مدیریت فناوری اطلاعات
دانشکدهٔ مدیریت دانشگاه تهران

دورهٔ ۷، شمارهٔ ۲
تابستان ۱۳۹۴
ص. ۴۰۶- ۳۸۵

فروشگاه اینترنتی هوشمند:
سیستم پیشنهاددهندهٔ مبتنی بر تحلیل رفتار کاربران

محمدرضا کریمی علویجه[۱]، شیوا عسکری[۲]، سیروان پرسته[۳]

چکیده: سیستم‌های پیشنهاددهنده‌ای که با ارائهٔ پیشنهادها شخصی‌سازی می‌شوند و به کاربران در یافتن محصولی که علاقه دارند، کمک می‌کنند، می‌توانند در ترغیب مشتریان به خرید از وب‌سایت و در نتیجه موفقیت فروشگاه‌های آنلاین، نقش کلیدی ایفا کنند. روش پالایش همکارانه، یکی از موفق‌ترین روش‌های به‌کاررفته در این سیستم‌ها است که توانایی ارائهٔ پیشنهادهایی نزدیک به نظر کاربران را دارد، اما با افزایش تعداد کاربران و محصولات، با مشکلاتی مانند شروع سرد و مقیاس‌پذیری مواجه می‌شوند. به همین دلیل در این پژوهش روش جدیدی معرفی شده است که ضمن به‌کارگیری الگوریتم روش پالایش همکارانهٔ مبتنی بر کاربر به‌مثابهٔ رویکرد پایه، از ترکیب وزن‌دار خوشه‌بندی کاربران بر اساس اطلاعات جمعیت‌شناختی آنها نیز برای دستیابی به نتایج بهتر از سیستم، استفاده کرده است. نتایج پیاده‌سازی الگوریتم نشان داد رویکرد ارائه‌شده، ریشهٔ میانگین مربعات خطای کمتری دارد که به‌معنای عملکرد بهتر و دقت بیشتر آن است و پیش‌بینی‌های حاصل از آن با ترجیح و سلیقهٔ کاربران همخوانی بیشتری دارد.

واژه‌های کلیدی: ***اطلاعات جمعیت‌شناختی، خوشه‌بندی، رویکرد پالایش همکارانه، سیستم پیشنهاددهنده.***

۱. استاد گروه مدیریت بازرگانی دانشکدهٔ مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران
۲. کارشناس‌ارشد مدیریت بازرگانی ـ گرایش بازاریابی، دانشگاه علامه طباطبائی، تهران، ایران
۳. کارشناس‌ارشد مهندسی کامپیوتر ـ گرایش هوش مصنوعی، دانشگاه علم و صنعت، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۳/۱۲/۱۱
تاریخ پذیرش نهایی مقاله: ۱۳۹٤/۰۳/۲۰
نویسندهٔ مسئول مقاله: شیوا عسکری
E-mail: askari921@atu.ac.ir

مقدمه

فروش گستردهٔ محصولات با ورود فروشگاه‌ها و خرده‌فروشان آنلاین، بسیار آسان‌تر شده است؛ به‌طوری که نسبت به فروشگاه‌های فیزیکی، درصد فروش بیشتری را به خود اختصاص داده‌اند. اما مشتریان برای یافتن محصولات مد نظرشان از بین بسیاری از محصولات و رده‌های این فروشگاه‌ها، باید به جست‌وجو بپردازند (اندرسون و هیلارال، ۲۰۰۹). به بیانی، مشتریان زمان زیادی را برای مشاهدهٔ وب‌سایت فروشگاه‌ها و جمع‌آوری اطلاعات ارزشمند صرف می‌کنند؛ در حالی‌که محتوای بسیاری از وب‌سایت‌ها با نیاز مشتریان همخوانی ندارد که این مسئله بر اطمینان و علاقهٔ مشتریان برای خرید اینترنتی تأثیر می‌گذارد (زنگ، ۲۰۰۹). برای مشتریان انتخاب فروشگاهی که مناسب‌ترین محصول را در اختیار آنها قرار دهد، کار دشواری است. یکی از راهبردهای مؤثر در بازاریابی، این است که با ارائهٔ پیشنهادهای خاص برای هر کاربر، آنها را به سمت خود جلب کنند؛ به‌گونه‌ای که با رضایت از بین انبوه فروشگاه‌های آنلاین به خرید از فروشگاه مد نظر ترغیب شوند. از سویی این پیشنهادها به شناسایی و تحلیل ویژگی‌های کاربران نیاز دارد (ایواتا، ۲۰۰۸). هنگامی که کاربران درگیر یافتن و انتخاب محصولات، خدمات و اطلاعات‌اند، موارد بسیاری باید در نظر گرفته شود و به سیستمی نیاز است که کاربر را حمایت کند (واندن بالک، لکرون، دوکاروز و فوز، ۲۰۱۳). یکی از راه‌های غلبه بر مشکلات ذکرشده و کمک به مشتریان در تصمیم‌گیری، توسعهٔ سیستم‌های پیشنهاددهنده به‌منظور ارائهٔ خدمات شخصی‌سازی‌شده است. این سیستم‌ها در فروشگاه‌های اینترنتی پیاده‌سازی می‌شوند و پیشنهادهای شخصی‌سازی‌شده‌ای را با توجه به ترجیح و علاقهٔ کاربر، ارائه می‌کنند (گوپتا و گاژ، ۲۰۱۴). سیستم پیشنهاددهنده، سازوکار قدرتمندی برای حل مشکل سرریز اطلاعات در فروشگاه‌های اینترنتی است. شکل ۱ معماری سایت فروش مجهز به سیستم پیشنهاددهنده را نشان می‌دهد.

شکل ۱. معماری وب‌سایت مجهز به سیستم پیشنهاددهنده

بر این اساس، سیستم‌های پیشنهادی با روند رو به افزایشی در وب‌سایت‌های تجاری به‌کار گرفته می‌شوند. مشهورترین سیستم پیشنهادی را وب‌سایت آمازون پیاده‌سازی کرده است که از رفتار خرید مشتریان و نظر و علایق آنها استفاده می‌کند و محصولاتی را متناسب با سلیقهٔ کاربران مانند کتاب، پیشنهاد می‌دهد (لیندن و اسمیت و یورک، ۲۰۰۳). سیستم پیشنهاددهندهٔ خوب، محصولاتی را به مشتریان معرفی می‌کند که به احتمال زیاد مشتریان به آن علاقه‌مندند، اما برای خرید آن برنامه‌ریزی ندارد و نمی‌تواند مشتریان را به خرید بیشتر تشویق کند. این خریدهای بی‌برنامه در فروشگاه‌های آنلاین، به اندازهٔ فروشگاه‌های سنتی و فیزیکی رخ نمی‌دهد (پک و چیلدرز، ۲۰۰۶). سیستم‌های پیشنهاددهنده، ارتباطی را با ارزش افزوده بین کاربران و وب‌سایت پایه‌ریزی می‌کنند. هرچه کاربر از وب‌سایت بیشتر خرید کند، به سیستم برای ارائهٔ پیشنهادهای بهتر کمک می‌کند و این کار فروش را افزایش می‌دهد (اندرسون و هیلارال، ۲۰۰۹). این سیستم‌ها دانش مشتریان و وفاداری آنها را افزایش می‌دهد که این راهبردی ضروری برای کسب‌وکار در محیط تجارت الکترونیک به‌شمار می‌رود (یون، هوستلر، گو و گویماراس، ۲۰۱۳). سیستم پیشنهاددهنده بر آن است تا محصولات فروشگاه‌های آنلاین را با توجه به سلیقهٔ مشتری غربال کند. سیستم بر پایهٔ پروفایل فعلی مشتری ساخته می‌شود و به همین دلیل، امکان پیداکردن محصول مد نظر کاربر را افزایش می‌دهد. از این رو، برای درک نیاز مشتریان و پاسخ مناسب به این نیازها، ضروری است از الگوریتم‌های کارآمدی استفاده شود. از این رو پژوهش حاضر قصد دارد با مطالعه و تحقیق در این حوزه و ارائهٔ الگوریتمی برای بهبود سیستم پیشنهادهندهٔ مبتنی بر راهکار پالایش همکارانهٔ مبتنی بر کاربر، به این مسئله پاسخ دهد و از این طریق پیشنهادهایی ارائه دهد که با سلیقه و ترجیح مشتریان همخوانی بیشتری داشته باشد.

