نشریه علمی‌ترویجی محاسبات نرم
سال هفتم/ شمارهٔ اول/ بهار و تابستان ۱۳۹۷/ صفحهٔ ۵۰ـ۶۱

دریافت مقاله: ۱۳۹۶/۰۶/۲۸
پذیرش مقاله: ۱۳۹۸/۱۱/۱۴

طراحی سیستم پیشنهاددهندهٔ ترکیبی برای حل مشکل شروع سرد به کمک خوشه‌بندی و الگوریتم ازدحام ذرات آشوبی

نفیسه موسی‌الرضایی گلیان[۱]، جواد حمیدزاده[۲*]

[۱] دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه بین‌المللی امام رضا(ع)، مشهد، ایران
n.moosarrezayi@imamreza.ac.ir
[۲] استادیار، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی سجاد، مشهد، ایران
j_hamidzadeh@sadjad.ac.ir

چکیده: امروزه یکی از مهم‌ترین چالش‌های افزایش اطلاعات، یافتن اطلاعات مورد علاقه از بین انبوه داده‌هاست. به این موضوع در طراحی سایت‌های تعاملی همواره توجه شده است. سیستم‌های پیشنهاددهنده برای حل این مسئله به وجود آمده‌اند تا به کاربران برای رسیدن به اطلاعات مورد نظرشان کمک کنند؛ اما این سیستم‌ها محدویت‌هایی دارند. یکی از مهم‌ترین چالش‌های پیش‌روی سیستم‌های پیشنهاددهنده، مشکل شروع سرد است. این مشکل زمانی به وجود می‌آید که یک کاربر (قلم‌داده) جدید وارد سیستم می‌شود. عدم وجود اطلاعات قبلی از این کاربر (قلم‌داده) باعث می‌شود سیستم نتواند به‌طور عادی پیشنهادها را تولید کند. در این مقاله برای حل مشکل شروع سرد کاربر، روش جدیدی به کمک ترکیب مدل‌های مبتنی بر محتوا و فیلترمشارکتی ارائه شده است. در این روش، فهرست پیشنهادی دارای ویژگی‌هایی مانند کیفیت بالای قلم‌داده‌های پیشنهادی و تنوع آن‌هاست که دامنهٔ اطلاعات دریافتی از کاربر را به‌سرعت گسترش می‌دهد، به همین دلیل کاربران را سریع‌تر از حالت شروع سرد خارج می‌کند. همچنین با استفاده از اطلاعات دموگرافیک کاربر، سعی شده قلم‌داده‌های فهرست پیشنهادی به نحوی انتخاب شوند که به علایق کاربر نزدیک‌تر باشند تا دقت بیشتر شود. نتایج ارزیابی روش پیشنهادی نشان می‌دهد میزان خطای **MAE** و **RMSE** نسبت به روش‌های موجود تا حد مطلوبی کاهش یافته است.

واژه‌های کلیدی: سیستم پیشنهاددهنده، مشکل شروع سرد، روش مبتنی بر محتوا، روش فیلتر مشارکتی، تنوع.

۱. مقدمه

با توجه به رشد روزافزون صفحات وب و به‌تبع آن تجارت الکترونیک، میزان داده هایی که کاربران با آن‌ها مواجه می‌شوند، روزبه‌روز افزایش می‌یابد؛ به همین دلیل جست‌وجو و انتخاب قلم‌داده های[1] مورد علاقه از بین انبوهی از اطلاعات، کاری سخت و زمان‌بر است. از آنجایی که در دنیای وب رقابت برای جذب و حفظ کاربران (مشتریان) افزایش یافته است، اهمیت رضایت‌مندی کاربران از صفحات وب موجب به وجود آمدن سیستم‌های پیشنهاددهنده[2] شد. این سیستم‌ها علاوه بر اینکه با پیشنهادهای خود به کاربران موجب رضایت‌مندی آنان می‌شوند، با دریافت بازخوردهای آن‌ها، به توسعه‌دهندگان صفحات وب نیز کمک می‌کند تا طراحی فضای کاری خود را بهینه و مطابق سلایق کاربران خود به‌روز کنند. سیستم‌های پیشنهاددهنده سیستم های اطلاعاتی هستند که با کمک منبع داده‌های مختلف مانند اطلاعات شخصی کاربران، عقاید آن‌ها و از این قبیل سعی می‌کنند کاربران را در پیداکردن و انتخاب موارد مورد علاقه‌شان مانند کتاب [۱]، فیلم [۲]، برنامه‌های تلویزیونی [۳]، موسیقی [۴]، آموزش الکترونیک [۵]، تجارت الکترونیک [۶]، صفحات وب [۷] و... یاری کنند.

۱ـ۱. تعریف مسئله و نوآوری

یکی از مشکلات اساسی پیش روی سیستم‌های پیشنهاددهنده که حجم زیادی از تحقیقات را متوجه خود کرده، مشکل شروع سرد[3] می‌باشد که بر دو قسم است: شروع سرد کاربر جدید و شروع سرد قلم‌دادهٔ جدید. مشکل شروع سرد کاربر جدید زمانی رخ می‌دهد که یک کاربر جدید وارد سیستم می‌شود، درحالی‌که هیچ اطلاعاتی از علایق و سلایق، سابقهٔ خرید، بازدید یا امتیازدهی از وی موجود نیست. بنابراین سیستم برای تهیهٔ پیشنهاد برای این کاربر دچار مشکل می‌شود. به همین ترتیب مشکل شروع سرد قلم‌دادهٔ جدید نیز زمانی اتفاق می‌افتد که یک قلم‌دادهٔ جدید به سیستم اضافه می‌شود درحالی‌که تاکنون هیچ کاربری به این قلم‌داده امتیازی نداده و بازدیدی نداشته است. بنابراین این قلم‌داده نیز در فهرست‌های پیشنهادی سیستم جایی نخواهد داشت.

کارهای انجام‌شده برای حل مشکل شروع سرد کاربر جدید به سه دستهٔ کلی تقسیم می‌شوند [۸]:

۱. استفاده از مجموعه داده‌های کمکی مانند برچسب‌های اجتماعی، نظریات کاربران و...؛

۲. انتخاب شبیه‌ترین گروه کاربران؛

۳. تهیهٔ پیشنهادات بر اساس روش‌های ترکیبی.

در این مقاله سعی شده است مشکل شروع سرد کاربر جدید به کمک روش‌های ترکیبی مبتنی بر محتوا و فیلتر مشارکتی حل شود. به همین منظور برای تهیهٔ فهرست پیشنهادی سعی شده است از معیارهای تأثیرگذار در پیشنهادات سیستم های پیشنهاددهنده، شامل تنوع، خوشبختی و شباهت به نحو مناسبی استفاده شود. مزیت روش پیشنهادی نسبت به سایر روش‌های موجود این است که علاوه بر کاهش تعداد دفعات مشارکت کاربر در تکمیل اطلاعات، با بهینه‌سازی فهرست‌های پیشنهادی با معیارهای ذکرشده، هر بار دامنهٔ گسترده‌تری از علایق کاربر به دست می‌آید که این امر موجب خروج سریع‌تر کاربر از حالت شروع سرد می‌شود. این کاهش زمان نقش مهمی در تعیین میزان رضایت کاربر از سیستم دارد.

ساختار ادامهٔ مقاله به شرح زیر است: در بخش دوم ابتدا به بررسی پیشینهٔ سیستم‌های پیشنهاددهنده پرداخته می‌شود. سپس در بخش سوم مفاهیم اولیهٔ سیستم‌های پیشنهاددهنده، شامل اهداف این سیستم‌ها، مدل‌های پایه‌ای آن‌ها و ویژگی‌های یک فهرست پیشنهادی خوب مرور می‌شوند. پس از آن روش پیشنهادی در بخش چهارم ارائه شده و آزمایش‌ها و ارزیابی مدل نیز در بخش پنجم توضیح داده شده است. در بخش آخر نیز نتیجه‌گیری و کارهای آتی بیان شده است.

