مجله تحقیق در عملیات در کاربردهای آن، سال نوزدهم، شماره سوم (پیاپی ۷۴)، پاییز ۰۱، ص ص ۱۵-۱
شاپا ۷۲۸۶-۲۲۵۱

تکامل خوشه‌های کاربران بر اساس تغییر سلیقه آگاه بر زمان در سیستم پیشنهاددهنده

راهله قوچان نژادنورنیا[۱]، مهرداد جلالی[۲*]

۱- دانشجوی دکتری، گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

۲- دانشیار، گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

رسید مقاله: ۵ آذر ۱۳۹۹

پذیرش مقاله: ۵ اردیبهشت ۱۴۰۰

چکیده

فراوانی حجم داده‌ها در اینترنت، برای کاربران مسایلی را به وجود آورده است و باعث سردرگمی برای یافتن اطلاعات مورد نیاز آنها شده است. همچنین سلیقه و اولویت کاربران با گذشت زمان تغییر می کند. سیستم‌های پیشنهاددهنده می توانند در یافتن اطلاعات مفید به کاربران کمک کنند. با توجه به تغییرات علایق، سیستم‌ها باید بتوانند تکامل پیدا کنند. جهت تکامل، کاربران خوشه‌بندی می‌شوند و برای تعیین کاربران هم‌خوشه، به امتیازاتی که کاربر به آیتم‌ها داده است، توجه می‌شود. پارامتر زمان در روش پیشنهادی الگوریتم ژنتیک-تبرید تدرجی (SAGA) این مقاله مورد توجه قرار گرفته است که بتواند اولویت‌بندی کاربر را بر اساس زمان بهبود دهد. در روش پیشنهادی، با استفاده از الگوریتم تکاملی ممتیک، خوشه‌ها در طول زمان بهبود می‌یابند، و بر اساس خوشه‌بندی انجام‌شده، پیشنهادات مناسبی به کاربر ارایه می‌شود. همچنین این سیستم، با استفاده از ویژگی‌های آیتم برای مساله‌ی آیتم شروع سرد، و اطلاعات دموگرافیک کاربر برای مساله‌ی کاربر شروع سرد، خوشه‌بندی تکاملی بهینه را انجام می‌دهد. روش پیشنهادی با استفاده از مجموعه داده Movielens ارزیابی شده است و نتایج تجربی نشان می‌دهد که روش پیشنهادی SAGA با صحت ۰/۸۹، نسبت به روش‌های موجود عملکرد بهتری در صحت پیش‌بینی و پیشنهادات به کاربران دارد.

کلمات کلیدی: الگوریتم تکاملی، ممتیک، شروع سرد، آگاه با زمان.

۱ مقدمه

در ابتدا به تعاریفی از اصطلاحات مورد استفاده پرداخته می‌شود. الگوریتم تکاملی ممتیک، به الگوریتمی گفته می‌شود که از ترکیب الگوریتم‌های تکاملی با یادگیری خود، عملکرد بهینه‌تری از خود نشان می‌دهند. همچنین مساله شروع سرد کاربر، به کاربری گفته می‌شود که تازه در سیستم ثبت‌نام می کند و پیش‌زمینه‌ای از اطلاعات و

* عهده‌دار مکاتبات

آدرس الکترونیکی: jalali@mshdiau.ac.ir

[DOI: 10.52547/jamlu.19.3.1]