استفاده از روش پالایش همکارانهٔ مبتنی بر حافظه، یکی از روش‌های موفق و شناخته‌شده در توسعهٔ سیستم‌های پیشنهاد دهنده است. در این رویکرد برای پیشنهاد کالا به مشتری، از نظر کاربران دیگری که آن کالا را تهیه کرده‌اند، استفاده می‌شود. یکی از چالش‌های پیش روی این روش، ورود کاربران جدید و نبود پیشینهٔ اطلاعاتی از آنها و یافتن کاربرانی است که سلیقه و ترجیح مشابهی با کاربران جدید دارند؛ در این صورت، سیستم نمی‌تواند پیشنهادی در اختیار آنها قرار دهد. از این رو پژوهش حاضر قصد دارد از اطلاعات فردی کاربران، به‌منظور بهبود نتایج سیستم پیشنهاددهنده استفاده کند. بدین منظور، کاربران بر اساس اطلاعات فردی خوشه‌بندی می‌شوند و معیار جدیدی برای محاسبهٔ شباهت با ترکیب وزن‌دار شباهت خوشه‌ای و شباهت پالایش همکارانه ایجاد می‌شود. برای پیشگیری از مشکل یادشده، می‌توان شباهت خوشه‌ای

کاربران جدید را با وزن بزرگ‌تری همراه کرد و برعکس آن، برای کاربران قدیمی‌تر از وزن شباهت خوشه‌ای کاست. انتظار می‌رود، سیستم جدید در ارائهٔ پیشنهادهایی هم‌راستا با سلیقهٔ کاربران، بهتر عمل کند.

پیشینهٔ پژوهش

سیستم‌های پیشنهاددهنده می‌توانند با استفاده از انواع اطلاعاتی که از رفتار کاربران در اختیار دارند، الگوهایی را استخراج کنند و محصولاتی را به کاربران پیشنهاد دهند (گوپتا و گاژ، ۲۰۱۴). این اطلاعات عبارت‌اند از: دفعات تکرار خرید مشتری و وفاداریش، زمان شروع یا پایان مشاهدهٔ وب‌سایت فروشگاه، نوع کالاهایی که قبلاً کاربر پسندیده، زمان خرید کالا توسط کاربر و همچنین اطلاعاتی مانند رتبه‌بندی، پیشینهٔ خرید و اطلاعات کالاها (ایواتا، ۲۰۰۸).

مطالعات ایواتا (۲۰۰۸) نشان داد فروشگاه با درک علاقه و سلیقهٔ اخیر کاربران، می‌تواند پیشنهادها مؤثرتری ارائه دهد؛ زیرا علایق و سلایق کاربران روزبه‌روز تغییر می‌کند. با پیشنهاد محصولاتی که اغلب توسط کاربران وفادار خریداری شده است، فروشگاه می‌تواند فروش را افزایش دهد؛ چرا که به شمار کاربران وفادار افزوده می‌شود. همچنین فروشندگان می‌توانند در تعیین قیمت محصولات انعطاف‌پذیری بیشتری داشته باشند (گوپال و همکاران، ۲۰۰۶).

فروشگاه‌های آنلاین فراوانی از سیستم‌های پیشنهاددهنده برای ارائهٔ خدمات و محصولاتی مانند کتاب، فیلم، موزیک، مقاله، کارتون، جوک و صفحات وب استفاده می‌کنند. در وب‌سایت‌های فارسی با استفاده از شباهت ساده و معنایی میان جمله و کلمه (عباسی، منصفی و استیری، ۱۳۹۱) و به‌کارگیری ریشهٔ یک کلمهٔ کلیدی به جای خود آن، توانسته‌اند مقاله‌های مشابه علاقهٔ پیشین کاربر را با دقت بالایی پیشنهاد دهند (استیری، کاهانی و پورغلامی، ۱۳۹۱).

برای پیاده‌سازی سیستم پیشنهاددهنده، سه رویکرد اصلی را می‌توان به شرح زیر نام برد (آداماویشز و توژیلین، ۲۰۰۵):

۱. پالایش برپایهٔ محتوا[1]؛
۲. پالایش همکارانه[2]؛
۳. روش ترکیبی[3].

در پالایش بر اساس محتوا، کالاهایی که مشخصات مشابهی دارند، به‌گونه‌ای یکسان درجه‌بندی می‌شوند (اندرسون و هیلارال، ۲۰۰۹). این رویکرد به در دسترس‌بودن مشخصات

1. Content based
2. Collaborative filtering
3. Hybrid method

کالاها و پروفایل کاربران مرتبط با این مشخصات، بستگی دارد. در این رویکرد مشخصات پروفایل کاربر با تحلیل محتوا از کالاهایی که قبلاً مشاهده‌کرده یا به آنها رتبه‌ای داده، به‌دست می‌آید و برای هر یک از آنها واژهٔ کلیدی در نظر می‌گیرد و محصولاتی را با بیشترین شباهت به کالاهای موجود در پروفایل کاربر، به او پیشنهاد می‌دهد. در واقع مواردی مشابه به آنچه کاربر پیش‌تر ترجیح داده است به او پیشنهاد می‌شود (لیمستول و استورمر، ۲۰۰۷).

روش پالایش همکارانه از دو بخش تشکیل می‌شود: ۱. پالایش اطلاعات و انتخاب اطلاعات درست از مجموعهٔ داده و ۲. بخش مشترک که اطمینان حاصل می‌کند، اطلاعات استخراج‌شده از بخش قبل بین همهٔ کاربران مشترک است. این روش بر این فرض استوار است که کاربرانی که سلیقه‌های مشابهی دارند، رتبهٔ شبیه به هم می‌دهند (لیمستول و استورمر ، ۲۰۰۷). برای مثال در این روش، اگر کاربر اول به محصولات «الف» و «ب» علاقه داشته باشد و کاربر دوم که شبیه کاربر اول است، کالاهای «الف»، «ب» و «ج» را پسند کند؛ به‌دلیل این شباهت، به احتمال زیاد کاربر اول کالای «ج» را هم می‌پسندد. از این رو کالای «ج» به او پیشنهاد می‌شود.

در روش ترکیبی سعی بر آن است ترکیبی از روش‌های پیشین به‌کار گرفته شود و از این طریق بر محدودیت‌های این روش‌ها غلبه شود و ضمن بهره‌بردن از مزیت آنها، سیستم را بهینه‌سازی کند (اندرسون و هیلارال، ۲۰۰۹).

روش پالایش همکارانه، روش متداولی برای سیستم‌های پیشنهادی است و در آن، پیشنهادها بر پایهٔ رتبه‌بندی سایر کاربران در سیستم ساخته می‌شود (گوپتا و گاژ، ۲۰۱۴). در این روش بر اساس کاربران مشابه و رتبه‌هایی که به کالاها داده شده است، ماتریس‌هایی ایجاد می‌شود که سطرهای آن را m کاربر و ستون‌های آن را n کالا شکل می‌دهد (شبیب، نعمت‌بخش و عقیلی دهکردی، ۱۳۸۶).

جدول ۱. نمونهٔ ماتریس حاصل از محصولات و کاربران

	محصول ۱(i۱)	محصول ۲ (i۲)	محصول ۳ (i۳)	محصول ٤ (i٤)
کاربر ۱ (u۱)	۳	۵	۲	۱
کاربر ۲ (u۲)	۳	۴	۲	۲
کاربر ۳ (u۳)	۳	۵	۲	۴
کاربر ۴ (u۴)	۳	۵	۴	۵

الگوریتم‌های این روش به دو دسته تقسیم می‌شود: الگوریتم مبتنی بر مدل[1] و الگوریتم مبتنی بر حافظه[2].

در روش مبتنی بر مدل برای محاسبه و استخراج پیشنهادها، داده‌های پروفایل کاربران به‌صورت آفلاین بازخوانی و پردازش می‌شود که این امر به بهبود محاسبات منجر می‌شود (لیمستول و استورمر، ۲۰۰۷). در این روش، مدل بر اساس رتبه‌بندی مجموعهٔ داده، ساخته می‌شود. از روش مبتنی بر مدل هنگامی استفاده می‌شود که حجم مجموعهٔ داده بسیار زیاد است. سیستم پیشنهادی ابتدا با استفاده از روش‌های آماری، مدل پیش‌بینی‌کننده را به‌کمک بخشی از مجموعهٔ داده ـ که مربوط به فعالیت‌های قبلی کاربران است ـ می‌سازد و از آن برای ایجاد پیشنهادها استفاده می‌کند. الگوریتم‌هایی که برای روش مبتنی بر مدل استفاده می‌شوند عبارت‌اند از: روش احتمالی بیزین، مدل‌های خوشه‌بندی و شبکه‌های وابستگی و شبکه‌های عصبی (سو و خوش‌گفتار، ۲۰۰۹).

