مقاله پژوهشی

بهینه‌سازی و پیش‌بینی برنامه‌های مورد علاقه کاربران با استفاده از رویکرد فیلترینگ مشارکتی و الگوریتم فاخته

رضا مولایی فرد، جواد محمدزاده و پیام یاراحمدی

چکیده: در این تحقیق به ارائه روشی به منظور بهبود سیستم‌های توصیه‌گر برنامه موبایل با استفاده از فیلترینگ مشارکتی و الگوریتم فراابتکاری فاخته پرداخته می‌شود که برای خوشه‌بندی داده‌ها از الگوریتم SW-DBSCAN استفاده شده که این الگوریتم توانست میزان کارایی ۹۹٪ را در بخش خوشه‌بندی به دست آورد که از سایر الگوریتم‌های مشابه توانست عملکرد بهتری را به دست آورد. همچنین برای بهینه‌سازی داده‌ها از الگوریتم فاخته استفاده شده که این الگوریتم توانست عملکرد ۹۸٪ را به دست آورد و نسبت به الگوریتم‌های کرم شب‌تاب، گرگ خاکستری و بهینه‌سازی ذرات عملکرد بهتری را به دست آورد. برای قسمت پیش‌بینی نیز از الگوریتم شبکه‌های عصبی استفاده شده که در این بخش نیز توانست عملکرد قابل قبولی نسبت به سایر الگوریتم‌های مشابه به دست آورد که در نهایت با استفاده از سیستم توصیه‌گر مبتنی بر فیلترینگ مشارکتی این اطلاعات در اختیار کاربر قرار می‌گیرد. یکی از مشکلاتی که سیستم‌های توصیه‌گر با آن مواجه هستند، مشکل داشتن معنای یکسان است. مشکل داشتن معنای یکسان وقتی رخ می‌دهد که یک آیتم با دو یا چند نام نشان داده شود؛ نام‌هایی که معنای مشابهی دارند. در چنین مواردی، سیستم توصیه‌گر نمی‌تواند تشخیص دهد که این نام‌ها نشان‌دهنده اقلام متفاوتی هستند یا همگی به یک آیتم یکسان اشاره دارند که در این روش سعی به برطرف‌کردن آن شد. همچنین طبق تحقیقات صورت‌گرفته، استفاده از این روش پیشنهادی تا ۹۴٪ می‌تواند نیازهای کاربر را به‌درستی تشخیص داده و پیشنهادهای مناسبی را به کاربر پیشنهاد دهد.

کلیدواژه: **سیستم توصیه‌گر، برنامه موبایل، فیلترینگ مشارکتی، الگوریتم فاخته، داده‌کاوی.**

۱- مقدمه

با توجه به رشد روزافزون اپلیکیشن‌های موجود در سطح وب، وجود سیستمی که بتواند اطلاعات مورد نیاز کاربران را از میان حجم عظیم داده‌ها که روزبه‌روز نیز در حال افزایش می‌باشند، استخراج و به کاربر توصیه کند، لازم و ضروری به نظر می‌رسد. یکی از بهترین و مهم‌ترین راه‌ها برای انجام این کار استفاده از سیستم‌های توصیه‌گر می‌باشد. یک سیستم توصیه‌گر با قابلیتی که در جمع‌آوری اطلاعات مربوط به سلایق، علایق و اولویت‌های کاربران و دسته‌بندی و تفسیر آنها دارد، امکانی را فراهم می‌آورد که کاربران با صرف زمان و انرژی کمتر به اطلاعات مورد نظر خود دسترسی پیدا کنند [۱]. سیستم‌های توصیه‌گر یا پیشنهاددهنده زیرمجموعه‌ای از سامانه پالایش اطلاعات می‌باشد که به دنبال پیش‌بینی امتیاز یا اولویتی است که کاربر به یک آیتم (داده، اطلاعات، کالا و ...) خواهد داد. در سال‌های اخیر سیستم‌های توصیه‌گر بسیار متداول شده و در حوزه‌های مختلفی مورد استفاده قرار گرفته‌اند. برخی از کاربردهای معروف این سیستم‌ها در ادامه آمده‌اند: موسیقی، صفحات وب، اخبار، برنامه و بازی‌های موبایل، کتاب‌ها و مقالات، جستجو و شبکه‌های اجتماعی. سیستم توصیه‌گر ازجمله ابزارهایی است که می‌تواند کاربران را در محیط‌های الکترونیکی به سمت یافتن اطلاعات، خدمات و آیتم‌های مورد نظرشان هدایت کند [۲]. سیستم‌های توصیه‌گر با قابلیت کشف علایق کاربران و پیش‌بینی اولویت آنها، آیتم‌هایی را که احتمال می‌رود مورد توجه کاربر باشند، از بین حجم بالای داده‌ها پالایش کرده و آنها را به کاربر پیشنهاد دهد که می‌تواند باعث صرفه جویی در وقت شود. از طرف دیگر این سیستم‌ها توانایی ذخیره و تحلیل رفتارهای گذشته کاربر، خدمات و اطلاعاتی را که مورد توجه کاربران نبوده و احتمالاً به آنها علاقه‌مند هستند را نیز استنتاج کرده و نتایج جالب توجهی به کاربران ارائه می‌کند. در واقع سیستم‌های توصیه‌گر یکی از ابزارهای اصلی غلبه بر مشکل افزونگی اطلاعات بوده و با داشتن قدرت تحلیل رفتارهای کاربر، مکمل هوشمندی برای مفاهیم بازیابی و پالایش اطلاعات است. در سیستم‌های توصیه‌گر برای هر کاربر، یک نمایه کاربر توسعه داده می‌شود که شامل اطلاعاتی از قبیل اطلاعات شخصی وی، اقلام بازدیدشده توسط وی، رتبه‌بندی‌های صورت‌گرفته توسط وی و اقلام خریداری‌شده و دانلودشده توسط وی است و این نمایه با این اطلاعات به‌روزرسانی می‌شود. به‌طور مشابه، هر آیتم دارای ویژگی‌ها و خاصیت‌های منحصربه‌خود است که این موارد از طریق تعامل کاربر با سیستم به‌طور ضمنی یا به‌صورت صریح به‌روزرسانی می‌شود. این امکان وجود ندارد که کاربر تمام اقلام موجود در کاتالوگ را رتبه‌بندی کند و بدین ترتیب برخی از اقلام به هیچ وجه رتبه‌بندی نمی‌شوند؛ بنابراین یک سیستم توصیه‌گر باید از روش‌های مناسبی برای فیلترکردن استفاده کند تا چنین اقلامی نیز در نهایت به‌طور ضمنی یا به‌صورت صریح رتبه‌بندی شوند و در نتیجه مورد پیشنهاد قرار بگیرند [۳] و [۴]. یکی از مشکلات اساسی سیستم‌های توصیه‌گر، مشکل داشتن معنای یکسان است. این مشکل وقتی رخ می‌دهد که یک آیتم با دو یا چند نام نشان داده شود؛ نام‌هایی که معنای مشابهی دارند. در چنین مواردی، سیستم توصیه‌گر نمی‌تواند تشخیص دهد این نام‌ها نشان‌دهنده اقلام متفاوتی هستند یا همگی به یک آیتم یکسان اشاره دارند. امکان وجود انواع این گونه تغییرات در هنگام استفاده از شرح

این مقاله در تاریخ ۲۶ آبان ماه ۱۴۰۳ دریافت و در تاریخ ۶ فروردین ماه ۱۴۰۴ بازنگری شد.

رضا مولایی فرد (نویسنده مسئول)، گروه مهندسی کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران، (email: molaefard@gmail.com).

جواد محمدزاده، گروه مهندسی کامپیوتر، دانشکده هوش مصنوعی، واحد کرج، دانشگاه آزاد اسلامی، کرج، ایران، (email: mohamadzadeh@kiau.ac.ir).

پیام یاراحمدی، گروه مهندسی کامپیوتر، دانشگاه ملی تاراس شوچنکو کیف، کیف، اوکراین، (email: yarahmadi.pay@gmail.com).

و توصیف اقلام بسیار بیش از حد معمول است و استفاده زیاد از واژگان مترادف و هم‌معنا باعث کاهش عملکرد سیستم‌های توصیه‌گر فیلترکننده مشارکتی می‌شود. از آنجا که محتویات اقلام به‌طور کامل نادیده گرفته می‌شوند، سیستم توصیه‌گر وجود ارتباط مخفی بین اقلام را در نظر نمی‌گیرد. به همین دلیل است که اقلام جدید تا زمانی که توسط کاربران رتبه‌بندی نشده‌اند، پیشنهاد داده نمی‌شوند [۵]. برای رفع مشکل داشتن معنای یکسان، از روش پیش‌بینی توسط شبکه عصبی و بهینه‌سازی اطلاعات توسط الگوریتم فراابتکاری فاخته استفاده کردیم تا بتوانیم مشکل موجود را حل کنیم و سیستم توانست عملکرد قابل قبولی برای غلبه بر مشکل موجود و پیش‌بینی صحیح برنامه‌ها به کاربران بپردازد. هدف این پژوهش بهینه‌سازی پیشنهادهای مربوط به توصیه برنامه‌ها و بازی‌های موبایل با توجه به شخصیت کاربر و همچنین درصد دقت و میزان فراخوانی بالای توصیه‌ها به کاربر می‌باشد.