علایق کاربر موجود نمی‌باشد. شروع سرد آیتم نیز به آیتم‌هایی دلالت می‌کند که جدیدا در سیستم ثبت شده‌اند و هنوز کاربران به این آیتم جدید امتیازی نداده‌اند. اطلاعات دموگرافیک، مشخصات شخصی هست که کاربر در سیستم ثبت می‌کند، این اطلاعات مربوط به سن، جنسیت، محل زندگی و شغل آنها می‌باشد و خوشه‌بندی را بر اساس اطلاعات پروفایل ثبت‌شده، انجام می‌دهد. همچنین لحظه زمانی صفر به زمان فعلی و لحظه زمانی $t+1$ به لحظه بعدی دلالت دارد. در ابتدا که کاربر وارد سیستم می‌شود شروع سرد آن بررسی می‌شود و سپس با استفاده از ماتریس تعیین می‌شود که هر کاربر چه امتیازی به آیتم داده است. در لحظه اول اعضا استخراج می‌شود و در لحظه بعدی نیز تحلیل می‌شود که آیا سلیقه کاربر طی زمان تغییر کرده است یا خیر. با الگوریتم k-means خوشه‌بندی انجام می‌شود و جمعیت اولیه ایجاد می‌شود تا الگوریتم ممتیک وارد عمل‌شده و با محاسبه تابع هدف و بررسی این‌که آیا سلیقه کاربر در لحظه T با لحظه اول $t=0$ تغییر کرده است یا خیر، فاز خوشه‌بندی به پایان برسد. پس از عمل خوشه‌بندی و محاسبه شباهت، همسایگان مناسب کاربر هدف استخراج می‌شوند و برترین آیتم‌های مناسب به کاربر هدف پیشنهاد می‌شود.

در این مقاله با استفاده از الگوریتم ممتیک تکاملی که از ترکیب الگوریتم ژنتیک با الگوریتم تبرید تدریجی به‌دست آمده است به خوشه‌بندی کاربران و آیتم‌ها پرداخته شده است تا دقت سیستم افزایش یابد. سیستم به پیش‌بینی سلیقه کاربران می‌پردازد که سیستم ارایه‌شده می‌تواند با کاربر یا آیتم شروع سرد نیز مقابله کند. در ادامه مقاله در بخش دوم به کارهای مرتبط می‌پردازد. در بخش سوم روش پیشنهادی توصیف شده است. ارزیابی روش پیشنهادی در بخش چهارم انجام شده است. در نهایت در بخش پنجم نتیجه‌گیری مطرح می‌شود.

۲ کارهای مرتبط

روشی ارایه شده است که بر اساس تقسیم نوع فیلم کار می‌کند و از الگوریتم بهینه‌سازی ازدحام ذرات برای انتخاب همسایه‌های آیتم هدف کار می‌کند [۱]. با کاهش ابعاد ویژگی به خوشه‌بندی آیتم‌ها پرداخته شده است و تکاملی در خوشه‌بندی با گذشت زمان اتفاق نمی‌افتد. همچنین آیتمی که به تازگی وارد سیستم می‌شود را نمی‌تواند خوشه‌بندی کند، زیرا با استفاده از امتیازاتی که کاربران به آیتم‌ها می‌دهند عمل می‌کند [۱]. با استفاده از ویژگی‌های مرتبط بر اساس آنتروپی نظارت‌شده، اعضای یک گروه در شبکه را شناسایی می‌کنند و با قوانین ارتباط و طبقه‌بندی سلسله مراتبی، به حذف اعضای نامرتبط می‌پردازد [۲]. با شناسایی و حذف اعضای ناهمگن، پیشنهاداتی را به کاربران ارایه می‌دهد ولی کاربران را با به روزرسانی خوشه‌ها در گروه جدید قرار نمی‌دهد تا خوشه‌ها تکامل یابند [۲]. در این مقاله، در روش پیشنهادی SAGA، داده‌ها با خوشه‌بندی در زمان‌های مختلف تکامل پیدا می‌کنند و به صورت پویا رفتار می‌کنند. خوشه‌بندی را در شروع کار (لحظه فعلی)، بدون در نظر گرفتن بعد زمان انجام می‌دهد؛ زیرا در ابتدا زمانی از فعالیت کاربر نگذشته است تا بتواند سلیقه کاربر را شناسایی کند؛ و ممکن است کاربر با گذشت زمان، به آیتم دیگری علاقه‌مند شود، این سیستم نمی‌تواند این تغییر را در خوشه‌بندی اثر دهد و منجر به پیشنهادات نامناسب به کاربر می‌شود [۳]. بنابراین، بزرگ‌ترین مساله، بهبود دقت سیستم، همراه با تغییر نیازهای کاربر و رتبه‌بندی محتوا است. سیستم‌های پیشنهاددهنده‌ی معمولی، نمی‌توانند نرخ

[DOI: 10.52547/jamlu.19.3.1]