۲. کارهای گذشته

در مقالهٔ [۱۰] یک نوع استفادهٔ جدید از داده‌های دموگرافیک کاربر به‌جای استفاده از تاریخچهٔ امتیازات برای تهیهٔ پیشنهادات در نظر گرفته می‌شود تا از به وجود آمدن مشکل کاربر جدید جلوگیری شود. برای این منظور یک چارچوب برای ارزش‌دهی استفادهٔ صفت‌های دموگرافیک مختلف مانند سن، جنسیت و شغل برای تولید پیشنهادها ارائه شده است. در همین راستا در مقاله [۱۱] استفاده از مارکوف چندبعدی زمینهٔ تصادفی، قبل از فاکتورگیری ماتریس پیشنهاد شده است. در مارکوف زمینهٔ تصادفی، به صفت‌ها (مانند سن، شغل کاربران، سبک، تاریخ انتشار قلم‌داده‌ها) به‌عنوان محل کار و پروفایل مخفی و متغیرهای تصادفی توجه می‌شود.

در مقالهٔ [۱۲] یک روش فیلتر مشارکتی مبتنی بر تئوری

1. Item
2. Recommender systems
3. Cold start problem

مجموعه‌های ناهموار برای حل مشکل شروع سرد ارائه شده است که امتیازات تهی[4] را در کاندیدها پیش‌بینی می‌کند و نتایج را با استفاده از همسایه های کاربر به دست می‌آورد. در مقالهٔ [۲۶] نیز یک روش جدید با ترکیب فیلترکنندهٔ مشارکتی و مجموعهٔ فازی ناهموار ارائه شده است که در آن، شباهت بر اساس فاصلهٔ اقلیدسی محاسبه می‌شود. در واقع با استفاده از مجموعه‌های فازی ناهموار، از اطلاعات کاربرانی که پیش‌بینی بهتری دارند، استفاده می‌شود. همچنین، از الگوریتم رقابتی استعماری آشوبی، برای تنظیم پارامترهای روش پیشنهادی استفاده شده است.

در مقالهٔ [۱۳] یک استراتژی مقایسهٔ امتیازات جدید برای شکست موانع سرد ارائه شده است که تمرکز اصلی آن بر تهیهٔ یک تنظیم جزئی‌نگر از پروفایل‌های مخفی کاربران شروع سرد است که به کمک کشف تفاوت‌های بین کاربران سرد و گرم انجام می‌گیرد و از تفاوت بین نتایج امتیازات مورد انتظار و نتایج واقعی برای ساخت پروفایل‌های مخفی کاربران سرد استفاده می‌کند.

در مقالهٔ [۱۴] نیز پیشنهاد داده شده است که به کاربران ثبت‌نام نشده در سیستم یک مدل استنتاج طبیعی بر اساس قوانین عدم قطعیت ارائه داده شود که به آن‌ها اجازه می‌دهد تا خودشان پیشنهادهای خودشان را استنتاج کنند. این فرایند به صورت ریاضی به کمک یک مدل احتمالی که استدلال رو به جلو مبتنی بر قوانین را شبیه‌سازی می‌کند رسمی (فرمولیزه) شده است.

در تحقیق [۱۵] پیشنهاد شده یک روش تشویقی به کاربران ارائه شود که آن‌ها را به امتیازدهی قلم‌داده‌ها ترغیب کند. در این پژوهش الگوهای مختلف بازی‌مانند[5] روی کاربران آزمایش می‌شوند تا تأثیرشان بر میانگین امتیازاتی که از کاربران به دست می‌آید مشخص شود.

در مقالهٔ [۱۶]، یک تعریف ریاضی سیستمی از سیستم‌های پیشنهاددهندهٔ فازی، شامل تحلیل‌های تئوری ویژگی‌ها و عملگرهای جبری، مورد بحث قرار گرفته و یک روش فیلتر مشارکتی فازی کاربرمحور ترکیبی[6]، که درجهٔ شباهت نهایی را با مجموع درجهٔ شباهت فازی بین کاربران بر اساس اطلاعات دموگرافیک و درجهٔ شباهت کاربرمحور سخت (که از تاریخچهٔ امتیازات محاسبه می‌شود)، برای رسیدن به دقت بالا پیش‌بینی ارائه کرده و در ادامه در مقالهٔ [۱۷] نیز یک روش ترکیبی که HU-FCF++ نامیده شده است ارائه می‌شود که به‌وسیلهٔ تجمیع چندین روش موجود، برای ترکیب مزایای آن‌ها و حذف معایب آن‌ها، با اضافه کردن چندین رویه، مشکلات روش‌های فعلی مانند تعداد بهینهٔ خوشه‌ها، معیار شباهت، کاربران غیرمرتبط و انتخاب مقدار عضویت را حل می‌کند.

از سوی دیگر در مقالهٔ [۱۸] روش ExcUseMe ارائه شده است، الگوریتمی برای انتخاب کاربران در یک کاوش برخط پیشنهاددهندهٔ فیلتر مشارکتی، که بر تعاملات باینری تکیه می‌کند. هدف ExcUseMe انتخاب کاربرانی است که به نظر می‌رسد به قلم‌داده جدید علاقه‌مندند و بنابراین انتظار می‌رود که بازخورد ایجاد کنند. برای پیدا کردن چنین کاربرانی، ExcUseMe تلاش می‌کند در مشخصه‌های معنادار مورد علاقهٔ کاربران جست‌وجو کند. این عمل با انتخاب کاربرانی با علاقه‌مندی‌های گسسته شروع می‌شود تا زمانی که اولین بازخورد دریافت شود. هنگامی که یک کاربر بازخورد را تهیه کرد، ExcUseMe کاربرانی که به آن کاربر شبیه هستند و انتظار می‌رود روی قلم‌داده جدید بازخورد ایجاد کنند را انتخاب می‌کند.

در مقالهٔ [۱۹] نیز یک روش مبتنی بر فیلتر مشارکتی ارائه شده که با بهره‌گیری از اطلاعات شخصی کاربر در سه مرحله اجرا می‌شود: الف. فیلتر مشارکتی مبتنی بر شخصیت که مستقیماً مدل پیش‌بینی پیشنهادها را به‌وسیلهٔ به همکاری گرفتن اطلاعات شخصی کاربر بهبود می‌بخشد؛ ب. آموزش فعال مبتنی بر شخصیت که از اطلاعات شخصی برای شناسایی داده‌های مفید اضافه در دامنهٔ پیشنهاد هدف که از کاربران استخراج شده است استفاده می‌کند؛ ج. پیشنهاد دامنه‌های پیوندی مبتنی بر شخصیت که از اطلاعات شخصی برای استفاده بهتر از اطلاعات اولویتی کاربر از دامنه‌های کمکی، که می‌تواند برای خنثی کردن کمبود داده‌های اولویتی کاربر در دامنهٔ اصلی استفاده شود، بهره می‌برد. در زاویهٔ دیگری از نگاه به حل مشکل شروع سرد، مقالهٔ [۲۰] یک چارچوب الگوریتمیک بر اساس فاکتورگیری ماتریس ارائه داده که به‌طور همزمان از اطلاعات شباهت بین کاربران و قلم‌داده‌ها برای حل مشکل شروع سرد بهره می‌برد. برخلاف سایر روش‌های موجود، این الگوریتم دو جنبهٔ مشکل شروع سرد را از هم جدا می‌کند: الف. کامل کردن زیرماتریس امتیاز که به‌وسیلهٔ مستثنی کردن کاربران و قلم‌داده های شروع سرد از ماتریس امتیاز اصلی تولید می‌شود؛ ب. هدایت دانش از امتیازات موجود به کاربران و قلم‌داده‌های شروع سرد با استفاده از اطلاعات جانبی.

در مقالهٔ [۲۱] نیز یک روش ترکیبی ارائه شده که پیشنهاددهندهٔ

4. null
5. Gamification
6. HU-FCF :A Hybrid User-based Fuzzy Collaborative Filtering method

مبتنی بر فیلتر مشارکتی را با اطلاعات دموگرافیک ترکیب کرده است. روش بر اساس الگوریتم موجود SCOAL[7] است و یک روش ترکیبی پیشنهاد را ارائه داده که می‌تواند مشکل شروع سرد را در غیاب اطلاعات مشارکتی حل کند.