ساختار مقاله به این صورت است: بخش ۲ به بررسی سایر روش‌های مشابه و کارهای پیشین که در این زمینه صورت گرفته‌اند، می‌پردازد. در بخش ۳ مدل سیستم و روش پیشنهادی در طی چند مرحله اصلی که شامل پیش‌پردازش داده‌ها و خوشه‌بندی داده‌ها می‌باشد، آمده است. در بخش ۴ به پیش‌بینی داده‌ها با استفاده از الگوریتم شبکه‌ها عصبی پرداخته می‌شود. در بخش ۵، سیستم توصیه‌گر برای توصیه به کاربر و در بخش ۶ شبیه‌سازی روش پیشنهادی و نتایج به‌دست‌آمده آورده شده است. نهایتاً مقاله با نتیجه‌گیری در بخش ۷ خاتمه می‌یابد.

۲- پیشینه پژوهش

ابوصالح و همکاران در مقاله‌ای در سال ۲۰۲۱ به ارائه روشی به منظور توصیه برنامه‌های موبایل برای کاربران پرداختند. این محققان در مقاله خود سعی کردند که با توجه به عرصه توسعه برنامه‌های تلفن همراه و افزایش پیچیدگی‌های فناوری تلفن‌های هوشمند، توصیه‌های کاربردی را بر روی سیستم عامل موبایل ارائه دهند. در روش پیشنهادی از یک سیستم تلفیقی و سریع برای توصیه نرم‌افزارها به کاربران استفاده کردند. در روش پیشنهادی از یک ماژول استنتاج دانش دامنه شامل وب معنایی مختلف حاصل از فناوری و استفاده از سبک هستی‌شناسی با استفاده از پروفایل کاربر استفاده شده است. در این روش از مدل‌سازی کاربر چندبعدی و پرس‌وجو برای بهبود و افزایش نتایج و توصیه به کاربران استفاده گردیده است. نتایج حاصل از این تحقیق حاکی از بهبود توصیه در این روش می‌باشد [۶].

یانگ و همکاران در مقاله‌ای در سال ۲۰۲۵ به ارائه یک سیستم توصیه‌گر برای پیشنهاد اپ‌های موبایل پرداختند. این مطالعه یک سیستم توصیه پیشرفته را ارائه می‌کند که معماری حافظه کوتاه‌مدت، بلندمدت (LSTM) تقویت‌شده توجه را با یک شبکه عصبی بازگشتی دوطرفه برای بهینه‌سازی استخراج ترجیحات کاربر ترکیب می‌کند. مکانیسم توجه، چارچوب سنتی LSTM را با اولویت‌بندی نقاط داده‌های زمانی بحرانی غنی می‌کند؛ در نتیجه هم دقت پردازش ترجیحی بلندمدت و هم قابلیت تفسیر مدل را بهبود می‌بخشد [۷].

یو و همکاران در مقاله‌ای در سال ۲۰۲۵ به تقویت توصیه‌های اپلیکیشن موبایل با داده‌های آموزشی جمع‌سپاری‌شده با استفاده از یادگیری ماشینی و یادگیری عمیق پرداختند. در چشم‌انداز دیجیتالی که به‌سرعت در حال تحول است، توصیه‌های شخصی برای افزایش تجربه کاربر ضروری شده است. مدل‌های یادگیری ماشینی الگوهای رفتاری کاربر را تجزیه و تحلیل می‌کنند تا محتوای سرگرمی، آموزشی یا تجارت الکترونیک مرتبط را پیشنهاد کنند. دستگاه‌های تلفن همراه جمع‌آوری داده‌های آموزشی را از طریق جمع‌سپاری آسان‌تر می‌کنند که فرصت‌های جدیدی را برای بهبود الگوریتم‌های توصیه اپلیکیشن باز می‌کند. این مقاله روش‌های ارزشمندی را برای توصیه‌های دانشجویی مقیاس‌پذیر و سیستم‌های آموزشی ارائه می‌کند و مزایای DL را نسبت به CF در مدیریت مجموعه داده‌های پراکنده و حساس به زمان برجسته می‌کند. هدف این مطالعه آن است که برنامه‌ها را بر اساس الگوهای استفاده از برنامه به دانشجویان دانشگاه بر اساس دسته‌بندی توصیه کند. داده‌ها برای ارزیابی این ۸۰۶ دانشجوی دانشگاه برای آموزش مدل‌های فیلتر مشارکتی (CF) و یادگیری عمیق معاصر (DL) مورد استفاده قرار گرفت. نتایج نشان می‌دهند واحدهای بازگشتی دروازه‌ای (GRU) بهترین گزینه برای پیشنهادهای سفارشی‌شده در زمان واقعی هستند؛ زیرا ظرفیت آنها برای شبیه‌سازی تعاملات متوالی و تنظیم با تغییر رفتار کاربر است [۸].

ندیم و سلیمان در مقاله خود در سال ۲۰۲۴ به توسعه یک سیستم توصیه‌کننده شخصی برای کاربران پیشرفته تجربه در زمینه اپلیکیشن‌های موبایل پرداختند. این مقاله از یک چهارچوب مبتنی بر توصیه استفاده می‌کند؛ بدین صورت که این چارچوب از مدل‌های یادگیری ماشینی بهره می‌برد که می‌توانند داده‌های کاربر را به طور مؤثر بر روی دستگاه‌های تلفن همراه، پردازش و در عین حال کارایی محاسباتی و دقت پیش‌بینی را متعادل کنند. جنبه‌های کلیدی سیستم شامل تقسیم‌بندی کاربر، تولید توصیه‌های بلادرنگ و حلقه‌های بازخورد پیوسته است که با ترجیحات کاربر در حال تکامل سازگار است. علاوه بر این به ملاحظات حفظ حریم خصوصی برای اطمینان از اینکه داده‌های کاربر به طور ایمن مدیریت می‌شوند و مطابق با استانداردهای حفاظت از داده‌ها مورد استفاده قرار می‌گیرند، توجه می‌شود. این رویکرد مسیری را برای تبدیل برنامه‌های کاربردی تلفن همراه به پلتفرم‌های تعاملی‌تر و شخصی‌سازی‌شده ارائه می‌دهد و استاندارد جدیدی برای تجربه کاربر در اکوسیستم‌های تلفن همراه ایجاد می‌کند [۹].

گونگ و همکاران در مقاله‌ای در سال ۲۰۲۰ به ارائه روشی به منظور توصیه‌های تلفن همراه پرداختند. آنها بر اساس چهارچوب ابر به لبه که به شکل آبشار هستند، جایی که نتایج توصیه‌شده با محاسبه در سرور ابری لبه منتقل می‌شود از این ایده برای روش پیشنهادی خود استفاده کردند. با وجود اثربخشی، پهنای باند شبکه و تأخیر بین سرور ابری و لبه ممکن است باعث تأخیر در بازخورد سیستم و درک کاربر شود؛ پس محاسبه زمان واقعی در حاشیه می‌تواند به ترجیحات کاربر باارزش‌تر کمک کند و بنابراین توصیه‌های رضایت‌بخش‌تری ارائه دهد. این روش اولین تلاش برای طراحی و پیاده‌سازی سیستم توصیه‌کننده جدید بر روی لبه است که به درک کاربر در زمان واقعی و بازخورد سیستم در زمان واقعی دست می‌یابد. همچنین، مدل‌سازی توالی رفتار کاربر ناهمگن و بازآرایی متناسب با توجه به شبکه رفتاری کاربر، پیشنهادهایی ارائه کند تا علایق متنوع کاربر را جلب کرده و نتایج توصیه‌ها را بر این اساس تنظیم کند [۱۰].