تغییرات علایق کاربران را با توجه به اولویت‌های کاربر، در نظر بگیرند. حفظ دقت، همراه با اعمال تغییرات مورد نیاز کاربر، کار دشواری است. گنجاندن بعد زمان در سیستم‌های پیشنهاددهنده، می‌تواند در حل بسیاری از مسایل، کمک کند [۴]. در [۵] نشان داده شد که چگونه تغییرات در یک دوره از زمان، می‌تواند بر دقت، تاثیر گذارد. الگوریتم‌های تکاملی با بعد زمان، بهبود خوبی را در روند توصیه ارایه می‌کنند. در [۳]، یک مفهوم از خوشه‌بندی تکاملی ارایه داده شده است که فرآیند خوشه‌بندی را در پیدا کردن کاربران مشابه و خوشه‌های بهتر، بسیار موثر نشان داده است و می‌تواند دقت را در فرآیند توصیه، همچنین در مقیاس‌پذیری، بهبود دهد. البته با تغییر اولویت کاربران، ثابت رفتار می‌کند [۳].

در [۶] الگوریتم تکاملی با نام الگوریتم فرهنگی استفاده شده است که از مؤلفه‌های دانش در یک فضای باور برای هدایت جستجو در فضای جمعیت استفاده می‌کند. پس از خوشه‌بندی، همسایگان مستقیم که بیشترین اعتماد را در هر جامعه دارا می‌باشند، جمع می‌کند. از این مقادیر رتبه‌بندی‌شده به‌روز، برای طراحی الگوریتم فرهنگی استفاده می‌شود که همسایگان مستقیم را در نظر دارد و با صحت پایین‌تری نسبت به روش پیشنهادی SAGA عمل می‌کند. در [۷] رویکرد نظم اجتماعی که شامل اطلاعات شبکه های اجتماعی است برای بهره‌مندی از سیستم‌های پیشنهادی با اطلاعات اعتماد بین کاربران، ترکیب می‌شود. سوابق اعتماد و رتبه‌بندی (برچسب‌ها) برای پیش‌بینی، در ماتریس کاربر استفاده می‌شوند. از یک الگوریتم برای انتخاب مسیر قابل اطمینان برای تولید پیشنهادات استفاده می‌کند [۷]. از همبستگی بین کاربران هر گروه، برای انتخاب نزدیک‌ترین همسایه برای پیش‌بینی رتبه‌بندی استفاده می‌شود [۸]. الگوریتمی با عنوان فیلتر مشارکتی مبتنی بر همبستگی کاربر و خوشه تکاملی مطرح‌شده است که از الگوریتم تبرید تدریجی استفاده نمی‌کند و دارای خطای میانگین نسبتا بالایی نسبت به روش پیشنهادی در این مقاله می‌باشد [۸]. در [۹] یک مدل مخفی مارکوف سلسله مراتبی برای استخراج زمینه نهفته از داده‌ها ارایه شده است. زمینه‌های نهفته از الگوهای پنهان بدون نظارت تشکیل شده‌اند. موارد انتخاب‌شده کاربر در یک بازه زمانی معین به عنوان دنباله در نظر گرفته می‌شود و همچنین فرض می‌کند که این عوامل پویا هستند و باید از تعامل کاربر با سیستم استنباط شود. با این‌که بازه‌های زمانی را در نظر می‌گیرد ولی از الگوریتم تکاملی جهت بهبود خوشه‌بندی بهره نمی‌برد و دقت سیستم خیلی پایین می‌باشد. در [۱۰] محتوای میکروبلاگ یک بازخورد عمومی در شخصی‌سازی دارد و پراکندگی محتوا در میکروبلاگ مشکلاتی را برای پیشنهادات به‌وجود می‌آورند که احتیاج به تگ‌هایی دارد که کاربران درباره محتوا زده‌اند، و ماتریس کاربر-تگ بپیشنهادات مناسبی را بدون در نظر گرفتن بعد زمان به کاربران ارایه می‌دهد. در [۱۱] به امتیازاتی که کاربر در زمان‌های مختلف به آیتم‌ها می‌دهد توجه نمی‌شود. در [۱۲] از اطلاعات دموگرافیک و ویژگی‌های آیتم استفاده نمی‌شود تا بتواند خوشه‌بندی را بهبود بخشد. در [۱۳] از روش ترکیبی یادگیری عمیق برای چالش شروع سرد آیتم و شروع سرد کاربر استفاده شده است که سیستم پیشنهاددهنده برای آیتم‌ها و کاربران موجود در سیستم را بررسی نمی‌کند.