۳. مفاهیم

در این بخش ابتدا اهداف سیستم‌های پیشنهاددهنده بیان و سپس مدل‌های پایه‌ای این سیستم‌ها بررسی شده‌اند و در نهایت کیفیت فهرست پیشنهادی نیز مطرح شده است.

۱.۳. اهداف سیستم‌های پیشنهاددهنده

یک سیستم پیشنهاددهنده چندین هدف دارد که در ادامه بیان می‌شود [۹].

ارتباط[8]: مشخص‌ترین هدف عملیاتی یک سیستم پیشنهاددهنده، پیشنهاد قلم‌داده‌هایی است که با کاربر مرتبط هستند.

جدید بودن[9]: سیستم پیشنهاددهندهٔ وب وقتی کمک‌کننده است که کاربر قبلاً پیشنهادها را ندیده باشد. تکرار پیشنهاد قلم‌داده‌های مشهور می‌تواند باعث نارضایتی کاربران و به تبع آن کاهش فروش محصولات شود.

خوشبختی[10]: خوشبختی یا نعمت غیرمترقبه، یعنی پیشنهاد قلم‌داده‌ای که کاربر انتظار آن را ندارد، بنابراین برخلاف سایر پیشنهادهای معمولی می‌تواند تصادفی باشد. خوشبختی با جدید بودن متفاوت است؛ زیرا در خوشبختی، پیشنهادها کاربر را غافلگیر می‌کند. خوشبختی این مزیت را دارد که می‌تواند دامنهٔ فروش را افزایش دهد و یا یک گرایش جدید از علاقه‌مندی را در کاربر ایجاد کند. افزایش خوشبختی به‌دلیل کشف محدوده‌های جدیدی از علاقه‌مندی در کاربران معمولا فواید طولانی‌مدت و استراتژیکی برای تجار دارد.

افزایش تنوع پیشنهادات[11]: وقتی فهرست پیشنهادی شامل قلم‌داده‌های متفاوت باشد، این احتمال بیشتر می‌شود که کاربر حداقل یکی از آن‌ها را بپسندد. تنوع باعث می‌شود اطمینان حاصل شود که کاربر از دیدن پیشنهادهای تکراری قلم‌داده‌های شبیه به هم خسته نمی‌شود.

۲.۳. مدل‌های پایه‌ای سیستم‌های پیشنهاددهنده

روش‌هایی که از تعاملات کاربر-قلم‌داده مانند امتیازات یا رفتار خرید استفاده می‌کنند، روش‌های فیلتر مشارکتی[12] نامیده می‌شوند و روش‌هایی که از اطلاعات صفتی دربارهٔ کاربران و قلم‌داده‌ها استفاده می‌کنند، روش‌های مبتنی بر محتوا[13] هستند.

در روش سوم که مبتنی بر دانش[14] نامیده می‌شود، پیشنهادها بر اساس نیازهای آشکارا مشخص‌شدهٔ کاربر است. به جای استفاده از امتیازات قبلی و یا داده‌های خرید، از دانش خارجی و محدودیت‌ها برای ساخت پیشنهادات استفاده می‌شود.

بعضی سیستم‌های پیشنهاددهنده این جنبه‌های مختلف را برای ساخت سیستم‌های ترکیبی با هم ترکیب می‌کنند [۹].

۳.۳. کیفیت

در بسیاری از موارد، قلم‌داده‌های پیشنهادی بسیار به هم شبیه‌اند. همان طور که قبلاً گفته شد، شباهت زیاد قلم‌داده‌های پیشنهادی این ریسک را بیشتر می‌کند که اگر کاربر پربازدیدترین قلم‌داده را نپسندد، هیچ‌یک از قلم‌داده‌های پیشنهادی دیگر را نیز نمی‌پسندد. بنابراین، این حالت احتمال انتخاب‌های درست در فهرست را کاهش می‌دهد که موجب نارضایتی کاربر خواهد شد.

یک راه‌حل برای حل این مسئله این است که وقتی قرار است k قلم‌داده پیشنهاد داده شود تعداد بیشتری انتخاب $b>k$ در نظر گرفته شود. حال از بین آن‌ها به‌طور تصادفی k قلم‌داده انتخاب شوند. این راه‌حل در عمل به‌خوبی نتیجه نخواهد داد. یک راه‌حل بسیار مؤثرتر این است که به‌طور حریصانه k قلم‌داده از بین b قلم‌داده برتر انتخاب شود. در این روش با b مورد شبیه به هدف شروع کرده و به‌طور افزایشی یک مجموعهٔ متنوع ساخته می‌شود [۹].

۴. روش پیشنهادی

در این مقاله، ابتدا طبق روش‌های مبتنی بر دانش، سعی می‌شود از کاربر جدید، برای شناسایی علاقه‌مندی‌هایش کمک گرفته شود. برای این منظور ابتدا یک فهرست پیشنهادی اولیه، که سعی شده با در نظر گرفتن معیارهای یک پیشنهاد خوب، طوری تهیه شود که بتواند دامنهٔ گسترده‌تری از علایق کاربر را مشخص کند، به وی ارائه می‌شود. همان‌طور که قبلاً اشاره شد معیارهای مختلفی برای پیشنهادها وجود

7. Simultaneous Co-Clustering and Learning
8. Relevance
9. Novelty
10. Serendipity
11. Increasing recommendation diversity
12. collaborative filtering
13. content-based
14. knowledge-based

دارند که از مهم‌ترین آن‌ها می‌توان ارتباط، تنوع و خوشبختی را نام برد. پس از اینکه بازخوردهای کاربر نسبت به فهرست پیشنهادی اولیه دریافت شد، به کمک روش‌های مبتنی بر محتوا (بر اساس اطلاعات دموگرافیک کاربر) و فیلتر مشارکتی (بر اساس اطلاعاتی که از امتیازات داریم)، سعی می‌شود علایق و سلایق کاربر استخراج و از آن‌ها در تهیهٔ فهرست‌های بعدی استفاده شود. تا زمانی که کاربر از حالت شروع سرد خارج شود، همچنان از کاربر کمک می‌گیریم و پس از آن، سیستم به حالت عادی بازمی‌گردد.

۱.۴. آماده‌سازی داده‌ها

برای بهبود روش پیشنهادی، ابتدا کاربران بر اساس اطلاعات دموگرافیکشان به روش K-means خوشه‌بندی می‌شوند و تعداد خوشه‌ها به‌صورت خودکار بهینه می‌شود. برای تعیین و ارزیابی بهترین تعداد خوشه‌ها از تابع Silhouette Coefficient(SC) [25] استفاده شده است. برای مجموعه داده‌های ما تعداد بهینهٔ خوشه‌ها دو خوشه تعیین شد. پس از ورود کاربر جدید، با توجه به اطلاعات دموگرافیک کاربر، خوشه‌ای که بیشترین درجهٔ شباهت را به آن دارد تعیین شده و سایر محاسبات بر اساس آن انجام می‌شود. سپس ماتریس شباهت قلم‌داده‌ها نیز تهیه می‌شود. نحوهٔ محاسبهٔ شباهت میان قلم‌داده‌ها در این مقاله که در آن از اطلاعات غیر عددی استفاده می‌شود بدین صورت است که به‌ازای یکسان بودن مقدار دو صفت، تابع شباهت آن صفت عدد یک و در صورت یکسان نبودن، عدد صفر را برمی‌گرداند. این مقادیر شباهت کلی دو قلم‌داده را تولید می‌کنند. روند تعیین شباهت دو قلم‌داده *A(a1,a2...,ak)* و *B(b1,b2...,bk)* شامل k ویژگی در ادامه آورده شده است.

```
Sim-Function
Input: (A,B,k)
sim=0;
For i:=1 to k do
      If  ai=bi then
          sim=sim+1;
similarity= sim/k;
return similarity;
```

سپس از طریق محاسبه روی داده‌های کاربران-قلم‌داده‌ها، میانگین امتیازات هر قلم‌داده محاسبه می‌شود. این میانگین‌ها در ستون جدیدی که در مجموعه داده‌های قلم‌داده‌ها اضافه می‌شود ذخیره می‌شوند. سپس همین عمل به‌ازای قلم‌داده‌های درون هر خوشه تکرار می‌شود و در ستون‌های جداگانه به داده‌ها اضافه می‌شوند.