پیمیندیس و همکاران در مقاله‌ای در سال ۲۰۱۹ به ارائه روشی برای توصیه برنامه‌ها در موبایل با استفاده از روش‌های داده‌کاوی پرداختند. این محققان در مقاله خود، برنامه‌های کاربردی متعددی مانند تجارت الکترونیک، پلتفرم‌های ویدئویی و شبکه‌های اجتماعی توصیه‌های شخصی را به کاربران خود ارائه می‌دهند که باعث بهبود تجربه کاربر و درآمد فروشندگان شده است. توسعه سیستم‌های توصیه‌گر بیشتر بر پیشنهاد الگوریتم‌های جدید متمرکز شده که توصیه‌های دقیق‌تری ارائه می‌دهند. با این حال، استفاده از دستگاه‌های تلفن همراه و رشد سریع اینترنت و

زیرساخت‌های شبکه، ضرورت استفاده از سیستم‌های توصیه‌گر تلفن همراه را به همراه داشته است. پیوندهای بین سیستم‌های توصیه‌کننده وب و تلفن همراه با نحوه بهبود توصیه‌ها در محیط‌های تلفن همراه توضیح داده شده است. این کار بر شناسایی پیوندهای بین وب و سیستم‌های توصیه‌کننده تلفن همراه متمرکز است و جهت‌های محکمی را در آینده ارائه می‌دهد که هدف آنها منجر به یک حوزه یکپارچه توصیه تلفن همراه می‌شود [۱۱].

ین و همکاران در مقاله‌ای در سال ۲۰۱۸ به بررسی روش‌های توصیه برنامه موبایل به کاربران پرداختند. با توجه به پراکندگی شدید ماتریس برنامه کاربر و بسیاری از برنامه‌های تازه‌ظهور، چالش‌های شدیدی ایجاد می‌کند و باعث می‌شود که روش‌های مبتنی بر فیلترینگ مشارکتی در عملکرد توصیه‌های آنها به میزان قابل توجهی کاهش یابد. علاوه بر این برخلاف موارد سنتی، برنامه‌ها حق دسترسی به منابع شخصی کاربران را دارند که ممکن است منجر به خطر امنیت یا نشت حریم خصوصی شود؛ بنابراین انتخاب برنامه‌ها توسط کاربران نه‌تنها از علایق شخصی آنها، بلکه از ترجیحات حریم خصوصی آنها نیز متأثر است. علاوه بر این، ترجیحات حریم خصوصی کاربر با دسته برنامه متفاوت است. با توجه به چالش‌های فوق، در این روش یک مدل تولیدکننده افزودنی پراکنده تلفن همراه پیشنهاد شد تا با در نظر گرفتن علایق کاربر و ترجیحات حریم خصوصی کاربر در این مقاله، برنامه‌ها را توصیه کند. در این روش مجموعه داده‌های مقیاس بزرگ و واقعی از ۳۶۰-App store بزرگ‌ترین پلتفرم برنامه Android در چین جمع‌آوری شده و آزمایش‌های گسترده‌ای بر روی آن انجام داده شده است. نتایج تجربی نشان می‌دهند که روش پیشنهادی به طور مداوم و قابل توجهی از سایر روش‌های موجود پیشرفته پیشی می‌گیرد که بر اهمیت بهره‌برداری از ترجیحات حریم خصوصی کاربر آگاه از طبقه و داده‌های محتوای چندمنظوره برنامه در توصیه برنامه شخصی‌شده دلالت دارد [۱۲].

مشکل داشتن معنای یکسان وقتی رخ می‌دهد که یک آیتم با دو یا چند نام نشان داده شود، نام‌هایی که معنای مشابهی دارند [۱۳]. در چنین مواردی، سیستم توصیه‌گر نمی‌تواند تشخیص دهد این نام‌ها نشان‌دهنده اقلام متفاوتی هستند یا همگی به یک آیتم یکسان اشاره دارند. به عنوان مثال، یک رویکرد فیلترکردن مشارکتی مبتنی بر حافظه با دو عنوان «comedy movie» و «comedy film» به صورت متفاوت برخورد خواهد نمود، در صورتی که هر دو به معنای فیلم کمدی هستند. امکان وجود انواع این گونه تغییرات در هنگام استفاده از شرح و توصیف اقلام بسیار بیش از حد معمول است و استفاده زیاد از واژگان مترادف و هم‌معنا باعث کاهش عملکرد سیستم‌های توصیه‌گر فیلترکننده مشارکتی می‌شود. از آنجا که محتویات اقلام به طور کامل نادیده گرفته می‌شوند، سیستم توصیه‌گر وجود ارتباط مخفی بین اقلام را در نظر نمی‌گیرد. به همین دلیل است که اقلام جدید تا زمانی که توسط کاربران رتبه‌بندی نشده‌اند، پیشنهاد داده نمی‌شوند. برای رفع مشکل داشتن معنای یکسان، روش‌های مختلفی از جمله هستی‌شناسی [۱۴]، روش‌های تجزیه مقدار تکین [۱۵] و شاخص‌گذاری معنای مخفی مورد استفاده قرار بگیرند.

در مسائلی که اطلاعات خوبی از کاربر (مثل سن، جنسیت، شغل و ...) وجود دارد اما در مورد آیتم‌ها دچار کمبود اطلاعات هستیم یا به دست آوردن ویژگی‌های آیتم‌ها دشوار است، سیستم‌های توصیه‌گر فیلتر مشارکتی بسیار مناسب هستند. در این روش بر اساس شباهت رفتاری و عملکردی کاربرانی که در گذشته الگوی رفتاری مشابهی با کاربر فعلی داشته‌اند، پیشنهادها ارائه می‌شود. روش فیلتر مشارکتی بر این فرض استوار است که کاربرانی که یک سری نظرهای مشابه درباره یک آیتم دارند، درباره آیتم‌های دیگر هم نظرهای مشابه دارند. فیلتر مشارکتی خود به دو دسته تقسیم می‌شود: مبتنی بر کاربر و مبتنی بر آیتم [۱۶].

۳- روش پیشنهادی

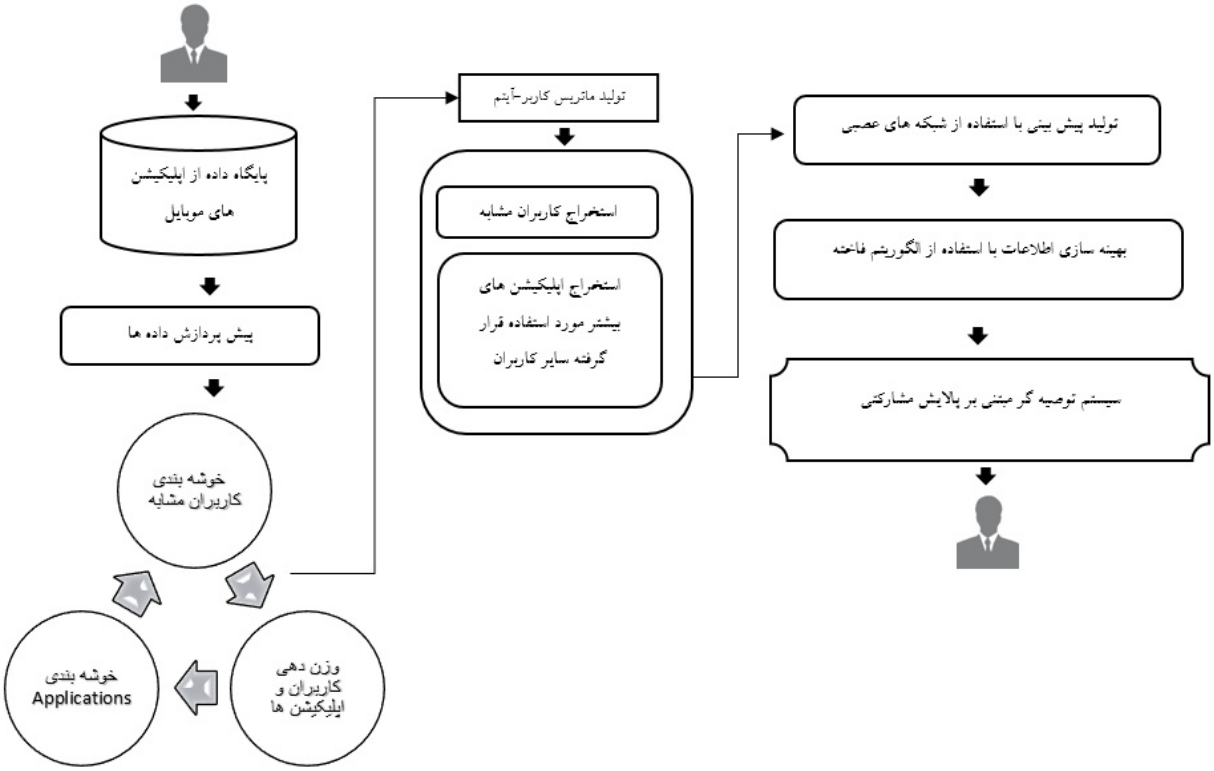
در روش پیشنهادی به ارائه روش جدیدی به منظور بهبود سیستم‌های توصیه‌گر در زمینه توصیه اپلیکیشن‌های مورد علاقه کاربران به کاربران پرداخته می‌شود. روش پیشنهادی بدین صورت می‌باشد که ما ابتدا یک پایگاه داده از جستجوهای مربوط به کاربران را جمع‌آوری می‌کنیم. سپس با استفاده از الگوریتم‌های داده‌کاوی، سعی در استخراج موارد مورد علاقه کاربر و سپس پیشنهاد این داده‌ها به کاربر داریم. رویکرد پیشنهادی سعی دارد تا منابع یادگیری را به فرد یادگیرنده با در نظر گرفتن ترجیحات وی و تاریخچه جستجوهای قبلی وی و سایر کاربران پیشنهاد دهد؛ تاریخچه‌ای که از فایل‌های جستجوشده کاربر و سایر کاربران استخراج گردیده است. این رویکرد، سبک‌های یادگیری و روش‌های فیلترکردن مشارکتی را ترکیب می‌کند تا کیفیت پیشنهادها را ارتقا دهد. ما همچنین از رویکرد فیلترکردن مشارکتی برای بازبینی این پیشنهادهای اولیه استفاده کرده‌ایم. ایده این روش، ایجاد پیش‌بینی‌هایی در مورد ترجیحات فرد یادگیرنده بر اساس ترجیحات افراد دیگری است که مشابه با فرد یادگیرنده فعال هستند. در واقع با در نظر گرفتن پروفایل افراد یادگیرنده، گروهی از افراد یادگیرنده شناسایی می‌شوند که ترجیحات آنها با فرد یادگیرنده فعلی مطابقت دارند. سپس پروفایل فرد یادگیرنده با گروه‌های متفاوت افراد یادگیرنده مقایسه می‌شوند که این گروه‌ها پیش از این دسته‌بندی شده‌اند و مناسب‌ترین همسایگی شناسایی می‌شود. شکل ۱ نمایی از روش پیشنهادی را نشان می‌دهد.