روش مبتنی بر حافظه، روش اکتشافی است، هیچ نتیجهٔ از قبل محاسبه‌شده‌ای ندارد و برای پیش‌بینی، از ماتریس رتبه‌بندی کاربران و کالاها استفاده می‌کند. هر کاربر متعلق به گروهی از کاربران با علایق مشابه است. پیش‌بینی‌ها با شناسایی کاربران فعال مشابه که همسایه نامیده می‌شوند، صورت می‌گیرد (اندرسون و هیلارال، ۲۰۰۹). روش مبتنی بر حافظه به دو شکل انجام می‌شود:

۱. مبتنی بر محصول (کالا)[3]: شباهت بین کالاهایی که کاربران به آن رأی داده‌اند، بررسی می‌شود. برای مثال، اگر کاربر محصول «الف» را خریده است، به احتمال زیاد محصول «ب» را نیز که مشابه محصول «الف» است، خریداری می‌کند.
۲. مبتنی بر کاربر[4]: پس از شناسایی شباهت‌های کاربران، گروه‌هایی با بیشترین شباهت ایجاد می‌شود. با محاسبهٔ میانگین وزنی امتیازهای داده‌شده، ترجیح کاربر مشابه پیش‌بینی می‌شود (زنگ، ۲۰۰۹). در این روش با جمع امتیاز کاربران مشابه (ru, i)، امتیاز کاربر u به کالای i پیش‌بینی می‌شود.

رابطهٔ ۱) $$r_{u,i} = aggr\, r_{u',i} \quad , \quad u' \in U$$

1. Model based
2. Memory based
3. Item based
4. User based

U نشان‌دهندهٔ مجموع کاربران مشابه کاربر u است که به کالای i رتبه داده‌اند. در برخی از مطالعات تجمع به‌صورت زیر محاسبه می‌شود.

$$r_{u,i} = \frac{1}{n} \sum_{u' \in U} r_{u',i}$$ رابطهٔ ۲)

$$r_{u,i} = k \sum_{u' \in U} sim(u,u') \times r_{u',i}$$ رابطهٔ ۳)

که مضرب k به‌منزلهٔ فاکتور نرمال‌سازی به‌کار می‌رود و از رابطهٔ ۴ به‌دست می‌آید.

$$k = \frac{1}{\sum_{u' \in U} sim(u,u')}$$ رابطهٔ ۴)

در بیشتر مراجع از رابطهٔ ۳ برای تجمع استفاده می‌شود. هرچه کاربر u و u' به یکدیگر شبیه باشند، با نرخ وزنی بیشتری $r_{u',i}$ برای پیش‌بینی $r_{u,i}$ به‌کار می‌رود (شبیب و همکاران، ۱۳۸۶).

برای پیاده‌سازی روش مبتنی بر حافظه باید گام‌هایی طی شود. ابتدا برای هریک از کاربران ضریب وزنی با توجه به میزان شباهت به کاربر فعال در نظر گرفته شود. در گام دوم، شباهت بین کاربر مد نظر و سایر کاربران فعال محاسبه شود. روش‌های بسیاری مانند روش فاصلهٔ کسینوسی، تفاوت میانگین مربعات و ضریب همبستگی پیرسون، برای محاسبهٔ این شباهت استفاده می‌شود (دای و ییو گانگ، ۲۰۰۹). در روش کسینوس، زاویهٔ بین بردار رتبه‌بندی‌های دو کاربر محاسبه می‌شود، زاویهٔ کوچک‌تر نشان‌دهندهٔ شباهت بیشتر است.

$$sim(u,u') = \frac{\sum_{i=1}^{n} r_{ui} r_{u'i}}{\sqrt{\sum_{i=1}^{n} {r_{ui}}^2 \sum_{i=1}^{n} {r_{u'i}}^2}}$$ رابطهٔ ۵)

که در آن؛ r_{ui} رتبه‌بندی هریک از موارد i توسط کاربر u و n تعداد موارد رتبه‌بندی‌شده توسط هر دو کاربر است.

به‌کمک روش همبستگی پیرسون رابطهٔ دو کاربر به‌صورت خطی مشخص می‌شود:

$$sim(u,u') = \frac{\sum_{i=1}^{n} (r_{ui} - \bar{r}_u)(r_{u'i} - \bar{r}_{u'})}{\sqrt{\sum_{i=1}^{n} (r_{ui} - \bar{r}_u)^2 \sum_{i=1}^{n} (r_{u'i} - \bar{r}_{u'})^2}}$$ رابطهٔ ۶)

i مجموعه‌ای از کالاها است که توسط دو کاربر رتبه‌بندی شده است، $r_{u'i}$ رتبه‌بندی کاربر u برای کالای i است و $\overline{r}_u$ میانگین رتبه‌بندی کاربر است.

در سیستم‌های پیشنهاد دهنده از روشی برای محاسبهٔ همسانی کاربران استفاده می‌شود که به کارایی سیستم کمک می‌کند. اغلب از رابطهٔ ۵ برای محاسبهٔ شباهت بین کاربران استفاده می‌شود.

معمولاً سیستم‌های پیشنهاددهندهٔ مبتنی بر روش‌های متداول، با مشکلات اساسی مواجه می‌شوند که کیفیت پیش‌بینی‌های تولیدشدهٔ مبتنی بر روش پالایش همکارانه را کاهش می‌دهد (شافر، کنستان وریدل، ۲۰۰۱). از جملهٔ این مشکلات، می‌توان به ماتریس‌های خلوت در داده‌ها اشاره کرد. در واقع با وجود کاربران بسیار زیاد، معمولاً رتبه‌بندی کامل یا کافی از محصولات در دست نیست و این مشکل سبب می‌شود در ماتریس‌های حاصل از کاربر و محصولات، خانه‌های بسیار زیادی خالی بماند که در پردازش اطلاعات، محاسبات و تعیین موقعیت همسایه‌ها، اختلال ایجاد می‌کند (اندرسون و هیرالال، ۲۰۰۹). به‌دنبال این مشکل، مسائلی دیگری چون مشکلات مربوط به محصولات یا کاربران جدید پدیدار می‌شود که به مسئلهٔ شروع سرد[1] معروف است. در این وضعیت، برای محصولات جدید یا کاربران تازه‌وارد، رتبه‌بندی وجود ندارد و این مانع یافتن گروه‌های مشابه آنها می‌شود و نمی‌توان متناسب با ترجیح کاربران جدید، محصولاتی را پیشنهاد داد (سو و خوش‌گفتار، ۲۰۰۹).

از مشکلات دیگر، می‌توان به مقیاس‌پذیری اشاره کرد که به‌دنبال افزایش حجم داده ایجاد می‌شود و به کاهش کارایی سیستم می‌انجامد (دای، یی و گانگ، ۲۰۰۹). در واقع زمانی که تعداد کاربران و کالاها به میزان زیادی افزایش می‌یابد، الگوریتم‌های متداول پالایش همکارانه عملکرد خوبی از خود نشان نمی‌دهند (سو و خوش گفتار، ۲۰۰۹).

هر سیستم پیشنهاددهنده باید بتواند به‌سرعت در مقابل احتیاجات آنلاین واکنش نشان دهد و پیشنهادهای را برای همهٔ کاربران، بدون توجه به پیشینهٔ خرید و رتبه‌بندی آنها، ارائه کند؛ این مسئله به‌شدت از مقیاس‌پذیری سیستم مبتنی بر پالایش همکارانه تأثیر می‌پذیرد (لیندن و اسمیت و یورک، ۲۰۰۳). برای غلبه بر مشکلات ذکرشده، در مطالعات گذشته رویکردهای متفاوتی برای بهبود روش پالایش همکارانه به‌کارگرفته شده است که از آن جمله می‌توان به در نظر گرفتن تغییر علایق کاربران در بازه‌های زمانی مختلف (زنگ، ۲۰۰۹)، خوشه‌بندی محصولات و ایجاد درخت تصمیم برای ارتباط‌دادن محصولات جدید با محصولات موجود (چو، کیم وکیم،

1. Cold start problem

۲۰۰۲) یا ترکیب روش‌های پالایش برپایهٔ محتوا و پالایش همکارانه (چوی، کیم، سو ویو، ۲۰۱۲) در پاسخگویی به مسئلهٔ شروع سرد محصولات، اشاره کرد.