۲.۴. تعیین فهرست پیشنهاد اولیه

با ورود کاربر جدید به سیستم، ابتدا خوشه‌ای که به آن تعلق دارد مشخص می‌شود. سپس یک فهرست ده‌تایی از قلم‌داده‌ها برای امتیازدهی به وی ارائه می‌شود. طول این فهرست به کمک الگوریتم بهینه‌سازی ازدحام ذرات آشوبی برای ما به دست آمده است و در هر پژوهش با توجه به شرایط موجود به‌صورت خودکار بهینه خواهد شد. برای پر کردن این فهرست اولیه، دو معیار ارتباط و تنوع از اهمیت زیادی برخوردارند.

نحوهٔ انتخاب قلم‌داده‌های فهرست پیشنهاد اولیه بدین صورت است که ابتدا قلم‌داده‌ها بر اساس امتیازات بالایی که کسب کرده‌اند در فهرست پیشنهادی ۱۰ قلم‌داده‌ای قرار می‌گیرند. به همین منظور ابتدا ۴ قلم‌دادهٔ برتر از نظر امتیاز کلی را در فهرست قرار می‌دهیم. (این عدد نیز به کمک الگوریتم ازدحام ذرات آشوبی بهینه شده است). از آنجایی که یکی از معیارهای لازم برای فهرست پیشنهادی در سیستم پیشنهاددهنده، متنوع بودن پیشنهادهاست، برای تهیهٔ یک فهرست متنوع از انواع قلم‌داده‌ها لازم است تا علاوه بر در نظر گرفتن میزان محبوبیت آن‌ها در بین سایر کاربران، به میزان تفاوت آن‌ها نیز توجه کنیم. برای نیل به این منظور، در ادامه از روی ستون امتیازهای قلم‌داده‌های درون خوشهٔ کاربر جدید، که در مرحلهٔ قبل تهیه کردیم، اولین قلم‌داده‌ای که بیشترین امتیاز را دارد درون فهرست قرار می‌دهیم، سپس یکی یکی قلم‌داده‌های امتیازبالا (کلاس محبوب‌تر) را برداشته، کیفیت آن‌ها را نسبت به قلم‌داده‌های درون فهرست محاسبه می‌کنیم و قلم‌داده‌ای که بیشترین کیفیت را دارا باشد به فهرست اضافه می‌شود. این کار را تا جایی ادامه می‌دهیم که فهرست پیشنهادی پر شود. همان طور که پیش‌تر نیز اشاره شد، دلیل استفاده از کیفیت این است که در بین قلم‌داده‌های کلاس محبوب، تفاوت بیشتر قلم‌داده‌ها با هم از تفاوت اندک امتیازاتشان مهم‌تر است؛ زیرا ارائهٔ قلم‌داده‌های متفاوت به کاربر، دامنهٔ اطلاعات ما از سلایق وی را گسترش می‌دهد و کمک می‌کند که کاربر زودتر از حالت شروع سرد خارج شود. همچنین ریسک نامطلوب بودن تمامی فهرست پیشنهادی از نظر کاربر را کاهش می‌دهد. بنابراین در این مرحله از معیار کیفیت برای ایجاد تعادل بین امتیاز بالای قلم‌داده‌ها و تفاوت قلم‌داده‌های درون فهرست استفاده می‌کنیم.

در ادامهٔ کار، این فهرست به کاربر ارائه و از وی درخواست می‌شود که به قلم‌داده‌ها امتیازدهی کند. بهتر است این درخواست به طرقی که برای کاربر جذاب‌تر است و یا روش‌های بازی‌مانند انجام

شود تا وی به امتیازدهی بیشتر تشویق شود [۱۵]. با دریافت امتیازات کاربر، اولین اطلاعات از علایق وی به دست می‌آید.

۳.۴. تعیین فهرست پیشنهاد ثانویه

حال با استفاده از این اطلاعات فهرست پیشنهادی بعدی تشکیل می‌شود. به‌ازای هر قلم‌داده‌ای که کاربر به آن امتیاز بالا داده است (کلاس محبوب) ۱۰ قلم‌داده از شبیه‌ترین قلم‌داده‌ها را از طریق ماتریس شباهت مشخص می‌کنیم. سپس در بین این ۱۰ قلم‌داده، شباهت را با دخیل کردن امتیاز مجددا محاسبه می‌کنیم؛ یعنی با توجه به امتیازی که کاربر به قلم‌داده مورد نظر در فهرست پیشنهاد قبلی داده و میانگین امتیازی که آن قلم‌داده دارد، شباهت را مجدداً این بار با اضافه کردن معیار امتیاز محاسبه می‌کنیم و در محاسبات بعدی برای پرکردن فهرست پیشنهادی از آن استفاده می‌کنیم. برای تمام قلم‌داده‌هایی که کاربر به آن‌ها امتیاز محبوب داده است، این عمل را تکرار می‌کنیم.

از سوی دیگر، اگر کاربر به قلم‌داده‌هایی امتیاز ضعیف داده بود از طریق ماتریس شباهت به‌ازای هر قلم‌داده، ۱۰ قلم‌داده‌ای را که فاصلهٔ بیشتری با آن دارد انتخاب می‌کنیم. در نهایت از بین قلم‌داده‌های انتخاب‌شده از فهرست محبوب‌ها ۸۰٪ فهرست پیشنهادی بعدی را پر کرده و ۲۰٪ بقیه را از فهرست قلم‌داده‌هایی که با محاسبهٔ فاصله به دست آوردیم تکمیل می‌کنیم. نحوهٔ تکمیل بر اساس تأثیر دوجانبهٔ شباهت (شباهت با دخیل کردن امتیاز) به قلم‌دادهٔ مربوطه در فهرست قبلی و تفاوت نسبت به قلم‌داده‌های فهرست فعلی می‌باشد. به همین دلیل از بین قلم‌داده‌هایی که در مرحلهٔ قبل به دست آمد ابتدا قلم‌داده‌ای که بیشترین شباهت را داشت، در فهرست قرار داده و مانند قبل به کمک فرمول کیفیت بقیهٔ فهرست پر می‌شود. شایان ذکر است در صورتی که در هریک از کلاس‌های محبوبیت تعداد قلم‌داده‌های امتیاز داده شده کمتر از حد مورد نیاز باشد می‌توان این کمبود را از کلاس دیگر جبران کرد.

پس از تکمیل فهرست پیشنهاد بعدی، آن را به کاربر ارائه داده و امتیازات وی را دریافت می‌کنیم.

۴.۴. بررسی وضعیت خروج کاربر از حالت شروع سرد

حال در این مرحله باید بررسی شود که آیا کاربر از حالت شروع سرد خارج می‌شود یا نه. همان طور که مشاهده شد، ما در مرحلهٔ قبل به کمک امتیازاتی که در فهرست پیشنهادی اولیه از کاربر دریافت کرده بودیم، قلم‌داده‌های مورد علاقهٔ کاربر را شناسایی کردیم. ازاین‌رو امتیازاتی که این قلم‌داده‌ها دارند همان مقدار پیش‌بینی ما می‌باشند. با مقایسهٔ این مقادیر با امتیازاتی که از کاربر دریافت می‌شوند، می‌توان میزان خطا را محاسبه کرد. برای این کار کافی است اختلاف امتیاز قلم‌داده‌های فهرست پیشنهادی ثانویه را با امتیازاتی که کاربر به آن‌ها داده محاسبه کرده و میانگین آن را حساب کنیم. در صورتی که این میانگین از حد آستانهٔ اختلاف (که در هر کاربرد مقدار متفاوتی خواهد داشت) کمتر باشد کاربر از حالت شروع سرد خارج شده و این بدین معنی است که از این پس سیستم بدون احتیاج به استفاده از روش‌های مبتنی بر دانش و مشارکت مستقیم کاربر می‌تواند برای وی فهرست پیشنهادی تولید کند. مراحل انجام روش پیشنهادی در شکل (۱) بیان شده است.

