران طراحی نمود. ایشان بیش از صد کتاب، مجله و مقاله کنفرانسی در حوزه کامپیوتر منتشر کرده است. نام‌برده به پنهان‌سازی اطلاعات، امنیت چندرسانه‌ای، تشخیص و قطعه‌بندی کاراکترهای فارسی و عربی، قطعه‌بندی اسناد فارسی و عربی، تصویربرداری پزشکی، پردازش سیگنال و تصویر و تحلیل موجک علاقه‌مند است. ایشان داور برخی از مجلات و کنفرانس‌های بین‌المللی است و در حال حاضر دانشیار دانشکده هوش مصنوعی دانشگاه آزاد اسلامی واحد تهران جنوب است.

مونا مرادی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد کامپیوتر به‌ترتیب در سال‌های ۱۳۷۹ و ۱۳۸۳ در دانشگاه آزاد اسلامی به پایان رسانده است. مدرک دکتری ایشان نیز در همین حوزه می‌باشد. وی اکنون استادیار دانشکده مهندسی کامپیوتر دانشگاه آزاد اسلامی می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: هوش مصنوعی، معماری کامپیوتر، طراحی مدارهای محاسباتی و نانو الکترونیک.