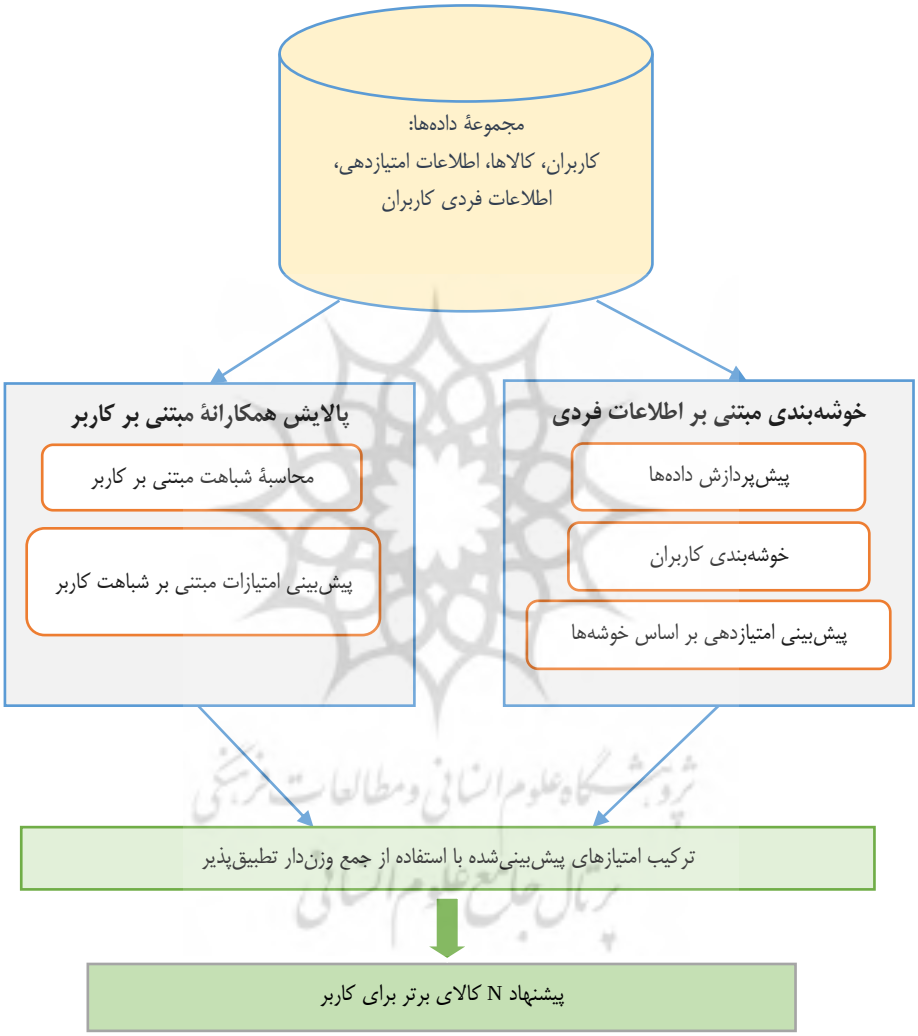
در پژوهش حاضر تلاش شده است با استفاده از خوشه‌بندی کاربران بر اساس اطلاعات جمعیت‌شناختی، گامی در جهت بهبود روش پالایش همکارانه برداشته شود. به بیانی، در این پژوهش سیستم پیشنهاددهندهٔ ترکیبی مبتنی بر پالایش همکارانهٔ کاربر محور ارائه می‌شود که از اطلاعات فردی کاربران به‌منظور خوشه‌بندی آنها استفاده می‌کند و از خوشه‌های ایجاد شده برای بهبود پیشنهادها بهره می‌برد.

با توجه به اینکه روش‌های متداول پالایش همکارانه، فقط رتبه‌دهی همسایگان را در نظر می‌گیرند و به اطلاعات جمعیت‌شناختی کاربران توجهی ندارند (دای و همکاران، ۲۰۰۹) و همچنین رویکردهای مبتنی بر مدل مقیاس‌پذیری، مقیاس‌پذیری بیشتر و دقت کمتری دارند، اما رویکردهای مبتنی بر حافظه، مقیاس‌پذیری کمتر و دقت بیشتری دارند (لیانگ، بو و جان، ۲۰۰۹)؛ در این پژوهش، روش خوشه‌بندی بر اساس اطلاعات فردی کاربران که یکی از روش‌های مبتنی بر مدل است، به‌کار گرفته می‌شود. با انتخاب این رویکرد، روش پالایش همکارانهٔ مبتنی بر حافظه بهبود می‌یابد و رویکردی ترکیبی مبتنی بر کاربر اتخاذ می‌شود. بر این اساس اگر دو کاربر در ویژگی‌هایی مثل سن یا تخصص مشابهت داشته باشند، نسبت به بعضی از محصولات سلیقهٔ مشابهی دارند که این مسئله در رتبه‌بندی‌های کاربران منعکس نمی‌شود (دای و همکاران، ۲۰۰۹). مدل‌های خوشه‌بندی، عملکرد و مقیاس‌پذیری آنلاین بهتری نسبت به پالایش همکارانه دارند؛ زیرا کاربر به جای مقایسه با همهٔ مشتریان، تنها با تعدادی محدود از کاربران هم‌خوشه مقایسه می‌شود. در خوشه‌بندی، مشتریان بر اساس معیار شباهت به بخش‌هایی (گروه‌هایی) تقسیم می‌شوند. هدف این الگوریتم هدایت کاربران به بخش‌هایی با بیشترین شباهت است. بنابراین، از این مرحله از خریدها و رتبه‌بندی‌های مشتریان در هر گروه، برای ایجاد پیشنهادها استفاده می‌شود (لیندن و همکاران، ۲۰۰۳).

کاربران بر اساس یکی از معیارهای شباهت خوشه‌بندی می‌شوند. میزان شباهت دو کاربر به یکدیگر، بر اساس مقادیر برخی ویژگی‌های توصیف‌کننده آنها محاسبه می‌شود. این ویژگی‌ها می‌تواند اطلاعاتی مانند جنسیت، سن، محل سکونت، تحصیلات و شغل باشد که برای همهٔ کاربران در قالب ماتریسی نگهداری می‌شود. هر سطر از ماتریس مذکور برداری از ویژگی‌های یک کاربر است. در خوشه‌بندی، تلاش می‌شود کاربرانی که بیشترین شباهت را دارند، در خوشه‌های همسان قرار گیرند و معیار شباهت نیز با محاسباتی تعیین می‌شود که روی بردار ویژگی کاربران انجام می‌گیرد (ووزالیس و مارگریتز، ۲۰۰۴).

روش‌شناسی پژوهش

در این بخش به‌صورت گام‌به‌گام مراحل سیستم پیشنهاددهنده این پژوهش شرح داده مـی‌شـود. شکل ۲ سیستم پیشنهادی پژوهش حاضر را نشان می‌دهد.



شکل ۲. معماری سیستم پیشنهادی

روش پیشنهادی رویکردی ترکیبی از سیستم پیشنهاددهندهٔ پایه با رویکرد خوشه‌بندی بـرای محاسبهٔ شباهت کاربران است. در روش پایه، از پالایش همکارانهٔ مبتنی بر کاربر استفاده می‌شود.

در این روش از تاریخچهٔ امتیازهای داده‌شدهٔ کاربران به کالاها، برای پیش‌بینی امتیاز اقلام دیده‌نشدهٔ کاربر دیگر استفاده می‌شود. از سویی، برای کاربران تازه‌وارد به وب‌سایت که تاریخچهٔ نظرها برای پیش‌بینی ترجیحات کاربر کافی نیست، از روش خوشه‌بندی مبتنی بر اطلاعات جمعیت‌شناختی کاربران استفاده می‌شود. فرض بر این است که کاربران بر اساس این اطلاعات می‌توانند ترجیحات مشابهی داشته باشند. در این روش، کاربران بر اساس شباهت اطلاعات فردی به خوشه‌هایی تقسیم می‌شوند. از مدل ایجادشده برای تعیین خوشهٔ کاربر جدید (خوشه‌ای با کمترین فاصله از مرکز) استفاده می‌شود. با محاسبهٔ شباهت درون‌خوشه‌ای (میان کاربر جدید و کاربران خوشهٔ تعلق‌یافته)، رأی کاربر به کالاها پیش‌بینی می‌شود. در نهایت تعدادی کالا (N) با بیشترین امتیاز پیش‌بینی‌شده به کاربر نشان داده خواهد شد. در ادامه مراحل معماری به‌صورت مختصر تشریح می‌شود.