پس از پیاده‌سازی روش پیشنهادی، پارامترهای طول فهرست پیشنهادی و تعداد قلم‌داده‌های انتخابی از محبوب‌ترین قلم‌داده‌ها به کمک الگوریتم ازدحام ذرات آشوبی بهینه شدند.

به همین منظور از تابع آشوبی Logistic [۲۴] برای اعمال آشوب در الگوریتم بهینه‌سازی ازدحام ذرات استفاده شد. طبق این روش در هر بار تکرار الگوریتم، سرعت ذرهٔ جدید به کمک فرمول (۱) به دست می‌آید:

$$V_{i,j}(t+1) = WV_{i,j}(t) + c_1 CM_{1,j}(t)\left[pbest_{i,j}(t) - X_{i,j}(t)\right] + c_2 CM_{2,j}(t)\left[gbest_i(t) - X_{i,j}(t)\right] \quad (۱)$$

که در آن، CM_1 و CM_2 طبق فرمول (۲) تابع Logistic است:

$$X_{n+1} = 0.5X_n(1-X_n) \quad (۲)$$

ذرهٔ اولیه شامل طول فهرست پیشنهادی ($x1$) و تعداد قلم‌داده انتخابی با معیار محبوبیت ($x2$) بوده و تابع هزینه، به دنبال کاهش میزان خطای RMSE و MAE است.

X: | X1 | X2 |

در ادامه، ساختار یک ذره آورده شده است.

Particle
particle.Position
particle.Cost
particle.Velocity
particle.Best.Position
particle.Best.Cost

روش پیشنهادی با طول فهرست و تعداد قلم‌داده انتخابی با معیار محبوبیتی که توسط الگوریتم ازدحام ذرات آشوبی تولید شدند، اجرا شده و در هر تکرار خطای آن به دست آمد. برای تعیین این اعداد، متغیرهای تولیدی الگوریتم ازدحام ذرات آشوبی به کمک تابع جزء صحیح تبدیل به اعداد صحیح شده و به‌صورت پارامتر به برنامهٔ روش

پیشنهادی داده شده است. در ادامه، قسمتی از روند اجرایی تابع هزینه، الگوریتم ازدحام ذرات آورده شده است.

Cost-Function
Input: (*x, P, M*)
*K=min (floor (1+M*x), M);*
K=max (K, P);
[MAE RMSE]=Proposed model (K);
Output: (*MAE error, RMSE error*)

که در آن M بیشترین مقدار طول فهرست و P کمترین مقدار آن است که در هر کاربرد متفاوت است و x نیز توسط الگوریتم ازدحام ذرات آشوبی تولید شده است.

بخش‌های دیگری از روند اجرایی برنامه نیز شامل به‌روزرسانی مکان ذرات و تعیین بهترین حالت محلی و بهترین حالت مطلق در ادامه آورده شده است.

Update Position
particle (i).Position = particle (i).Position + particle (i).Velocity;
Evaluation
particle (i).Cost = Cost Function (particle (i).Position);
Update Personal Best
If particle (i).Cost<particle (i).Best.Cost
particle (i).Best.Position=particle (i).Position;
particle (i).Best.Cost=particle (i).Cost;
Update Global Best
If particle (i).Best.Cost<GlobalBest.Cost
Global Best=particle (i).Best;

۵. آزمایشات

در این بخش، ابتدا توصیف مختصری از مجموعه داده استفاده شده آورده شده است و سپس به ارزیابی مدل پرداخته می‌شود.

۱.۵. مجموعهٔ داده

برای اعتبارسنجی روش پیشنهادی از مجموعه دادهٔ 1M و 100K MovieLens [۲۲] استفاده شده است. این مجموعه مربوط به اطلاعات و امتیازدهی کاربران به فیلم است که از سایت IMDB استخراج شده و در ادامه مشخصات این مجموعه آورده شده است. این مجموعه داده شامل امتیازاتی است که کاربران به فیلم‌ها داده‌اند (جدول۱) و هر کاربر حداقل ۲۰ فیلم را امتیازدهی کرده است. کاربران و فیلم‌ها به‌ترتیب از ۱ شماره‌گذاری شده‌اند و داده‌های امتیازدهی به‌صورت تصادفی پراکنده شده‌اند. اطلاعات دموگرافیک هر کاربر شامل شمارهٔ شناسایی، سن، جنسیت، شغل و زیپ کد است (جدول ۳). اطلاعات هر قلم‌داده نیز شامل شمارهٔ شناسایی، عنوان فیلم، تاریخ تولید و تاریخ انتشار و سبک فیلم شامل ۱۹ سبک مختلف است (جدول ۲).

جدول (۱): اطلاعات امتیازات کاربران به فیلم‌ها

user id	item id	rating

جدول (۲): اطلاعات فیلم‌ها

movie id	movie title	release date	video release date	IMDB URL
unknown	Action	Adventure	Animation	Children's
Comedy	Crime	Documentary	Drama	Fantasy
Film-Noir	Horror	musical	Mystery	Romance
Sci-Fi	Thriller	War	Western	

جدول (۳): اطلاعات کاربران

user id	age	gender	occupation	zip code

۲.۵. معیار ارزیابی

محیط استفاده‌شده برای پیاده‌سازی روش نرم‌افزار متلب بوده و برای ارزیابی میزان دقت روش از دو معیار MAE و RMSE استفاده شده است. این دو معیار از طریق فرمول‌های (۳) و (۴) به دست می‌آیند:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(X_{obs,i} - X_{model,i}) \quad (۳)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \quad (۴)$$

که در آن، $X_{obs,i}$ امتیاز واقعی کاربر و $X_{model,i}$ امتیاز پیش‌بینی‌شدهٔ روش پیشنهادی است.

۳.۵. ارزیابی

برای محاسبهٔ میزان خطای مدل، امتیازات واقعی که قلم‌داده‌های پیشنهادی فهرست ثانویه دارند همان مقدار پیش‌بینی در نظر گرفته شده‌اند. با مقایسهٔ این مقادیر با امتیازاتی که از کاربر دریافت می‌شوند، میزان خطا محاسبه می‌شود. برای این کار کافی است اختلاف امتیاز قلم‌داده‌های فهرست پیشنهادی ثانویه و امتیازاتی که کاربر به آن‌ها داده، محاسبه شود. در هر بار اجرای برنامه، میزان خطاها به همراه متغیرهای ورودی برنامه به‌عنوان خروجی نمایش داده می‌شوند که بهترین حالت طول فهرست ۱۰ و تعداد قلم داده انتخابی ۴ با میزان خطای MAE و RMSE به‌ترتیب ۰/۵۸ و ۰/۶۱ به دست آمد.

در جدول (۴) تعدادی از تکرارهای روش، با مقادیر مختلف متغیرهای طول فهرست پیشنهادی و تعداد قلم‌دادهٔ انتخابی از قلم‌داده‌های محبوب، به همراه خطای حاصل از هر تکرار آمده است. این جدول به‌ترتیب خطای MAE به‌صورت صعودی مرتب شده است.