۳-۱ داده‌ها

مجموعه داده‌های مورد استفاده در این تحقیق شامل یک پایگاه داده از موارد مورد جستجو قرار گرفته کاربران از تعدادی از اپلیکیشن‌های موجود در سطح وب می‌باشد. این پایگاه داده شامل تعداد ۸ هزار جستجو و دانلود اپلیکیشن‌های موبایل است که از این مجموعه پایگاه داده‌ای در حدود ۸۵۰ مورد به صورت تصادفی استخراج شده که از این پایگاه داده تصادفی برای قسمت ارزیابی و پیاده‌سازی روش پیشنهادی استفاده گردیده است. نتایج حاصل از ارزیابی و آنالیز مربوط به این پایگاه داده در ادامه نمایش داده شده است.

۳-۲ پیش‌پردازش داده‌ها

در اولین مرحله از روش پیشنهادی، ابتدا باید عملیات پیش‌پردازش داده‌ها را انجام دهیم. اگر داده‌ها را قبل از عملیات داده‌کاوی مورد پیش‌پردازش قرار دهیم عملکرد قابل قبول و قابل اعتمادتری به دست خواهیم آورد. از آنجا که نمی‌توانیم داده‌های خام را به الگوریتم‌های داده‌کاوی تزریق کنیم، لازم است که آنها را از شکل و حالت اولیه خود خارج کرده و به فرمی مناسب برای الگوریتم‌های داده‌کاوی تبدیل کنیم [۱۷]. الگوریتمی که برای پیش‌پردازش داده‌ها از آن استفاده می‌کنیم الگوریتم Feature selection یا انتخاب ویژگی می‌باشد. این الگوریتم به انتخاب زیرمجموعه ویژگی، فرایند شناسایی و تا حد امکان به حذف داده‌های نامرتبط و اضافی می‌پردازد. با استفاده از این الگوریتم ابعاد داده‌ها کاهش داده شده و الگوریتم‌های قادر خواهد بود تا عملکرد سریع‌تر



شکل ۱: نمایی از روش پیشنهادی.

و مؤثرتری داشته باشند. به طور کلی این الگوریتم دارای ۳ ویژگی است:

- ویژگی‌های مرتبط: ویژگی‌هایی که بر خروجی تأثیر می‌گذارند و بقیه نمی‌توانند نقش آنها را بر عهده بگیرند.
- ویژگی‌های نامرتبط: ویژگی‌هایی‌اند که هیچ تأثیری بر روی خروجی ندارند و مقادیر آن به صورت تصادفی برای هر مثال تولید می‌شود.
- ویژگی‌های اضافه: هرگاه یک ویژگی بتواند نقش دیگری را بازی کند (شاید ساده‌ترین راه برای مدل‌سازی افزونگی)، افزونگی وجود دارد.

الگوریتم‌های Feature selection به طور کلی دارای دو جزء هستند: یک الگوریتم انتخاب که زیرمجموعه‌های پیشنهادی از ویژگی‌ها را ایجاد می‌کند و سعی دارد که یک زیرمجموعه بهینه پیدا کند و یک الگوریتم ارزیابی که تعیین می‌کند«زیر» بودن یک زیرمجموعه ویژگی پیشنهادی چقدر خوب است و میزان خوبی را به الگوریتم انتخاب برمی‌گرداند. با این حال، بدون معیار توقف مناسب، فرایند انتخاب ویژگی ممکن است به طور کامل یا برای همیشه در فضای زیرمجموعه‌ها اجرا شود. معیارهای توقف می‌توانند عبارتند از اینکه آیا افزودن (یا حذف) هر ویژگی زیرمجموعه بهتری ایجاد نمی‌کند و اینکه آیا یک زیرمجموعه مطلوب با توجه به برخی از تابع ارزیابی به دست می‌آید.

۳-۳ خوشه‌بندی داده‌ها با الگوریتم SW-DBSCAN

در مرحله بعدی به‌منظور پیداکردن شباهت بین داده‌ها می‌باید که داده‌های خود را خوشه‌بندی کنیم. در سیستم‌های توصیه‌گر می‌توان از خوشه‌بندی برای دسته‌بندی‌کردن داده‌های مشابه با هم استفاده کرد تا جستجو و ساخت پیشنهادها فقط بر روی خوشه‌هایی انجام گردد که پیش‌بینی می‌شود و برای کاربر جالب‌تر خواهد بود. در نتیجه از تکنیک‌های خوشه‌بندی برای افزایش مقیاس‌پذیری الگوریتم‌های پشتیبان سیستم‌های توصیه‌گر استفاده گردیده است. این مسئله به‌خصوص در شرایطی که برای پیشنهاد آیتمی لازم است تمام فضای مسئله جستجو شود در کارایی الگوریتم تأثیرات مثبتی نشان داده است. روش خوشه‌بندی مورد استفاده در تحقیق، استفاده از نسخه بهبودیافته الگوریتم خوشه‌بندی DBSCAN یعنی SW-DBSCAN می‌باشد. SW-DBSCAN یک روش خوشه‌بندی مبتنی بر شبکه است که برای کاهش پیچیدگی زمانی و افزایش کارایی DBSCAN جهت به دست آوردن خوشه‌های دقیق‌تر می‌باشد که ما در این تحقیق خوشه‌های خود را از طریق این الگوریتم، خوشه‌بندی می‌کنیم. در ابتدا ما شبکه را به سلول‌های مختلف تقسیم می‌کنیم و سپس الگوریتم DBSCAN را با توجه به تعداد نقاط هر سلول بر روی سایر سلول‌ها اجرا می‌کنیم. سپس مختصات شبکه را تغییر می‌دهیم. اکنون یک شبکه جدید داریم که در آن همان روش قبل را بر روی آن اعمال خواهیم کرد. اکنون ما دو شبکه داریم. مرحله بعدی ادغام خوشه‌های حاصل در شبکه اول با توجه به شبکه دوم است. برای انجام‌دادن این کار، ما یک ماتریس ایجاد می‌کنیم که به ما کمک می‌کند تا مشخص کنیم کدام دسته‌ها باید ادغام شوند. سرانجام، ما خوشه‌های اولین شبکه را بر اساس ماتریس تولیدشده ادغام می‌کنیم. در نتیجه خوشه‌هایی را به ما می‌دهد که بسیار دقیق می‌باشند که از این داده‌ها می‌توان برای تشخیص علاقه کاربر به گروه خاصی از داده‌ها از آن استفاده کنیم [۱۸].

SW-DBSCAN می‌تواند خوشه‌هایی با شکل دلخواه پیدا کند. حتی می‌تواند خوشه‌ای را پیدا کند که به طور کامل توسط یک خوشه متفاوت احاطه شده است؛ البته این خوشه‌ها نباید به هم متصل باشند. با این وجود با افزایش پارامتر MinPts خطای حاصل از متصل‌بودن نقاط خوشه‌های مختلف کاهش می‌یابد.

SW-DBSCAN مفهوم نویز را در خود دارد و نسبت به داده‌های پرت مقاوم است.