انتخاب و پیش‌پردازش داده‌ها

در حوزهٔ داده‌کاوی و الگوریتم‌های یادگیری، بهره‌مندی از مجموعهٔ دادهٔ مناسب، از اهمیت زیادی برخوردار است. مطالعهٔ مجموعهٔ دادهٔ انتخاب‌شده از نظر کمی و کیفی و نیز انتخاب ویژگی‌های مناسب برای توصیف نمونه‌ها، یکی از گام‌های ضروری و مفصل در این حوزه محسوب می‌شود.

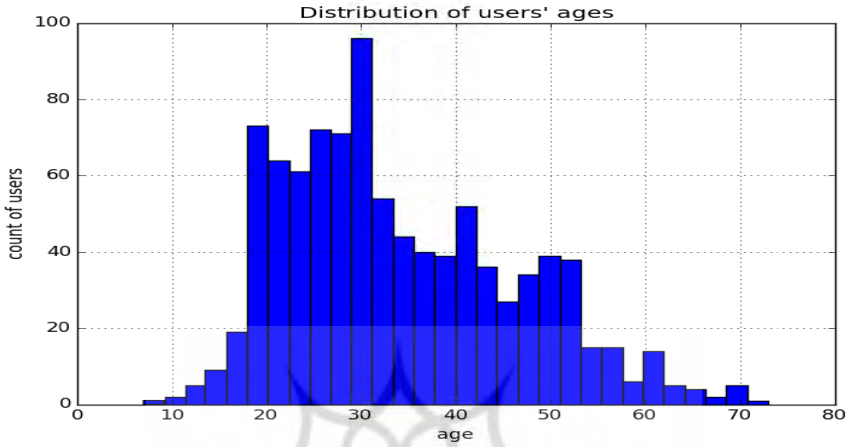
مجموعهٔ دادهٔ استفاده‌شده برای تحلیل نتایج این پژوهش، علاقه‌مندی کاربران به مجموعه‌ای از فیلم‌ها است. این مجموعهٔ داده، شامل اطلاعات فردی کاربران، فیلم‌ها، انواع و تعداد آنها و اطلاعات امتیازدهی است و در آن صدهزار رتبه‌بندی ستاره‌ای از ۹۴۳ کاربر برای ۱۶۸۲ فیلم وجود دارد که در آن مشتری می‌تواند به محصول از ۱ (کمترین) تا ۵ (بیشترین) امتیاز دهد. در این مجموعهٔ داده، هر کاربر دست کم به ۲۰ فیلم امتیاز داده است. اطلاعات مجموعهٔ داده در جدول ۲ آورده شده است.

جدول ۲. اطلاعات مجموعهٔ دادهٔ وب‌سایت فیلم

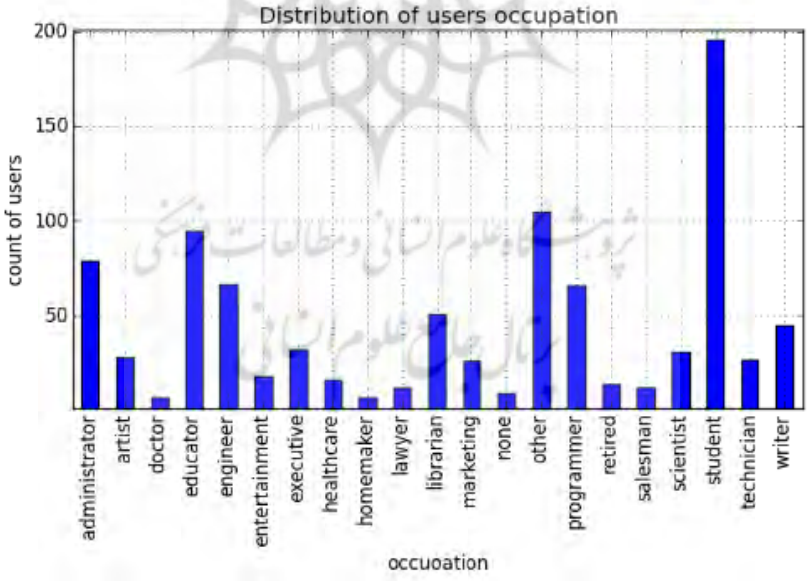
عوامل جمعیت‌شناختی	تعداد فیلم	تعداد کاربران	تعداد رتبه‌بندی
✓	۱۶۸۲	۹۴۳	۱۰۰۰۰۰

این مجموعهٔ داده، اطلاعات جمعیت‌شناختی کاربران شرکت‌کننده در رتبه‌بندی فیلم‌ها را نیز دربردارد. از میان ویژگی‌ها، این پژوهش ویژگی‌های سن، جنسیت و شغل را برای خوشه‌بندی کاربران انتخاب کرده است. توزیع فراوانی کاربران بر اساس ویژگی سن در شکل ۳- الف نشان داده شده است و مشاهده می‌شود کاربران ۱۷ تا ۶۱ سال، بیشترین فراوانی را به خود اختصاص

داده‌اند. همچنین توزیع فراوانی کاربران بر اساس سن و شغل در شکل ۳- ب مشاهده می‌شود. جدول ۳ نیز، میانگین آرای گروه‌های سنی به پنج فیلم نمونه را نمایش می‌دهد.



(الف)



(ب)

شکل ۳. نمودار توزیع فراوانی کاربران بر اساس سن (الف) و شغل (ب)

جدول ۳. میانگین آرای گروه‌های سنی به پنج فیلم نمونه

گروه سنی / عنوان فیلم	۱۰-۱۹	۲۰-۲۹	۳۰-۳۹	٤۰-٤۹	۵۰-۷۹
فیلم ۱ (۱۹۸۲)	۳/۸۵	۳/۵۷	۳/۷۷	۴/۱۴	۴/۳۸
فیلم ۲ (۱۹۸۰)	۴/۸۶	۴/۳۱	۴/۰۷	۴/۲۱	۴/۰۲
فیلم ۳ (۱۹۹۶)	۳/۷۱	۳/۵۹	۳/۶۴	۳/۵۰	۳/۸۰
فیلم ۴ (۱۹۹۶)	۳/۸۸	۳/۹۵	۴/۲۷	۴/۱۶	۴/۳۴
فیلم ۵ (۱۹۹۴)	۴/۰۰	۳/۸۲	۳/۹۰	۳/۷۶	۳/۹۱

اطلاعات هر کاربر در برداری نمایش‌داده می‌شود. مطالعهٔ ویژگی‌ها در مجموعهٔ داده‌ها، می‌تواند به شکلی جامع‌تر و با معیارهای دقیق‌تری انجام شود که با توجه به اهداف مقاله، از ارائهٔ نتایج بیشتر خودداری شده است.

برای محاسبهٔ شباهت‌ها و انجام روش خوشه‌بندی، باید داده‌ها به‌شکل مناسبی بازنمایی شوند. از این رو در مرحلهٔ پیش‌پردازش، از سه ویژگی یادشده بردارهایی با ۲۷ ویژگی بر اساس مدل ذکرشده ایجاد می‌شود (جدول ۴). به این ترتیب مجموعهٔ دادهٔ کاربران در قالب ماتریس ۹۴۳ × ۲۷ قرار می‌گیرد.

جدول ٤. نحوهٔ نگاشت اطلاعات فردی به فضای ویژگی

شمارهٔ ویژگی	محتوای ویژگی	شرح
۱	۱۹ =< سن	با قرار گرفتن کاربر در یکی از این بازه‌ها، برای ویژگی متناظر با آن بازه مقدار ۱ تعلق می‌گیرد و برای ویژگی‌ها دیگر مقدار صفر داده می‌شود.
۲	۲۹ = >سن > ۱۹	
۳	۳۹ = > سن > ۲۹	
۴	۴۹ = > سن > ۳۹	
۵	۴۹ < سن	
۶	مرد یا زن	مرد =۰ و زن =۱
۷-۲۷	شغل	ویژگی مشخص‌کنندهٔ شغل کاربر، عدد ۱ و سایر ویژگی‌های این بخش مقدار صفر می‌گیرند.

خوشه‌بندی کاربران

در روش پیشنهادی، از الگوریتم خوشه‌بندی K-Means استفاده می‌شود. K در این روش نمایش‌دهندهٔ تعداد خوشه‌ها است و در پیاده‌سازی‌ها، برای آن مقدار ۸ در نظر گرفته شده است. روند کلی این الگوریتم به‌صورت زیر است:

۱. K خوشهٔ اولیه ایجاد می‌شود؛
۲. فاصلهٔ هر کاربر با مراکز خوشه‌ها از طریق رابطهٔ کسیسنوسی محاسبه می‌شود؛
۳. کاربر به خوشه‌ای تعلق می‌گیرد که کمترین فاصله یا به بیانی، بیشترین شباهت را به آن دارد، این کار برای همهٔ کاربران تکرار می‌شود؛
۴. مراکز خوشه‌ها به‌روزرسانی می‌شود؛
۵. مراحل ۲ تا ۴ تا زمانی اجرا می‌شود که تغییری در مراکز خوشه‌ها ایجاد نشده باشد و کاربران در خوشهٔ خود جای گرفته باشند (برخین، ۲۰۰۶).