جدول (۴): نمونه مقادیر متغیرها در تکرارها

طول فهرست پیشنهادی	تعداد قلم‌داده‌های انتخابی از محبوب‌ها	خطای MAE	خطای RMSE
۱۰	۴	۰/۵۸	۰/۶۱
۱۰	۳	۰/۵۹	۰/۶۱
۱۰	۶	۰/۶۱	۰/۶۴
۱۰	۲	۰/۶۳	۰/۶۵
۲۲	۵	۰/۶۳	۰/۶۶
۲۶	۶	۰/۶۴	۰/۶۷
۲۸	۷	۰/۶۴	۰/۶۷
۲۰	۴	۰/۶۵	۰/۶۷
۳۰	۱۵	۰/۶۵	۰/۶۷
۲۰	۳	۰/۶۶	۰/۶۸
۲۰	۱۰	۰/۶۶	۰/۶۸
۳۰	۱۰	۰/۶۶	۰/۶۸
۲۰	۱۵	۰/۶۶	۰/۶۸
۲۰	۱۲	۰/۶۷	۰/۶۹
۲۸	۲۴	۰/۶۷	۰/۶۹
۳۰	۴	۰/۶۸	۰/۷۰
۳۰	۹	۰/۶۸	۰/۷۰

برای تعیین میزان داده‌های آموزش و آزمون، به‌ترتیب ۱۰، ۲۰ و ۳۰درصد داده‌ها به‌عنوان آزمون و بقیهٔ داده‌ها به‌عنوان آموزش در نظر گرفته شده‌اند که نتایج آن در جدول (۵) آمده است.

جدول (۵): ارزیابی روش پیشنهادی با درصد داده‌های آزمون مختلف

درصد داده‌های آزمون	خطای MAE	خطای RMSE
۱۰	۰/۵۹	۰/۶۱
۲۰	۰/۵۹	۰/۶۱
۳۰	۰/۵۸	۰/۶۱

با توجه به نتیجهٔ آزمایشات، برای ارزیابی نهایی روش، ۳۰درصد داده‌ها به‌صورت تصادفی به‌عنوان آزمون در نظر گرفته شده‌اند. جدول (۵) نشان می‌دهد اندازهٔ داده‌های آموزش و آزمون بر روی روش پیشنهادی تأثیر چندانی ندارد.

نتایج به‌دست‌آمده از ارزیابی روش پیشنهادی با نتایج چندین روش موجود مورد مقایسه قرار گرفتند. همان طور که نتایج این مقایسه در جدول (۶) مشاهده می‌شود، روش پیشنهادی توانسته خطای پیش‌بینی را تا حد مطلوبی کاهش دهد. این کاهش در مجموعه داده ۱M برای خطای RMSE حدود ۰/۳ و برای خطای MAE حدود ۰/۲ و در مجموعه داده ۱۰۰K حدود ۰/۰۲ برای خطای MAE بوده است.

جدول (۶): مقایسهٔ روش پیشنهادی با روش‌های موجود

مجموعه داده	روش	مقدار خطای MAE	مقدار خطای RMSE
۱M	روش Hernando et al [۱۴]	۰/۸	۰/۸۶۷
	روش RAPARE-MF [۱۳]	–	۰/۹
	روش پیشنهادی	۰/۵۸	۰/۶۱
100K	روش CME/DC [۲۱]	۰/۱۹	–
	روش پیشنهادی	۰/۱۷	۰/۶۲

نمودارهای مقایسهٔ میزان خطا در روش‌های مختلف در شکل‌های (۲) و (۳) نشان داده شده است.

همچنین برای اندازه‌گیری مقدار Precision، Recall و F-Measure امتیازات به دو کلاس محبوب (امتیازات ۴ و ۵) و غیرمحبوب (امتیازات ۳ و ۲ و ۱) طبقه‌بندی شدند. مقادیر این معیارها در جدول (۷) آمده است.

جدول (۷): مقادیر معیارهای Precision، Recall و F-Measure

	Recall	Precision	F-Measure
کلاس محبوب	۰/۹۱	۰/۴۲	۰/۶۹
کلاس غیرمحبوب	۰/۵۴	۰/۹۴	۰/۵۷

۴.۵. جمع‌بندی

همان طور که بیان شد، در این روش طول فهرست پیشنهادی و تعداد قلم داده‌هایی که به‌طور مستقیم از فهرست محبوب‌ها انتخاب می‌شوند توسط الگوریتم ازدحام ذرات آشوبی تعیین می‌شود که تعدادی از تکرارهای این الگوریتم به همراه نتایج خطای هر بار تکرار در جدول (۴) مشاهده می‌شود. طبق این بهینه‌سازی برای مجموعه‌داده فعلی در نهایت فهرست ده‌تایی و قلم داده چهارتایی مورد استفاده قرار گرفته است. این تعداد برای مجموعه‌داده‌های مختلف متفاوت خواهد بود. از سوی دیگر، برای تعیین بهترین میزان داده‌های آموزش و آزمون سه تکرار از برنامه اجرا شد که در نهایت با توجه به تفاوت ناچیز در نتایج، میزان ۳۰درصد دادهٔ آزمون در نظر گرفته شد (جدول ۵).

بررسی میزان خطا در مقایسه با برخی روش‌های موجود نشان‌دهندهٔ بهبود کارایی سیستم پیشنهاددهنده در روش پیشنهادی است. همچنین از مقایسهٔ نتایج به‌دست‌آمده از اجرای روش روی دو مجموعه دادهٔ مختلف (1M و 100K) مشخص می‌شود با وجود اینکه روش پیشنهادی به میزان داده‌های آموزش و آزمون بستگی چندانی ندارد، اما کیفیت و کمیت متفاوت مجموعه‌داده در نتیجه تأثیر محسوسی دارد (جدول ۶).

۶. نتیجه‌گیری و کارهای آتی

در این مقاله سعی شد سیستم پیشنهاددهنده‌ای طراحی شود که تا حد امکان، سریع‌تر مشکل شروع سرد را حل کند. روش پیشنهادی برای غلبه به این مشکل دارای مزایایی است که در ادامه به چند مورد از آن‌ها اشاره می‌شود.

سیستم‌های پیشنهاددهنده عمدتاً از سه روش به حل مشکل شروع سرد می‌پردازند که در این مقاله از ترکیب هر سه مدل مبتنی بر محتوا، مبتنی بر دانش و فیلتر مشارکتی استفاده شده است، بنابراین از مزایای هر سه مدل بهره برده شده است.

یکی از مواردی که در پژوهش‌های پیشین به آن توجهی نشده معیار خوشبختی است. توجه به این معیار می‌تواند علاوه بر سریع‌تر کردن روند گسترش دامنهٔ اطلاعات سیستم از سلایق کاربر (با توجه به جهش‌های نسبتاً تصادفی به قسمت‌های مختلف قلم‌داده‌ها)، با توجه به بالا بردن جذابیت سیستم، سطح همکاری کاربر را نیز بالا ببرد. برای در نظر گرفتن معیار خوشبختی، با پیشنهاد قلم‌داده‌های غیر شبیه به قلم‌داده‌هایی که کاربر دوست ندارد و امتیاز کمی به آن‌ها داده، احتمال غافلگیر شدن کاربر بیشتر از زمانی است که به کاربر فقط قلم‌داده‌های شبیه به قلم‌دادهٔ مورد علاقه‌اش پیشنهاد می‌شود. برای کاهش ریسک پیشنهادهای غیر مرتبط، تنها ۲۰٪ پیشنهادها به این روش اختصاص یافته است.