SW-DBSCAN فقط به دو پارامتر نیاز دارد و عمدتاً به ترتیب نقاط در پایگاه داده حساس نیست (با این حال، نقاطی که در مرز دو خوشه مختلف قرار دارند، ممکن است در صورت تغییر ترتیب نقاط، در خوشه

متفاوتی قرار بگیرند).

SW-DBSCAN برای استفاده با پایگاه‌های داده‌ای طراحی شده که می‌تواند پرس‌وجوهای با توجه به منطقه‌های فضای داده‌ها را تسریع کند [۱۹] تا [۲۲].

۳-٤ وزن‌دهی داده‌ها

در مرحله بعدی باید داده‌های خود را وزن‌دهی کنیم. ما برای این منظور، وزن رتبه‌دهی به هر فعالیت یادگیری را با استفاده از تابع امتیاز زیر تعریف نموده‌ایم.

۳-۵ استخراج کاربران مشابه

پس از اینکه کاربر جدید خوشه یا کلاس تعیین گردید، همسایگان آنکه شامل کاربران موجود در آن خوشه می‌باشد، استخراج می‌شوند. نظرات این همسایگان در پیشنهاد نهایی اپلیکیشن‌ها به کاربر جدید مؤثر هستند؛ اما همه همسایگان به یک اندازه به کاربر جدید شباهت ندارند و باید از یک معیار شباهت برای همسایگان نزدیک‌تر استفاده کرد. فرض می‌کنیم کاربران سیستم به‌صورت مجموعه $U = \{u_1, u_2, u_3, \ldots, u_m\}$ با خصوصیات $D = \{d_1, d_2, d_3, \ldots, d_n\}$ و مجموعه اپلیکیشن‌ها به‌صورت $I = \{i_1, i_2, i_3, \ldots, i_k\}$ تعریف شوند. سپس شباهت یک کاربر جدید و هر یک از همسایگان بر اساس رابطه زیر محاسبه می‌گردد

$$d(x,y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \qquad (۱)$$

۳-٦ تولید ماتریس کاربر- آیتم

برای انجام این کار C را به‌عنوان همه کاربران و S را به‌عنوان تمام آیتم‌ها (اپلیکیشن‌ها) که می‌تواند به کاربر پیشنهاد شود در نظر می‌گیریم. تابع سودمندی u، سودمندی آیتم s را برای کاربران c بیان می‌کند. مجموعه کل سفارش‌ها با R نشان داده می‌شود که به‌صورت $C \times R \to R$ تعریف می‌نماییم. سپس برای هر کاربر $c \in C$ آیت مثل $s \in S$ را که سودمندی کاربر را ماکسیمم نماید به‌صورت قراردادی طبق رابطه زیر تعریف می‌نماییم

$$\forall_c \in C \ , \ S_c = \arg\max u(C,S) \qquad (۲)$$

در سیستم توصیه‌گر باید ماتریسی از کاربران و آیتم‌ها ایجاد کنیم. در این روش هر کاربر به آیتمی رأی دهد، آن امتیاز در سلول مورد نظر در ماتریس ذخیره می‌شود. وقتی رأی یک کاربر مشخص شود از آن می‌توان برای تعیین پیشنهاد برای کاربران مشابه استفاده نمود. بدیهی است که این ماتریس تنک و خلوت خواهد بود. یک سیستم توصیه‌گر باید این پراکندگی‌ها را پیش‌بینی کرده و اگر امتیاز پیش‌بینی بالا بود، آن را به کاربر پیشنهاد دهد.

۳-۷ ایجاد نمایه کاربران

سیستم توصیه‌گر برای پیش‌بینی علایق و ارائه توصیه به کاربران، طراحی و پیاده‌سازی شده است. در هر یک از این سیستم‌ها با توجه به حوزه کاری و اهداف، مجموعه‌ای از تکنیک‌های ساخت، به‌روزرسانی و استخراج داده‌ها به کار گرفته شده است، ولی محور اساس در تمامی این سیستم‌ها پروفایل کاربر است. چگونگی ساخت پروفایلی که در ساخت توصیه‌ها استفاده خواهد شد، پروفایل پیش‌فرض سیستم برای کاربران، نحوه به‌روزرسانی اطلاعات پروفایل و منبع این به‌روزرسانی فاکتورهایی هستند که در طراحی یک سیستم توصیه‌گر جایگاه مهمی دارند

$$p(\theta) = EXP(\theta) + \alpha \times IMP(\theta) + \beta \times S(\theta) \qquad (۳)$$

که EXP امتیاز صریح داده‌شده توسط فرد یادگیرنده به هر شیء یادگیری θ است. IMP نیز امتیاز ضمنی و S امتیاز بُعد اجتماعی است. پارامترهای α و β برای نرمال‌سازی توابع IMP و S با واحد ۵ انتخاب شده‌اند. امتیاز صریح به صورت زیر ارائه شده است

$$IMP(\theta) = A(\theta) + B(\theta) + C(\theta) \qquad (۴)$$

A وقتی برابر یک است که در بوکمارک ذخیره شده باشد و در غیر این صورت برابر با صفر می‌باشد. تابع $B(\theta) = 1 - e^{-t}$ که t مدت زمان سپری‌شده توسط فرد یادگیرنده در طول شیء یادگیری θ و C فرکانس دسترسی فعالیت یادگیری است. نهایتاً تابع S که بُعد رتبه‌بندی اجتماعی است به صورت زیر تعریف می‌شود

$$s(\theta) = c \times e^{t} \qquad (۵)$$

و t' مدت زمان سپری‌شده در طول تمام ارتباطات همزمان و غیرهمزمان با استفاده از ابزار ارتباطی و انجمنی و c تعداد مشارکت‌ها و تعاملات با این ابزار می‌باشد.

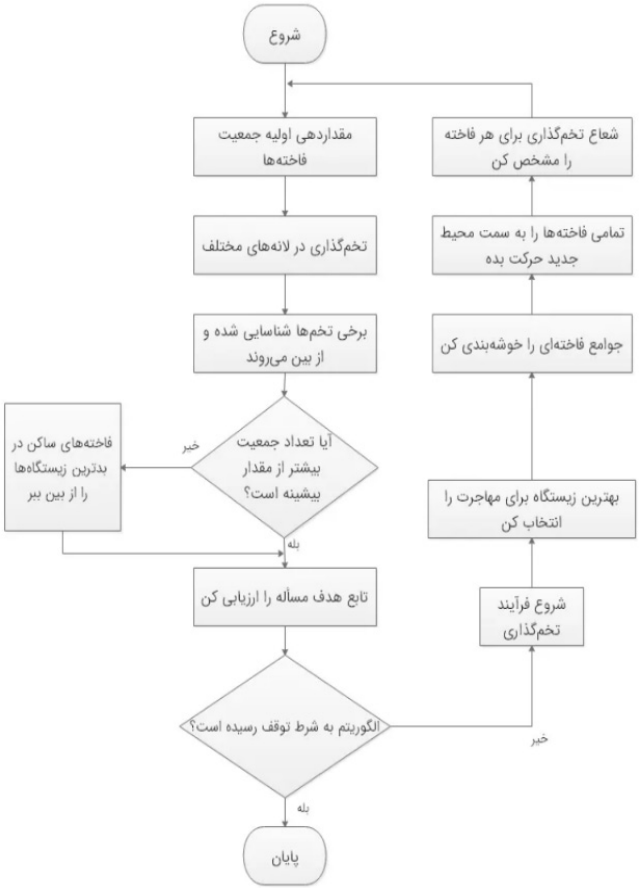
پس از وزن‌دهی به منابع یادگیری، ما یک مدل ترجیح برای هر فرد یادگیرنده به دست می‌آوریم که به صورت یک ماتریس رتبه‌بندی شیء یادگیری فرد یادگیرنده (LLOR)[1] با n ردیف و m ستون تعریف می‌شود که n نشان‌دهنده تعداد افراد یادگیرنده $L = \{l_1, l_2, \ldots, l_n\}$ و m نشان‌دهنده تعداد اشیای یادگیری $J = \{j_1, j_2, \ldots, j_m\}$ است.

این ماتریس از یک مقیاس رتبه‌بندی ۰ تا ۱۰ استفاده می‌کند: ۱۰ بدین معنی است که فرد یادگیرنده به شدت از شیء یادگیری انتخاب‌شده راضی است، ۵ نشان‌دهنده این است که فرد یادگیرنده نسبتاً راضی نیست، ۱ نشان می‌دهد که فرد یادگیرنده از شیء یادگیری به هیچ وجه راضی نیست و نهایتاً امتیاز صفر نیز نشان می‌دهد که شیء یادگیری هنوز به صورت صریح رتبه‌بندی نشده یا اصلاً هنوز مورد استفاده قرار نگرفته است [۲۳] و [۲۴].