به‌دنبال خوشه‌بندی، کاربران به گروه (خوشه) های مجزایی تقسیم می‌شوند که بر اساس ویژگی‌ها یاد شده بیشترین شباهت را با یکدیگر دارند.

پیش‌بینی امتیازدهی بر اساس خوشه

پس از خوشه‌بندی، به‌منظور پیش‌بینی نظر کاربر به یک فیلم، از میانگین وزن‌دار امتیازهای اعضای خوشهٔ کاربر به آن، استفاده می‌شود. وزن هر نظر بر اساس فاصلهٔ کاربر رأی‌دهنده نسبت به کاربر فعال، محاسبه می‌شود.

ترکیب امتیازهای تولیدشده

گام اصلی در روش ترکیبی ارائه‌شده، ترکیب شباهت خوشه‌ای و شباهت پالایش همکارانه به‌منظور تولید پیش‌بینی امتیاز کاربر به کالاها است. برای این کار، رابطهٔ پالایش همکارانهٔ مبتنی بر کاربر به‌منظور اعمال خوشه‌بندی بر اساس داده‌های فردی به‌شکل رابطهٔ ۷ تغییر داده می‌شود:

$$r_{u,i} = k \sum_{u' \in U} Hsim(u,u') \times r_{u',i}$$ رابطهٔ ۷)

که در آن Hsim به‌صورت جمع وزن‌داری از شباهت درون‌خوشه‌ای (clusSim) و شباهت مبتنی بر پالایش همکارانه (sim)، محاسبه می‌شود (رابطهٔ ۸).

$$Hsim(u,u') = (1-\alpha) \times sim(u,u') + \alpha \times clusSim(u,u')$$ رابطهٔ ۸)

پیشنهاد N کالا با امتیاز برتر

در این مرحله کالاها براساس امتیازهای پیش‌بینی‌شده مرتب می‌شوند و N کالا با بیشترین امتیاز به کاربر پیشنهاد خواهد شد. در واقع از میان کالاهای دیده‌نشدهٔ کاربر، این اقلام می‌توانند پیشنهادهای خوبی شمرده شوند.

یافته‌های پژوهش

محیط پیاده‌سازی

به‌منظور پیاده‌سازی سیستم پیشنهاددهنده، از چارچوب برنامه‌نویسی جاوا (JAVA) استفاده شده است. جاوا چارچوب و زبان برنامه‌نویسی متن‌باز[1] است. به همین دلیل می‌توان علاوه‌بر پیاده‌سازی روش‌های مد نظر در این محیط برنامه‌نویسی، از ابزارهای متن‌باز مرتبط نیز بهره برد. در زمینهٔ سیستم‌های پیشنهاددهنده و داده‌کاوی[2]، محققان ابزارهایی بر اساس جاوا ارائه کرده‌اند که از میان آنها می‌توان به ماهووت[3] و لنزکیت[4] اشاره کرد. از این رو، بر اساس مطالعات انجام‌گرفته، ابزار ماهووت برای پیاده‌سازی انتخاب شد. ماهووت ابزاری است که شرکت آپاچی[5] برای پیاده‌سازی الگوریتم‌های یادگیری ماشین و داده‌کاوی ارائه کرده است. در ادامه به نتایج پیاده‌سازی و آزمایش‌های اجراشده پرداخته می‌شود.

نتایج شبیه‌سازی

برای اعمال آزمایش‌ها، مجموعهٔ دادهٔ استفاده‌شده به دو دستهٔ آموزشی و آزمایشی تقسیم می‌شود. مجموعهٔ آموزشی برای ساخت مدل پیش‌بینی به‌کار می‌رود. پس از ایجاد مدل با بهره‌مندی از مجموعهٔ آزمایشی، دقت نتایج مدل ساخته‌شده، ارزیابی می‌شود. برای ارزیابی نتایج از روش ۱۰-fold استفاده می‌شود. در این روش، میانگین خطای ده بار تکرار آزمایش الگوریتم‌ها به‌دست می‌آید و با هم مقایسه می‌شود. در هر بار تکرار، ۹۰ درصد داده‌ها به‌مثابهٔ مجموعهٔ آموزشی، برای ساخت مدل به‌کار می‌رود و ۱۰ درصد باقی‌مانده، به‌منظور آزمون مدل استفاده می‌شود.

از آنجاکه در سیستم پیشنهاددهنده، هدف، نمایش اقلام با بیشترین امتیاز پیش‌بینی‌شده به کاربر است، برای مقایسه‌ها از $n \in \{۵,۱۰,۲۰\}$ فیلم برتر، در محاسبهٔ میزان خطای روش‌ها

---

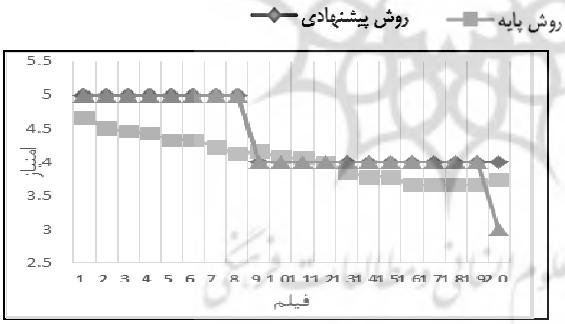
1. Open source
2. Data mining
3. Mahout
4. Lensekit
5. Apachi

استفاده می‌شود. میزان خطا بر اساس اختلاف مقدار پیش‌بینی‌شدهٔ روش از مقدار واقعی رأی کاربر، برآورد می‌شود. ریشهٔ میانگین مربعات خطا (RMSE) معیار خطای روش‌ها است که از رابطهٔ ۹ به‌دست می‌آید.

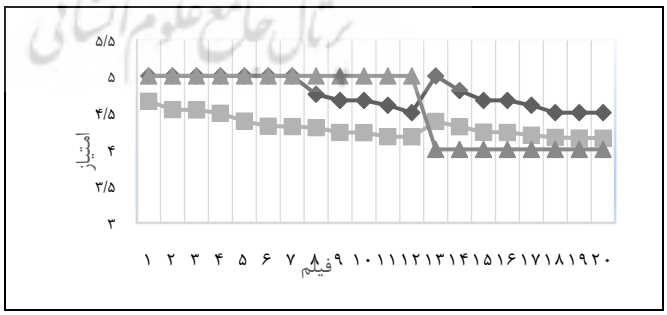
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{real,i} - X_{model,i})^2}{n}} \quad \text{رابطهٔ ۹)}$$

که در آن؛ $X_{real,i}$ امتیاز کاربر به کالای i و $X_{model,i}$ امتیاز پیش‌بینی‌شدهٔ کاربر به کالای i است.

خروجی روش پیشنهادی برای چهار کاربر تصادفی به دو روش مذکور، در شکل ۴ نشان داده شده است. مشاهده می‌شود روش پیشنهادی با بهره‌گیری از روش ترکیبی مبتنی بر خوشه‌بندی کاربران، توانسته است پیش‌بینی نزدیک‌تری به امتیازهای واقعی کاربران ارائه دهد. در حالت بهینه، روش پیشنهادی با میزان خطای صفر توانسته است ۲۰ فیلم را به کاربر شمارهٔ ۴۱۶ پیشنهاد دهد. خطای رخداده برای این کاربران در شکل ۵ نشان داده شده است. برای همهٔ کاربران نمونه، روش پیشنهادی نتایج بهتری داشته است.