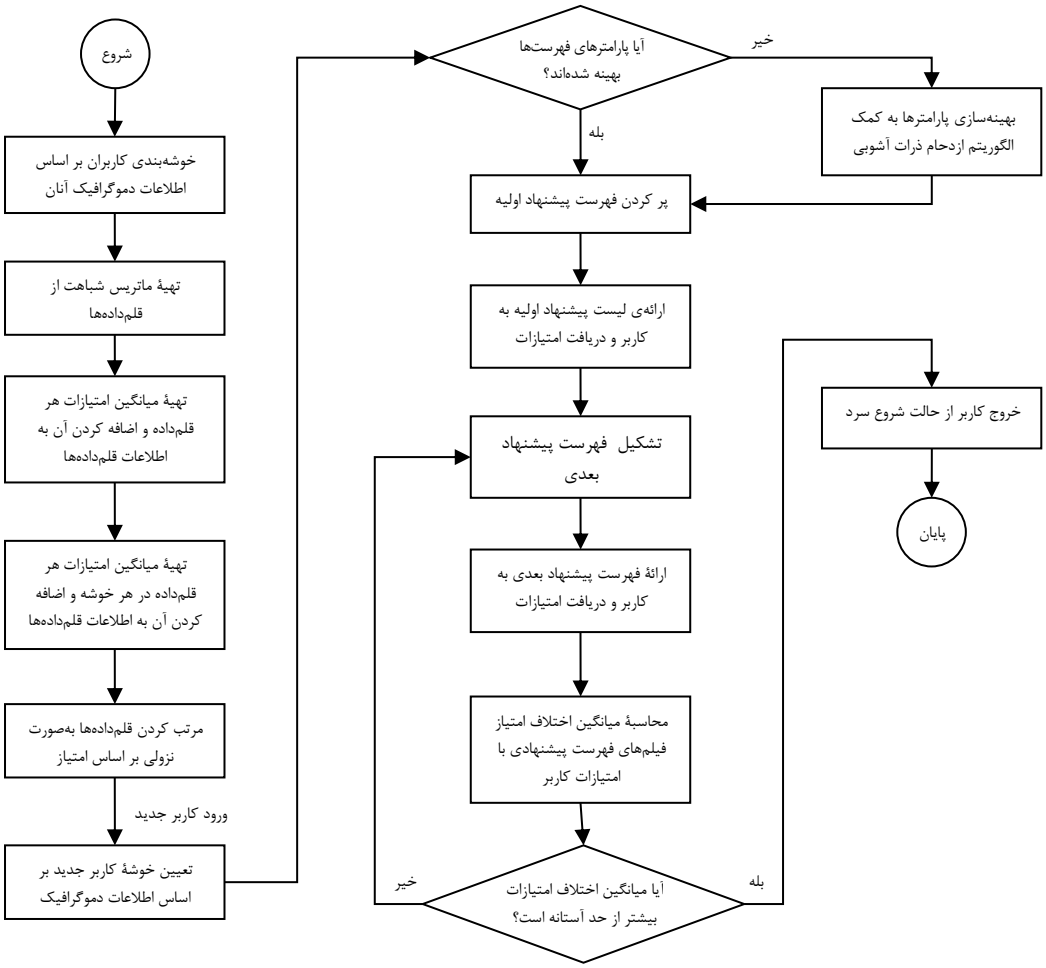
در این روش، هنگام پر کردن فهرست پیشنهادی، شباهت همزمان با در نظر گرفتن معیار امتیاز و اطلاعات محتوایی قلم‌داده‌ها محاسبه می‌شود. بدین منظور علاوه بر اینکه شباهت‌های قلم‌داده‌ها در نظر گرفته می‌شوند (روش مبتنی بر محتوا)، میانگین امتیازات قلم‌داده‌ها نیز در شباهت ثانویه دخیل می‌شود (روش مبتنی بر فیلتر مشارکتی). بنابراین همزمان از تمامی اطلاعاتی که از قلم‌داده‌های سیستم وجود دارد استفاده می‌شود.

در این روش معیار مهم دیگری که برای پر کردن فهرست پیشنهادی در نظر گرفته می‌شود، تنوع است. در روش‌های قبلی، فهرست پیشنهادی تنها بر اساس محبوبیت بالا و یا شباهت قلم‌داده‌ها به قلم‌داده‌های مورد علاقهٔ کاربر انتخاب می‌شدند. این روش‌های انتخاب باعث بروز مشکلاتی می‌شوند که از جملهٔ آن‌ها می‌توان به بالا رفتن ریسک نامطلوب بودن تمامی فهرست پیشنهادی برای کاربر نام برد. همچنین دامنهٔ اطلاعات سیستم از سلایق کاربر بسیار محدود شده و سرعت کسب اطلاعات بیشتر بسیار پایین می‌آید. در روش پیشنهادی این مقاله، برای حل این مشکل در تهیهٔ فهرست پیشنهادی اولیه، از آنجایی که هنوز هیچ اطلاعاتی از سلایق کاربر جدید موجود نیست نمی‌توان از روشی که برای معیار خوشبختی بیان شد استفاده کرد. بنابراین همان طور که بیان شد با در نظر گرفتن معیار تنوع، سعی شده است این مشکل تا حد زیادی رفع شود. نحوهٔ انجام آن نیز به این صورت است که پس از آنکه قسمتی از فهرست به کمک قلم‌داده‌های محبوب پر شد، برای گزینش سایر قلم‌داده‌ها به کمک معیار کیفیت بررسی می‌شود که تنها قلم‌داده‌هایی که میزان تفاوت تعیین‌شده‌ای با قلم‌داده‌های درون فهرست را دارند بتوانند گزینش شوند. این معیار تعادلی بین میزان محبوبیت قلم‌داده‌ها و میزان تفاوت قلم‌داده‌های درون فهرست با هم برقرار می‌کند که از شبیه بودن تمامی فهرست جلوگیری می‌کند.

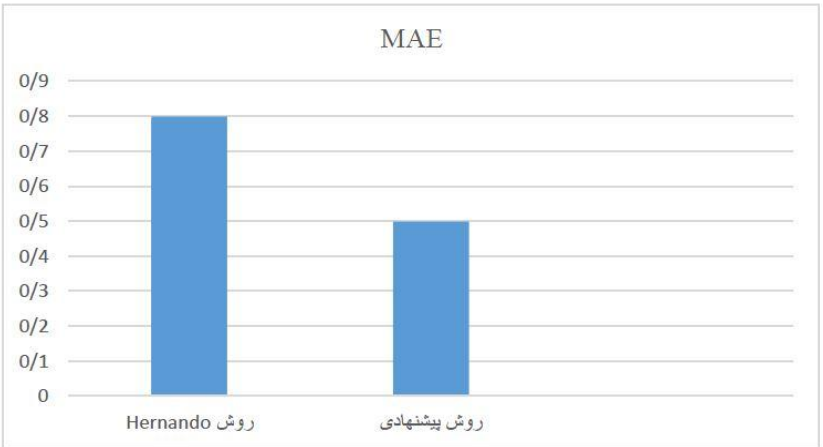
از آنجایی که مراحل اولیه و زمان‌بر این روش شامل خوشه‌بندی کاربران و تشکیل ماتریس شباهت تنها یک بار و در ابتدای کار انجام می‌گیرد، با افزایش حجم مجموعه‌داده، مشکلی برای انجام مراحل به وجود نمی‌آید و تنها یک بار محاسبات زمان‌بر انجام می‌شود و سایر مراحل از سرعت کافی برخوردار خواهند بود. بنابراین هنگامی که کاربر جدید وارد سیستم می‌شود، تمامی محاسبات زمان‌بر قبلاً انجام شده و کاربر در زمان دریافت پیشنهادات تأخیر زمانی را احساس نخواهد کرد که این خود از مزایای مهم یک سیستم پیشنهاددهنده است. همچنین در صورتی که به‌دلیل دریافت اطلاعات جدید مجموعه داده به مرور گسترش یابد، می‌توان به‌صورت دوره‌ای خوشه‌بندی و تشکیل ماتریس شباهت را انجام داد، تا میزان زمان کمتری صرف به‌روزرسانی سیستم شود. بنا به دلایلی که گفته شد، این سیستم به‌راحتی قابلیت گسترش برای استفاده‌های بلادرنگ و برخط را دارد؛ زیرا می‌توان مراحل زمان‌بر را به‌صورت برون‌خط انجام داد تا در هنگام ورود کاربر جدید تنها تشکیل فهرست‌های پیشنهادی به‌صورت برخط انجام شوند که زمان زیادی را به خود اختصاص نخواهند داد.

همچنین استفاده از فرمول‌های دقیق‌تر شباهت برای تشکیل ماتریس شباهت قلم‌داده‌ها و استفاده از مجموعه داده‌های کمکی برای بهبود و تکمیل این ماتریس برای تولید پیشنهادهای بهتر (از نظر کاربر) از دیگر گزینه‌های پیش رو برای کارهای آینده است. بهبود سایر پارامترهای سیستم فعلی در پژوهش‌های آتی نیز می‌تواند میزان خطا را تا حد مطلوبی بهبود بخشد.