در مرحله بعدی با استفاده از الگوریتم فراابتکاری فاخته، داده‌های خود را بهینه خواهیم کرد. الگوریتم فاخته یکی از جدیدترین و قوی‌ترین روش‌های بهینه‌سازی تکاملی است. این الگوریتم از نحوه زندگی پرنده‌ای به نام فاخته بهره گرفته و با یک جمعیت اولیه کار خود را شروع می‌کند. این جمعیت از فاخته، تعدادی تخم دارند که آنها را در لانه تعدادی پرنده میزبان خواهند گذاشت. تعدادی از این تخم‌ها شباهت بیشتری به تخم‌های پرنده میزبان دارند، شانس بیشتری به تخم‌های پرنده میزبان دارند، شانس بیشتری برای رشد و تبدیل‌شدن به فاخته بالغ را خواهند داشت. سایر تخم‌هایی که توسط پرنده میزبان شناسایی می‌شوند، از بین می‌روند. میزان تخم‌های رشدکرده، مناسب‌بودن لانه‌های آن منطقه را نشان می‌دهند. هرچه تخم بیشتری در یک ناحیه، قادر به زیست باشد و نجات پیدا کند، به همان اندازه، سود بیشتری به آن منطقه اختصاص داده می‌شود؛ بنابراین موقعیتی که در آن بیشترین تعداد تخم نجات یابند، پارامتری خواهد بود که فاخته‌ها قصد بهینه‌سازی آن را دارند [۲۵] تا [۲۷]. نمایی از الگوریتم فاخته در شکل ۲ قابل مشاهده است.

- برای حل یک مسئله بهینه‌سازی لازم است تا مقادیر متغیرهای مسئله به فرم یک آرایه شکل گیرند.
- در الگوریتم فاخته آرایه را *Habitat* یا محل سکونت می‌گوییم.

1. Learner Learning Object Rating



شکل ۲: نمایی از الگوریتم فاخته.

- در مسئله بهینه‌سازی N_{var} بعدی یک *Habitat* یک آرایه $1\times var$ خواهد بود که موقعیت فعلی زندگی فاخته‌ها را نشان می‌دهد. این آرایه به شکل زیر تعریف می‌شود

$$Habitat = [x_1, x_2, x_3, \ldots, x_{N_{var}}] \quad (۶)$$

- میزان مناسب‌بودن در *Habitat* فعلی با ارزیابی (f_p) در *Habitat* به دست می‌آید؛ بنابراین

$$Profit = f_p habitat = f_p(x_1, x_2, x_3, \ldots, x_{N_{var}}) \quad (۷)$$

- همان طور که مشاهده می‌شود، الگوریتم فاخته الگوریتمی است که تابع سود را ماکسیمم می‌کند.
- برای استفاده از الگوریتم فاخته برای حل مسائل کمینه‌سازی کافی است یک علامت منفی در تابع هزینه ضرب کنیم.
- برای شروع الگوریتم بهینه‌سازی یک ماتریس *Habitat* با سایز $N_{pop}\times N_{var}$ تولید می‌کنیم.
- سپس برای هر کدام از این *Habitat*ها تعدادی تخم تصادفی تخصیص می‌یابد.
- در طبیعت هر فاخته بین ۵ تا ۲۰ تخم می‌گذارد. این اعداد به‌عنوان حد بالا و حد پایین تخصیص تخم هر فاخته در تکرارهای مختلف استفاده می‌شوند.
- دیگر عادت فاخته حقیقی این است که آنها در یک دامنه مشخص تخم‌های خود را می‌گذارند.

٤- تولید پیش‌بینی با استفاده از شبکه‌های عصبی

در مرحله بعدی می‌باید به تولید پیش‌بینی برای سیستم پیشنهادی بپردازیم. روش مورد استفاده برای انجام‌دادن این کار استفاده از شبکه عصبی می‌باشد. عنصر کلیدی این روش، ایجادکردن ساختارهایی جدید برای سامانه پردازش اطلاعات است. این سیستم از شمار زیادی عناصر پردازشی فوق‌العاده به‌هم‌پیوسته با نام نورون تشکیل شده که برای حل یک مسأله با هم هماهنگ عمل می‌کنند و توسط سیناپس‌ها (ارتباطات الکترومغناطیسی) اطلاعات را منتقل می‌کنند. در این شبکه‌ها اگر یک سلول آسیب ببیند بقیه سلول‌ها می‌توانند نبود آن را جبران کرده و نیز در بازسازی آن سهیم باشند. این شبکه‌ها قادر به یادگیری هستند [۲۸] و [۲۹]. مثلاً با اعمال سوزش به سلول‌های عصبی لامسه، سلول‌ها یاد می‌گیرند که به طرف جسم داغ نروند و با این الگوریتم سیستم می‌آموزد که خطای خود را اصلاح کند. یادگیری در این سیستم‌ها به صورت تطبیقی صورت می‌گیرد؛ یعنی با استفاده از مثال‌ها وزن سیناپس‌ها به گونه‌ای تغییر می‌کند که در صورت دادن ورودی‌های جدید، سیستم پاسخ درستی تولید کند. شبکه عصبی شامل شبکه‌ای از عناصر پردازش ساده (نورون‌ها) است که می‌تواند رفتار پیچیده کلی تعیین‌شده‌ای از ارتباط بین عناصر پردازش و پارامترهای عنصر را نمایش دهد. منبع اصلی و الهام‌بخش برای این تکنیک، از آزمایش سیستم مرکزی عصبی و نورون‌ها (آکسون‌ها، شاخه‌های متعدد سلول‌های عصبی و محل‌های تماس دو عصب) نشأت گرفته است که یکی از قابل توجه‌ترین عناصر پردازش اطلاعات سیستم عصبی را تشکیل می‌دهد. در یک مدل شبکه عصبی، گره‌های ساده (به طور گسترده نورون، نئورون‌ها، «PE»ها (عناصر پردازش) یا واحدها) برای تشکیل شبکه‌ای از گره‌ها به هم متصل شده‌اند و به همین دلیل به آن، اصطلاح «شبکه‌های عصبی» اطلاق می‌شود. در حالی که یک شبکه عصبی نباید به خودی خود سازگارپذیر باشد، استفاده عملی از آن به‌واسطه الگوریتم‌هایی امکان‌پذیر است که جهت تغییر وزن ارتباطات در شبکه (به منظور تولید سیگنال مورد نظر) طراحی شده باشد. روش کار این الگوریتم در رابطه زیر نمایش داده شده است [۳۰] تا [۳۲]

$$ELR = a\times\frac{Number\ of\ current\ cuckoos\ eggs}{Total\ number\ of\ eggs}\times (Var_{hi} - Var_{low}) \quad (۸)$$

۱) مقادیری تصادفی به وزن‌ها نسبت می‌دهیم.
۲) پرسپترون را به تک‌تک مثال‌های آموزشی اعمال می‌کنیم. اگر مثال غلط ارزیابی شود مقادیر وزن‌های پرسپترون را تصحیح می‌کنیم.
۳) آیا تمامی مثال‌های آموزشی درست ارزیابی می‌شوند:
بله ← پایان الگوریتم
خیر ← به مرحله ۲ برمی‌گردیم.

برای یادگیری وزن‌های یک شبکه چندلایه از روش انتشار استفاده می‌کنیم. در این روش با استفاده از شیب نزول سعی می‌شود تا مربع خطای بین خروجی‌های شبکه و تابع هدف مینیمم شود. تعریف خطا طبق فرمول زیر محاسبه می‌شود

$$E(\vec{W}) \equiv \frac{1}{2}\sum_{d\in D}\sum_{k\in outputs}(t_{kd} - o_{kd})^2 \quad (۹)$$

منظور از *outputs* خروجی‌های مجموعه واحدهای لایه خروجی t_{kd} و o_{kd} مقدار هدف و خروجی متناظر با kامین واحد خروجی و مثال آموزشی d است. در پایان این مرحله پس از اجرای الگوریتم و کشف خطاهای احتمالی، این الگوریتم به پیش‌بینی خواهد پرداخت.

جدول ۱: مقایسه میزان فراخوانی و دقت روش پیشنهادی و روش‌های TAOBAO و MOBI-SAGE.

	Taobao	Mobi-sage	روش پیشنهادی
میزان فراخوانی	۸۹٪	۹۳٪	۹۴٪
میزان دقت	۹۰٪	۹۲٪	۹۳٪

جدول ۲: مقایسه میزان فراخوانی و دقت روش پیشنهادی با سایر الگوریتم‌های بهینه‌سازی فراابتکاری.