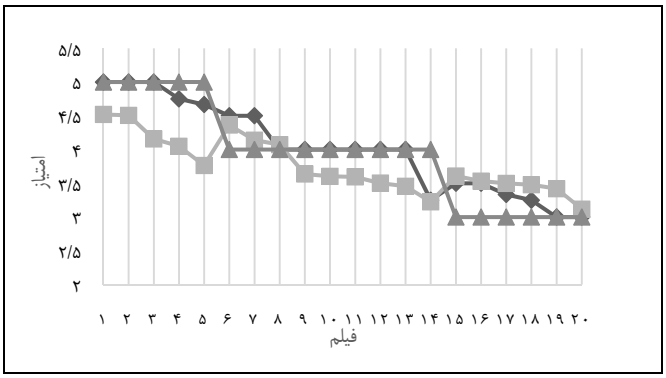
الف) کاربر ۱۹۶



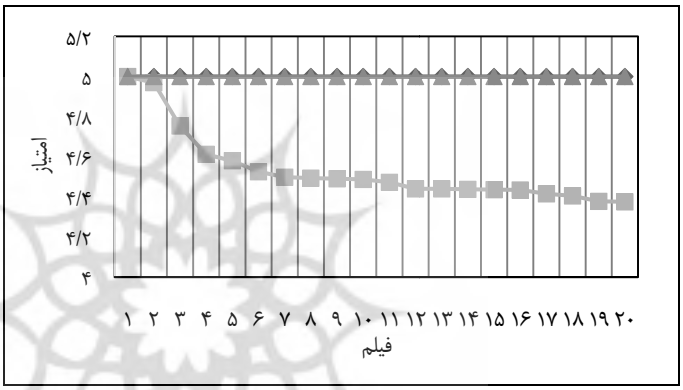
ب) کاربر ۱۱

شکل ٤. خروجی روش پیشنهادی برای چهار کاربر تصادفی

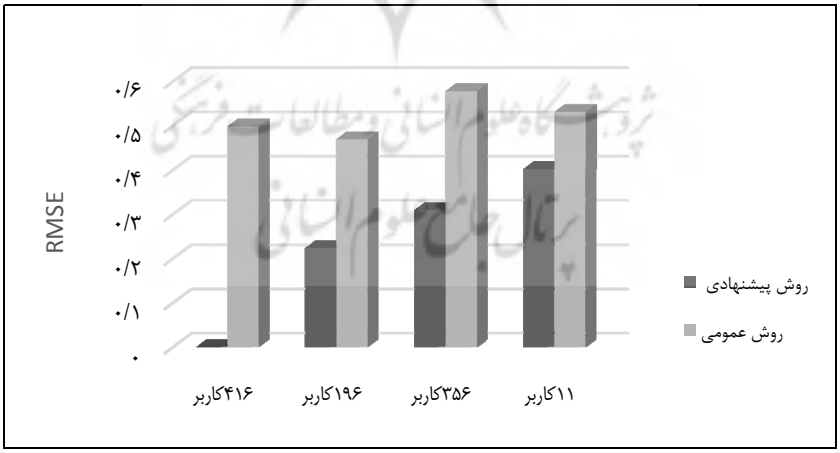
ج) کاربر ۳۵۶



د) کاربر ۴۱۶

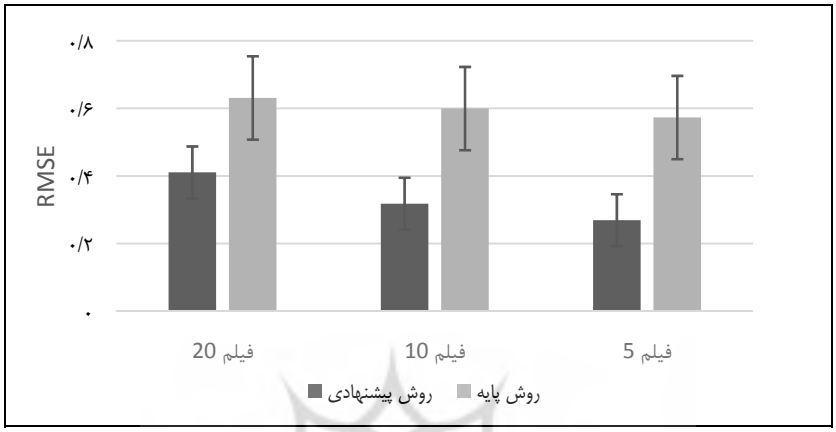


ادامهٔ شکل ٤.



شکل ۵. مقایسهٔ میزان خطای روش پیشنهادی و روش پایه برای کاربران نمونه

بر اساس رویکرد شرح داده‌شده، آزمایش‌هایی روی مجموعهٔ داده برای همهٔ کاربران صورت گرفت. نتایج بررسی‌ها در شکل ۶ نشان داده شده است.



شکل ٦. مقایسهٔ میزان خطای روش پیشنهادی و روش پایه

هریک از مقادیر RMSE، به‌صورت میانگین ده بار اجرای الگوریتم‌ها برای مجموعهٔ دادهٔ آزمایشی آورده شده است. مشخص است روش پیشنهادی در همهٔ حالت‌ها، نسبت به رویکرد پایه عملکرد بهتری دارد. برای ۲۰ فیلم پیشنهادشده، مقدار خطای روش پایه ۰/۶۳ است و برای روش پیشنهادی مقدار خطا ۰/۴۱ به‌دست آمده است. همچنین بر اساس یافته‌ها، نتایج الگوریتم ترکیبی برای فیلم‌های پیشنهادشده با بیشترین امتیاز ـ که در رتبه‌های نخست به کاربر نشان داده می‌شود ـ دقت به مراتب بیشتری نسبت به روش پایه دارد. برای اثبات تفاوت به‌دست‌آمده، از آزمون تی استفاده می‌شود. بر این اساس فرضیهٔ پژوهش چنین بنا می‌شود:

فرض صفر: میانگین خطای روش‌ها در ده بار اجرا با یکدیگر اختلاف معناداری ندارد.

فرض یک: اختلاف میانگین خطای روش‌ها در ده بار اجرا با یکدیگر معنادار است.

$$\begin{cases} H_0: \mu_d = 0 \\ H_1: \mu_d \neq 0 \end{cases}$$

در جدول ۵، نتایج اجرای آزمون تی درج شده است. همان‌طور که مشاهده می‌شود با توجه به اینکه در سطح اطمینان ۹۵ درصد، سطح معناداری کمتر از ۵ درصد است (sig = ۰/۰۰۲)، در سطح اطمینان ۹۵ درصد، فرض صفر رد می‌شود و فرض مقابل، یعنی فرض یک به تأیید می‌رسد؛ به این معنا که در میانگین خطاهای روش‌ها در دو روش پایه و پیشنهادی اختلاف وجود

دارد. همچنین با توجه به اینکه حد پایین و بالا در سطح اطمینان ۹۵ درصد مثبت است، می‌توان نتیجه گرفت میانگین خطای روش پایه بیشتر از میانگین خطای آن در روش پیشنهادی است.

جدول ۵. مقادیر تحلیل آزمون تی

	اختلاف زوج‌ها			سطح معناداری (sig)
	میانگین	اختلاف در سطح اطمینان ۹۵٪		
		پایین	بالا	
روش پیشنهادی	۰/۱۸۵۴۰۱۴۹۶	۰/۰۸۷۶۸۴۰۵	۰/۲۸۳۱۱۸۹۴۲	۰/۰۰۲

به بیان دیگر، تحلیل نتایج نشان می‌دهد سیستم پیشنهاددهنده با الگوریتم معرفی‌شدهٔ این پژوهش، می‌تواند با خوشه‌بندی کاربران بر اساس اطلاعات جمعیت‌شناختی، نتایج بسیار با کیفیت‌تری نسبت به الگوریتم‌های پایه ارائه کند. یکی از مشکلات روش‌های پالایش همکارانه، مسئلهٔ شروع سرد برای کاربرانی است که به تازگی وارد سایت شده‌اند و تاریخچهٔ نظر برای آنان وجود ندارد. در این حالت روش‌های سنتی نمی‌توانند عمل پیش‌بینی را انجام دهند. با این حال روش پیشنهادی مشکل شروع سرد را با خوشه‌بندی کاربران بر اساس اطلاعات جمعیت‌شناختی رفع می‌کند. بر این اساس می‌توان کالاهایی را به کاربر پیشنهاد داد که با میزان خطای کمتر، از جانب مشتری استقبال شود و در رضایت کاربران از وب‌سایت فروشگاه و وفاداری آنها تأثیرگذار باشد.

نتیجه‌گیری و پیشنهادها

با گسترش هرچه بیشتر فروشگاه‌های آنلاین و انبوه محصولات ارائه‌شدهٔ آنها، وب‌سایت‌ها باید در جست‌وجوی راهی باشند تا با درک علاقه‌مندی کاربران و ارائهٔ پیشنهادها شخصی‌سازی‌شده، آنان را به خرید از خود ترغیب کنند. ساخت پروفایل مشتریان، گام نخست در این راستا شمرده می‌شود که دربرگیرندهٔ اطلاعاتی چون ویژگی‌های جمعیت‌شناختی، سابقهٔ خرید، بازدید، نظرها و رتبه‌های داده‌شده به کالاهای فروشگاه است. گام بعدی، استفاده از سیستمی است که با پردازش هوشمند پروفایل کاربران، علاقه‌مندی آنان را یاد بگیرد. مطالعات نشان می‌دهد سیستم‌های پیشنهاددهنده این کار را به خوبی انجام می‌دهند، اما در عین حال با کاستی‌هایی نیز مواجه‌اند. از این رو باید با ارائهٔ الگوریتم‌های کاراتر، عملکرد سیستم و نتایج آنها را بهبود بخشید.