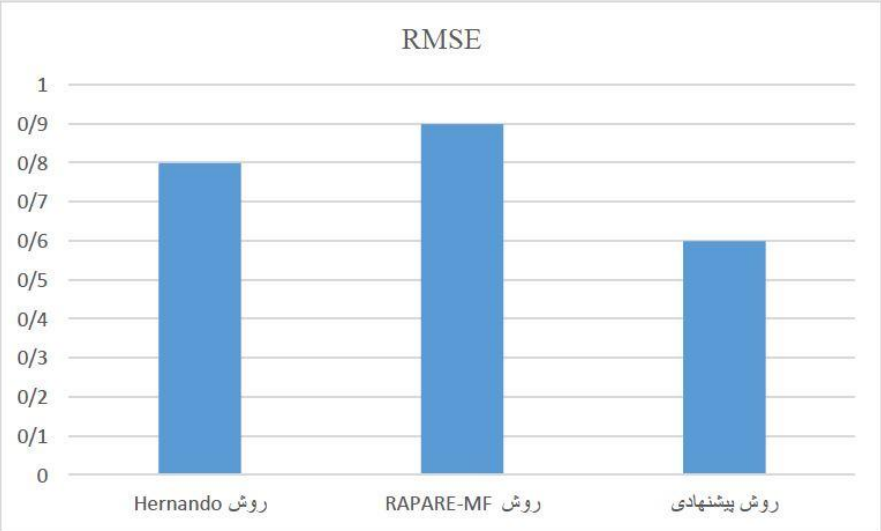
با جایگزینی مجموعه داده‌های مختلف، استفاده از این روش برای سایر سیستم‌های پیشنهاددهنده نیز به‌راحتی امکان‌پذیر است. همچنین می‌توان با اضافه کردن بخش‌هایی از روش پیشنهادی (مثلاً در نظر گرفتن معیارهای خوشبختی و تنوع و... در قسمت تولید فهرست) به سایر روش‌های موجود کارایی آن‌ها را افزایش داد.



شکل (۱): فلوچارت روش پیشنهادی



شکل (۲): مقایسهٔ خطای RMSE روش‌های موجود با روش پیشنهادی



شکل (۳): مقایسهٔ خطای MAE روش‌های موجود با روش پیشنهادی

مراجع

[1] Núñez-Valdéz E. R., Lovelle J. M. C., Martínez O. S., García-Díaz V., de Pablos P. O., and Marín C. E. M., "*Implicit feedback techniques on recommender systems applied to electronic books*", Computers in Human Behavior, Vol. 28, No. 4, pp. 1186-1193, 2012.

[2] Roy S. and Guntuku S. C., "*Latent Factor Representations for Cold-Start Video Recommendation*", In Proceedings of the 10th ACM Conference on Recommender Systems, pp. 99-106, ACM, 2016.

[3] Barragáns-Martínez A. B., Costa-Montenegro E., Burguillo J. C., Rey-López M., Mikic-Fonte F. A., and Peleteiro A., "*A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition*", Information Sciences, Vol. 180, No. 22, pp. 4290-4311, 2010.

[4] Lee S. K., Cho Y. H., and Kim S. H., "*Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations*", Information Sciences, Vol. 180, No. 11, pp. 2142-2155, 2010.

[5] Bobadilla J., Serradilla F., and Hernando A., "*Collaborative filtering adapted to recommender systems of e-learning*", Knowledge-Based Systems, Vol. 22, No. 4, pp. 261-265, 2009

[6] Castro-Schez J. J., Miguel R., Vallejo D., and López-López L. M., "*A highly adaptive recommender system based on fuzzy logic for B2C e-commerce portals*", Expert Systems with Applications, Vol. 38, No. 3, pp. 2441-2454, 2011.

[7] Nguyen T. T. S., Lu H. Y., and Lu J., "*Web-page recommendation based on web usage and domain knowledge*", IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 10, pp. 2574-2587, 2014.

[8] Son L. H., "*Dealing with the new user cold-start problem in recommender systems: A comparative review*", Information Systems, Vol. 58, pp. 87-104, 2016.

[9] Aggarwal C. C., *Recommender Systems*, Springer, 2016.

[10] Safoury L. and Salah A., "*Exploiting user demographic attributes for solving cold-start problem in recommender system*", Lecture Notes on Software Engineering, Vol. 1, No. 3, pp. 303, 2013.

[11] Peng F., Lu J., Wang Y., Yi-Da Xu R., Ma C., and Yang J., "*N-dimensional Markov random field prior for cold-start recommendation*" , Neurocomputing, Vol. 191, pp. 187-199, 2016.

[12] Huang C. B. and Gong S. J., "*Employing rough set theory to alleviate the sparsity issue in recommender system*", International Conference on Machine Learning and Cybernetics, Vol. 3, pp. 1610-1614, IEEE, 2008.

[13] Xu J., Yao Y., Tong H., Tao X., and Lu J., "*Ice-breaking: Mitigating cold-start recommendation problem by rating comparison*", In Proceedings of the 24th International Conference on Artificial Intelligence, pp. 3981-3987, AAAI Press, 2015.

[14] Hernando A., Bobadilla J., Ortega F., and Gutiérrez A., "*A probabilistic model for recommending to new cold-start non-registered users*", Information Sciences, Vol. 376, pp. 216-232, 2017.

[15] Feil S., Kretzer M., Werder K., and Maedche A., "*Using Gamification to Tackle the Cold-Start*

Problem in Recommender Systems", In Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, pp. 253-256, ACM, 2016.

[16] Son L. H., "*HU-FCF: a hybrid user-based fuzzy collaborative filtering method in recommender systems*", Expert Systems with Applications: An International Journal, Vol. 41, No. 15, pp. 6861-6870, 2014.

[17] Son L. H., "*HU-FCF++: A novel hybrid method for the new user cold-start problem in recommender systems*", Engineering Applications of Artificial Intelligence, Vol. 41, pp. 207-222, 2015.

[18] Aharon M., Anava O., Avigdor-Elgrabli N., Drachsler-Cohen D., Golan S., and Somekh O., "*ExcUseMe: Asking Users to Help in Item Cold-Start Recommendations*", In Proceedings of the 9th ACM Conference on Recommender Systems, pp. 83-90, ACM, 2015.

[19] Fernández-Tobías I., Braunhofer M., Elahi M., Ricci F., and Cantador, "*I Alleviating the new user problem in collaborative filtering by exploiting personality information*", User Modeling and User-Adapted Interaction, pp. 1-35, 2015.

[20] Barjasteh I., Forsati R., Masrour F., Esfahanian A. H., and Radha H., "*Cold-start item and user recommendation with decoupled completion and transduction*", In Proceedings of the 9th ACM Conference on Recommender Systems, pp. 91-98, ACM, 2015.

[21] Pereira A. L. V. and Hruschka E. R., "*Simultaneous co-clustering and learning to address the cold start problem in recommender systems*", Knowledge-Based Systems, Vol. 82, pp. 11-19, 2015.

[22] *GroupLens Research*, MovieLens, 2014, Available at: http://grouplens.org/datasets/movielens/.

[23] Ahn H. J., "*A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem*", Information Sciences, Vol. 178, No. 1, pp. 37-51, 2008.

[24] Alatas B., Akin E., and Ozer A. B., "*Chaos embedded particle swarm optimization algorithms*", Chaos, Solitons & Fractals, Vol. 40, No. 4, pp. 1715-1734, 2009.

[25] Rousseeuw P. J., "*Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*", Journal of computational and applied mathematics, Vol. 20, pp. 53-65, 1987.

[۲۶] حمیدزاده، جواد، صادق‌زاده، مریم، «**فیلترکنندهٔ مشارکتی فازی ناهموار مبتنی بر کاربر در سیستم‌های پیشنهاددهنده**»، مجله مهندسی برق دانشگاه تبریز، دورهٔ ۴۷، شمارهٔ ۲، صفحهٔ ۴۹۱_۵۰۰، ۱۳۹۶.