	RBF	KNN	PSO	روش پیشنهادی
میزان فراخوانی	٪۹۷	٪۹۰	٪۹۷	٪۹۸
میزان دقت	٪۸۹	٪۷۹	٪۹۸	٪۹۹

جدول ۳: مقایسه الگوریتم خوشه‌بندی SW-DBSCAN و سایر الگوریتم‌های خوشه‌بندی.

	C-Means	K-Means	DBSCAN	SW-DBSCAN
میزان کارایی	٪۶۹	٪۷۶	٪۹۹	٪۹۹

جدول ۴: مقایسه الگوریتم فاخته و سایر الگوریتم‌های فراابتکاری.

	الگوریتم بهینه‌سازی ذرات	الگوریتم گرگ خاکستری	الگوریتم کرم شب‌تاب	الگوریتم فاخته
میزان بهینه‌سازی	۹۵٪	۹۶٪	۹۷٪	۹۸٪

۵- پیشنهاد داده‌ها با استفاده از سیستم توصیه‌گر مبتنی بر فیلترینگ مشارکتی

در مرحله آخر با استفاده از یک سیستم توصیه‌گر مبتنی بر فیلترینگ مشارکتی، داده‌های حاصل از مراحل قبل را در اختیار کاربر قرار می‌دهیم. اغلب سیستم‌های توصیه‌گر کار خود را با یافتن کاربرانی که در اولویت‌ها و سوابق خرید اشتراکاتی با کاربر جاری دارند شروع می‌کنند. در ادامه با جمع‌آوری اطلاعات مربوط به آیتم‌های مورد علاقه این گروه و حذف آن دسته آیتم‌ها که قبلاً توسط کاربر خریداری شدند، لیستی از توصیه‌ها را ارائه می‌کند. به این گروه از سیستم‌ها، سیستم‌های مبتنی بر پالایش مشارکتی گویند و ازجمله پرکاربردترین سیستم‌ها در راستای تولید توصیه به کاربران هستند [۳۳] و [۳۴]. مکانیزم اصلی الگوریتم پالایش مشارکتی به این شرح است که با استفاده از معیار تشابه افراد، اولویت‌های گروه‌های زیادی از کاربران ثبت می‌شود. در ادامه کاربرانی که اولویت‌های یکسانی با کاربر جاری دارند، به‌عنوان همسایگان او انتخاب می‌شوند و سپس میانگین اولویت‌ها محاسبه شده و تابع اولویت نهایی سعی در توصیه آیتمی می‌کند که کاربر امتیازدهی نکرده است. نهایتاً تعدادی داده باقی خواهد ماند که این داده‌ها می‌توانند مورد علاقه کاربر باشند.

٦- ارزیابی روش پیشنهادی

اغلب برای اعتبارسنجی و ارزیابی سیستم های توصیه گر از معیارهایی مانند دقت و فراخوانی و کارایی استفاده می‌شود. در این تحقیق نیز برای ارزیابی قسمت‌های مختلف روش پیشنهادی از این معیارها استفاده شده است. دقت و فراخوانی در سیستم‌های توصیه‌گر با استفاده از دو رابطه زیر محاسبه می‌شود

$$Precision = \frac{\left|\{relevant\ item \cap retrived\ item\}\right|}{\left|\{retrived\ item\}\right|} \qquad (۱۰)$$

$$Recall = \frac{\left|\{relevant\ item \cap retrived\ item\}\right|}{\left|\{relevant\ item\}\right|} \qquad (۱۱)$$

نتایج حاصل از ارزیابی روش پیشنهادی بر اساس مقایسه دقت و فراخوانی روش پیشنهادی با روش‌های دیگر حاکی از برتری روش پیشنهادی نسبت به روش‌های مورد مقایسه است، به نحوی که روش پیشنهادی توانست به فراخوانی ٪۹۴ و دقت ۹۳٪ دست یابد. نتایج حاصل از دقت و فراخوانی روش پیشنهادی در جدول ۱ مشاهده می‌شود. در قسمت بهینه‌سازی نیز روش پیشنهادی توانست عملکرد ٪۹۸ و ٪۹۷ را به ترتیب در بخش های میزان فراخوانی و دقت به دست آورد. نتایج حاصل از میزان دقت و فراخوانی بهینه سازی با استفاده از الگوریتم های فراابتکاری را در جدول (۲) قابل مشاهده است. در قسمت میزان کارایی الگوریتم‌های خوشه‌بندی، الگوریتم پیشنهادی یعنی SW_DBSCAN توانست میزان کارایی ٪۹۹ را بدست آورد که حاکی از میزان عملکرد بالای این الگوریتم در مقایسه با سایر روش های مشابه است. نتایج حاصل از میزان کارایی خوشه‌بندی را در جدول ۳ مشاهده می‌شود. در نتایج حاصل از ارزیابی میزان بهینه‌سازی کلی روش پیشنهادی نیز توانست عملکرد ٪۹۸ را به‌دست آورد. نتایج حاصل از ارزیابی میزان بهینه‌سازی کلی روش پیشنهادی در جدول (۴) مشاهده می‌گردد.

۷- نتیجه‌گیری

سیستم‌های توصیه‌گر، سیستم‌هایی هستند که می‌توانند با گرفتن اطلاعات محدودی از کاربر و خصوصیاتی مانند موارد جستجوشده توسط کاربر در گذشته، لیستی از اقلام را در اختیار کاربر قرار دهند که می‌تواند مورد علاقه کاربر باشد. در این تحقیق به ارائه روشی به منظور بهبود سیستم‌های توصیه‌گر در زمینه اپ‌های موبایلی پرداخته شده است. روش کار بدین صورت می‌باشد که با استفاده از یک سیستم توصیه‌گر مبتنی بر فیلترینگ مشارکتی و با استفاده از روش‌های یادگیری ماشین سعی گردید سیستمی طراحی شود که بتواند مشکلات سیستم‌های قبلی را رفع و پیشنهادهای مناسبی به کاربر ارائه دهد. در این سیستم بعد از جمع‌آوری دیتاست مربوط به برنامه‌های مورد علاقه کاربر ابتدا عملیات پیش‌پردازش داده‌ها بر روی دیتاست مورد نظر انجام گرفت. سپس داده‌های خود را جهت ارزیابی میزان علاقه و شباهت آیتم‌ها به هم با استفاده از الگوریتم خوشه‌بندی SW-DBSCAN خوشه‌بندی کردیم که نتایج حاصل از ارزیابی میزان کارایی الگوریتم SW-DBSCAN حاکی از کارایی بالاتر این روش خوشه‌بندی نسبت به سایر روش‌های موجود بود که این الگوریتم توانست عملکرد ٪۹۹ را به دست آورد. این در حالی است که الگوریتم K-Means میزان کارایی ٪۷۶ و الگوریتم C-Means عملکرد ٪۶۹ را به دست آورد. سپس داده‌های به‌دست‌آمده را توسط الگوریتم فراابتکاری فاخته، بهینه کردیم که این الگوریتم در بخش بهینه‌سازی عملکرد ٪۹۸ را به دست آورد و این در حالی است که الگوریتم‌های کرم شب‌تاب، گرگ خاکستری و الگوریتم بهینه‌سازی ذرات به ترتیب ٪۹۷، ٪۹۶ و ٪۹۵ را به دست آوردند که نشان از عملکرد بهتر الگوریتم فاخته داشته است. در نهایت با استفاده از الگوریتم شبکه‌های عصبی جهت تولید پیش‌بینی پرداختیم که در این قسمت نیز این الگوریتم عملکرد قابل قبولی را به دست آورد و در مرحله آخر با استفاده از یک سیستم توصیه‌گر مبتنی بر فیلترینگ مشارکتی موارد مورد علاقه کاربر را استخراج کردیم و در اختیار کاربر هدف قرار دادیم. نتایج حاصل از ارزیابی روش پیشنهادی حاکی از میزان دقت و فراخوانی روش پیشنهادی نسبت به سایر

روش‌های موجود بود و با توجه به آمار و ارقام به‌دست‌آمده می‌توان گفت که روش پیشنهادی تا ۹۴٪ می‌تواند اطلاعات مورد علاقه کاربر را به‌درستی به وی پیشنهاد دهد.

در خصوص محورهای پیشنهادی در مورد پژوهش‌های آتی می‌توان موارد زیر را توصیه کرد:

- ترکیب چندین الگوریتم خوشه‌بندی به‌منظور خوشه‌بندی اطلاعات مربوط به کاربران؛
- بهره‌گیری از سایر الگوریتم‌های یادگیری ماشین پیش‌بینی؛
- ترکیب سایر روش‌های سیستم‌های توصیه‌گر با سایر روش‌ها.

مراجع

[1] S. Kim, *et al.*, "PubChem in 2021: new data content and improved web interfaces," *Nucleic Acids Research*, vol. 49, no. D1, pp. D1388-D1395, 2021.