در این پژوهش رویکردی ترکیبی از سیستم پیشنهاددهندهٔ پایه، روش پالایش همکارانهٔ مبتنی بر کاربر و روش خوشه‌بندی کاربران بر اساس ویژگی‌های جمعیت شناختی آنها، ارائه شد.

در روش پایه، از تاریخچهٔ امتیازهای کاربران به کالاهایی که مشاهده یا خریداری شده‌اند؛ برای پیش‌بینی امتیاز اقلام دیده‌نشدهٔ کاربر استفاده می‌شود، اما این روش برای کاربران و کالاهای جدید که چنین تاریخچه‌ای برای آنان موجود یا کافی نیست، با مشکلاتی روبه‌روست که به مسئلهٔ «شروع سرد» معروف است. برای حل این مشکل، در روش پیشنهادی از روش خوشه‌بندی بهره برده شد و کاربران بر اساس ویژگی‌های جمعیت‌شناختی به خوشه‌هایی تقسیم شدند. پس از خوشه‌بندی، به‌منظور پیش‌بینی نظر کاربر دربارهٔ هر فیلم، از میانگین وزن‌دار امتیازهای اعضای خوشهٔ کاربر به آن فیلم بر اساس فاصلهٔ او نسبت به اعضای خوشه، استفاده شد.

نتایج پیاده‌سازی نشان داد مجذور میانگین مربعات خطای الگوریتم پیشنهادی، به‌مراتب کمتر از الگوریتم پایه است. این مسئله به‌معنای آن است که سیستم با استفاده از این روش، پیشنهادهای بهتر و نزدیک‌تری به خواسته‌های کاربران ارائه می‌کند و پیش‌بینی‌ها به امتیازهای واقعی کاربران نزدیک‌تر است. در واقع استفاده از اطلاعات جمعیت‌شناختی در روش خوشه‌بندی، توانسته است نتایج سیستم پیشنهادی را بهبود بخشد. مطالعهٔ کارپن، جزیاسن و آساف (۲۰۱۱) تأیید می‌کند ویژگی‌های جمعیت‌شناختی کاربران، بر تصمیم‌گیری‌های خرید آنلاین آنها تأثیرگذار است. بنابراین با لحاظ‌کردن این ویژگی‌ها، می‌توان علایق کاربران را بهتر شناسایی و ردیابی کرد و احتمال خرید و رضایت کاربران را افزایش داد. نتایج حاکی از آن است که عملکرد سیستم در برابر مشکل شروع سرد بهبود یافته است و پیشنهادهای بهتری ارائه می‌کند.

برای مطالعات بعدی، پیشنهاد می‌شود تحلیل آماری دقیقی از ویژگی‌های جمعیت‌شناختی به عمل آید و با تعیین میزان تأثیر هریک از ویژگی‌ها بر امتیازهای کاربران، از آنها به‌صورت معیار وزن‌دار برای محاسبهٔ شباهت در خوشه‌بندی استفاده شود. همچنین می‌توان از اطلاعات تحلیل انواع محصولات وب‌سایت، مانند فیلم و تعیین ویژگی‌هایی آن، مانند نوع فیلم، بازیگران و کارگردانان و سایر ویژگی‌های تعیین‌کننده در انتخاب محصول و در نظر گرفتن ضریب اهمیت برای آنها در محاسبات استفاده کرد و با یادگیری علاقه‌مندی کاربران، پیشنهادهای بهتری ارائه داد.

یکی از محدودیت‌های این پژوهش، دسترسی نداشتن به مجموعهٔ دادهٔ مناسب برخوردار از ویژگی‌های جمعیت‌شناختی بود. در صورت وجود فروشگاه آنلاینی که امکان استفاده از اطلاعات یادشده در آن فراهم باشد، می‌توان نتایج پژوهش را با بررسی تأثیر سیستم پیشنهاددهنده بر فروش فروشگاه و درصد استقبال کاربران از کالاهای پیشنهادی، همراه کرد.

References

Abbasi, M., Monsefi, R. & Astiri, A. (2011). A hybrid recommender system using the natural process technique og Persian language. *The first international conference on Persian language processing, Semnan, Semnan University*, 6 September. (*in Persian*)

Adomavicius, G. & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6): 734-749.

Anderson, C. & Hiralall, M. (2009). Recommender systems for e-shops. *Business Mathematics and Informatics paper*. Vrije Universiteit, Amsterdam.

Astiri, A., Kahani, M. & Pour-Gholami, F. (2011). A survey of calculating the semantic similarity of words and sentences using a network of words. *The first international conference on Persian language processing*, Semnan, Semnan University, 6 September. (*in Persian*)

Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data* (pp. 25-71). Springer, Berlin Heidelberg. DOI: 10.1007/3-540-28349-8_2.

Cho, Y. H., Kim, J. K. & Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems with Applications*, 23(3): 329-342.

Choi, K., Yoo, D., Kim, G. & Suh, Y. (2012). A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. *Electronic Commerce Research and Applications*, 11(4): 309-317.

Dai, Y., Ye, H. & Gong, S. (2009). Personalized recommendation algorithm using user demography information. In *Knowledge Discovery and Data Mining, 2009. WKDD 2009. Second International Workshop on* (pp. 100-103). IEEE.

Gao, F., Xing, C., Du, X. & Wang, S. (2007). Personalized service system based on hybrid filtering for digital library. *Tsinghua Science & Technology*, 12(1): 1-8.

Gupta, J. & Gadge, J. (2014). A framework for a recommendation system based on collaborative filtering and demographics. In *Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on* (pp. 300-304). IEEE.

Iwata, T. (2008). *Probabilistic user behavior models in online stores for recommender systems*. Doctoral Thesis, Graduate School of Informatics, Kyoto University.

Josiassen, A., Assaf, A. G. & Karpen, I. O. (2011). Consumer ethnocentrism and willingness to buy: Analyzing the role of three demographic consumer characteristics. *International Marketing Review*, 28(6): 627-646.

Leimstoll, U. & Stormer, H. (2007). Collaborative recommender systems for online shops. *AMCIS 2007 Proceedings*, 156. Available in: http://aisel.aisnet.org /cgi/viewcontent.cgi?article=1666&context=amcis2007.

Liang, Z., Bo, X. & Jun, G. (2009). An Approach of Finding Localized Preferences Based-On Clustering for Collaborative Filtering. In *Web Information Systems and Mining, 2009. WISM 2009. International Conference on* (pp. 19-22). IEEE.

Linden, G., Smith, B. & York, J. (2003). Amazon. Com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1): 76-80.

Moghaddam, S. G. & Selamat, A. (2011). A scalable collaborative recommender algorithm based on user density-based clustering. *Data Mining and Intelligent Information Technology Applications (ICMiA), 2011 3rd International Conference on* (pp. 246-249). IEEE, 24-26 Oct.

Peck, J. & Childers, T. L. (2006). If I touch it I have to have it: Individual and environmental influences on impulse purchasing. *Journal of business research*, 59(6): 765-769.

Schafer, J. B., Konstan, J. A. & Riedl, J. (2001). E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5 (1-2): 115-153.

Shabib, N. & Nemat Bakhsh, M.A. & Dehkurdi, S.A. (2007). Recommende systems application in e-commerce. *4th National e-Commerce Conferenc*, Ministry of Commerce, Tehran. (*in Persian*)

Su, X. & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009 (4). DOI:10.1155/2009/ 421425.

Vandenbulcke, V., Lecron, F., Ducarroz, C., & Fouss, F. (2013). Customer segmentation based on a collaborative recommendation system: Application to a mass retail company. In *Proceedings of the 42nd Annual Conference of the European Marketing Academy*.

Vozalis, M. & Margaritis, K. G. (2004). Collaborative filtering enhanced by demographic correlation. In *AIAI Symposium on Professional Practice in AI, of the 18th World Computer Congress.*

Yoon, V. Y., Hostler, R. E., Guo, Z. & Guimaraes, T. (2013). Assessing the moderating effect of consumer product knowledge and online shopping experience on using recommendation agents for customer loyalty. *Decision Support Systems*, 55(4): 883-893.

Zeng, Z. (2009). An intelligent e-commerce recommender system based on web mining. *International journal of business and management*, 4(7): 10.