[2] Adiyanto and Nurasiah, "Customer relationship management (CRM) based on web to improve the performance of the company," *IAIC Trans. on Sustainable Digital Innovation*, vol. 1, no. 1, pp. 32-41, Oct. 2019.

[3] A. Lau, E. Tsui, and W. Lee, "An ontology-based similarity measurement for problem-based case reasoning," *Expert Systems with Applications*, vol. 36, no. 3, pt. 2, pp. 6574-6579, Apr. 2009.

[4] F. Beierle and T. Eichinger, "Collaborating with users in proximity for decentralized mobile recommender systems," in *Proc. IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, pp. 1192-1197, Leicester, UK, 19-23 Aug. 2019.

[5] M. Montaner, B. López, J. L. De La Rosa, "A taxonomy of recommender agents on the Internet," *Artificial Intelligence Review*, vol. 19, pp. 285-330, 2003.

[6] B. Abu-Salih, *et al.*, "Toward a knowledge-based personalised recommender system for mobile app development," *JUCS-J. of Universal Computer Science*, vol. 27, no. 2, pp. 208-229, 2021.

[7] C. Yang, *et al.*, "Enhancing mobile app recommendations through adaptive fusion of long-term stability and short-term interests," *Information Sciences*, vol. 699, Article ID: 121817, May 2025.

[8] D. Yu, X. Zhou, A. Noorian, and M. Hazratifard, "An AI-driven social media recommender system leveraging smartphone and IoT data," *The J. of Supercomputing*, vol. 81, no. 1, pp. 1-32, Jan. 2025.

[9] F. Nadeem and M. Suleman, *Developing a Personalized Recommender System for Enhanced User Experience in Mobile Applications*, Oct. 2024. DOI:10.13140/RG.2.2.19331.31528

[10] Y. Gong, *et al.*, "EdgeRec: recommender system on edge in mobile Taobao," in *Proc. of the 29th ACM Int. Conf. on Information & Knowledge Management*, pp. 2477-2484, Virtual Event Ireland, 19-23 Oct. 2020.

[11] E. Pimenidis, N. Polatidis, and H. Mouratidis, "Mobile recommender systems: identifying the major concepts," *J. of Information Science*, vol. 45, no. 3, pp. 387-397, Jun. 2019.

[12] H. Yin, *et al.*, "Mobi-SAGE-RS: a sparse additive generative model-based mobile application recommender system," *Knowledge-Based Systems*, vol. 157, pp. 68-80, Oct. 2018.

[13] H. Ranaivoson, A. Smets, and P. Ballon, "15 Challenges and opportunities for recommender systems in media markets," In U. Rohn, M. B. von Rimscha, and T, Raats (Eds.) *De Gruyter Handbook of Media Economics*, pp. 215-228, Boston: De Gruyter, 2024.

[14] E. Hasan, M. Rahman, C. Ding, J. X. Huang, and S. Raza, *Review-Based Recommender Systems: A Survey of Approaches, Challenges and Future Perspectives*, arXiv preprint arXiv:2405.05562, 2024.

[15] X. Chen, S. Wang, J. McAuley, D. Jannach, and L. Yao, "On the opportunities and challenges of offline reinforcement learning for recommender systems," *ACM Trans. on Information Systems*, vol. 42, no. 6, Article ID: 150, 26 pp., Nov. 2023.

[16] N. Alharbe, M. A. Rakrouki, and A. Aljohani, "A collaborative filtering recommendation algorithm based on embedding representation," *Expert Systems with Applications*, vol. 215, Article ID: 119380, Apr. 2023.

[17] S. A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, "Data preprocessing in predictive data mining," *The Knowledge Engineering Review*, vol. 34, 2019. doi:10.1017/S026988891800036X

[18] N. Ohadi, *et al.*, "SW-DBSCAN: a grid-based DBSCAN algorithm for large datasets," in *Proc. 6th Int. Conf. on Web Research,* pp. 139-145, Tehran, Iran, 22-23 Apr. 2020.

[19] Y. Jiang, C. Kang, Y. Shen, T. Huang, and G. Zhai, "Research on Argo data anomaly detection based on improved DBSCAN algorithm," in *Proc. China Conference on Wireless Sensor Networks*, pp. 44-54, Guangzhou, China, 10-12 Nov. 2022.

[20] R. Molaee Fard and P. Yarahmadi, "A recommendation system in the medical industry using SW-DBSCAN algorithm," *International J. of Web Research*, vol. 5, no. 1, pp. 82-87, Winter-Spring 2022.

[21] M. A. Sorkhi, E. Akbari, M. Rabbani, and H. Motameni, "A dynamic density-based clustering method based on K-nearest neighbor," *Knowledge and Information Systems*, vol. 66, pp. 3005-3031, 2024.

[22] I. De Moura Ventorim, D. Luchi, A. L. Rodrigues, and F. M. Varejão, "BIRCHSCAN: a sampling method for applying DBSCAN to large datasets," *Expert Systems with Applications*, vol. 184, Article ID: 115518, Dec. 2021.

[23] O. Bourkoukou and O. Achbarou, "Weighting based approach for learning resources recommendations," *JOIV: International J. on Informatics Visualization*, vol. 2, no. 3, pp. 104-109, 2018.

[24] B. Zhao, Y. Zheng, X. Yu, W. Song, "A study on product recommendation based on dynamically adjusted multimodal weights," in *Proc. IEEE 5th Int. Conf. on Software Engineering and Artificial Intelligence*, pp. 235-239, Fuzhou, China, 20-22 Jun. 2025.

[25] H. R. Boveiri, "An enhanced cuckoo optimization algorithm for task graph scheduling in cluster-computing systems," *Soft Computing*, vol. 24, no. 13, pp. 10075-10093, 2020.

[26] X. Cai, *et al.* "An under-sampled software defect prediction method based on hybrid multi-objective cuckoo search," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 5, Article ID: e5478, Mar. 2020.

[27] F. Rezaimehr and C. Dadkhah, "T&TRS: robust collaborative filtering recommender systems against attacks," *Multimedia Tools and Applications*, vol. 83, pp. 31701-31731, 2024.

[28] C. Yuan, Y. Yang, and Y. Liu, "Sports decision-making model based on data mining and neural network," *Neural Computing and Applications*, vol. 33, pp. 3911-3924, 2021.

[29] N. Zhou, "Database design of regional music characteristic culture resources based on improved neural network in data mining," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 103-114, Feb. 2020.

[30] A. Sahu, G. M. Harshvardhan, and M. K. Gourisaria, "A dual approach for credit card fraud detection using neural network and data mining techniques," in *Proc. IEEE 17th India Council Int. Conf.*, 7 pp., New Delhi, India, 10-13 Dec. 2020.

[31] H. Zhou, G. Sun, S. Fu, J. Liu, X. Zhou, and J. Zhou, "A big data mining approach of PSO-based BP neural network for financial risk management with IoT," *IEEE Access*, vol. 7, pp. 154035-154043, 2019.

[32] S. Wang, J. Cao, and P. Yu, "Deep learning for spatio-temporal data mining: a survey," *IEEE Trans. on Knowledge and Data Engineering*, vol. 34, no. 8, pp. 3681-3700, Aug. 2020.

[33] R. Molaee Fard, "Provide a video recommendation system using collaborative filtering and data mining methods," *J. of Advances in Computer Research*, vol. 12, no. 1, pp. 13-25, 2021.

[34] S. M. Z. Kashani and J. Hamidzadeh, "Feature selection by using privacy-preserving of recommendation systems based on collaborative filtering and mutual trust in social networks," *Soft Computing*, vol. 24, no. 15, pp. 11425-11440, Aug. 2020.

رضا مولایی فرد، دکترای رشته کامپیوتر در گرایش نرم افزار از دانشگاه علوم و تحقیقات تهران است و در حال حاضر عضو هیأت علمی دانشگاه آزاد اسلامی واحد دزفول است. نام‌برده در حال حاضر در زمینه تحلیل داده‌ها فعالیت دارد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: داده کاوی، یادگیری ماشین، بهینه‌سازی و الگوریتم‌های فراابتکاری.

جواد محمدزاده دکترای رشته کامپیوتر از دانشگاه تهران است و در حال حاضر عضو هیأت علمی دانشگاه آزاد اسلامی واحد کرج می‌باشد. زمینه‌های تحقیقاتی موردعلاقه ایشان عبارتند از: شبکه‌های پیچیده پویا، سیستم‌های توزیع‌شده و الگوریتم‌های متاهیورستیک.

پیام یاراحمدی دکترای رشته کامپیوتر از دانشگاه ملی تاراس شوچنکو کیف اوکراین است نام‌برده در حال حاضر در زمینه تحلیل داده‌ها فعالیت دارد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: یادگیری ماشین، سیستم‌های توزیع‌شده، رایانش ابری و اینترنت اشیاء.