[DOI: 10.52547/jamlu.19.3.1]

۳ روش پیشنهادی

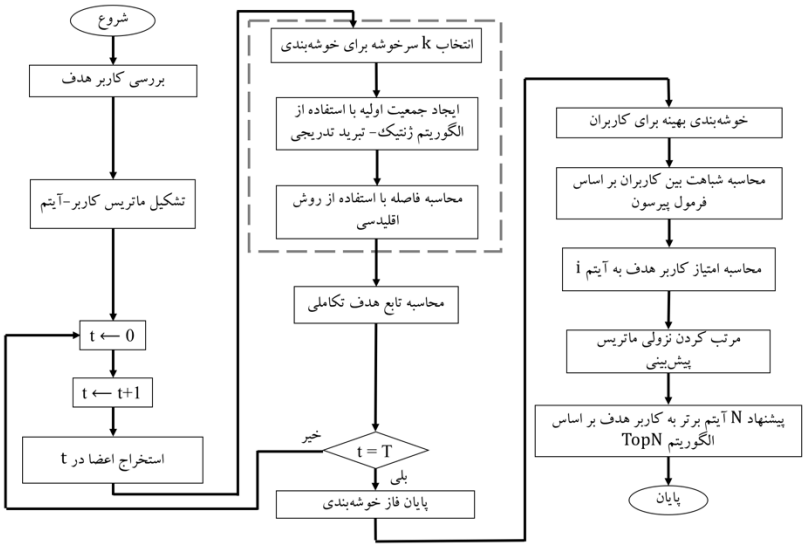
طراحی الگوریتم ممتیک کارا به جزییات مساله وابسته است. روش پیشنهادی در این مقاله، الگوریتم تکاملی ممتیکی است که الگوریتم ژنتیک-تبرید تدریجی[1] (SAGA) نام‌گذاری شده است. جمعیت اولیه الگوریتم ژنتیک شامل کاربران سیستم می‌شود که بایستی در خوشه مناسب خود قرار بگیرند. با اعمال عملگرهای ترکیب و جهش، خوشه‌های بهینه را استخراج می‌کند و پس از محاسبه فاصله کاربران با همسایگانش با استفاده از روش اقلیدسی و محاسبه تابع تناسب، بهترین خوشه‌بندی انجام می‌شود. فرآیند این روش با اجرای الگوریتم ژنتیک برای یافتن بهترین خوشه‌بندی شروع می‌شود و برای یافتن بهینه سراسری در خوشه‌بندی، الگوریتم تبرید تدریجی وارد کار می‌شود تا بهترین خوشه‌بندی را استخراج کند. یکی از مسایل بهینه‌سازی در الگوریتم‌های تکاملی یافتن تابع هدف بهینه می‌باشد. معمولا، الگوریتم‌های خوشه‌بندی سنتی در بهینه‌سازی تابع هدف سراسری شکست می‌خورند. در این مقاله، یک تابع هزینه برای بهینه‌سازی تابع هدف، پیشنهاد شده است، که نتایج دقیق خوشه‌بندی را، به ویژه برای مجموعه داده تکاملی، تولید می‌کند.

در شکل ۱، نمودار روش پیشنهادی رسم شده است که ابتدا جدید بودن کاربر را به عنوان شروع سرد بررسی می‌کند. چنانچه کاربر شروع سرد باشد، به اطلاعات دموگرافیک کاربر مراجعه می‌کند و براساس مشخصات موجود، عمل خوشه‌بندی را انجام می‌دهد تا بتواند با تکرار خوشه‌بندی بهترین گروه مناسب را برای کاربر جدید پیدا کند. در ادامه مبتنی بر همسایگان کاربر، به کاربر شروع سرد پیشنهادات مناسب را ارایه می‌کند. اما اگر کاربر هدف در سیستم حضور داشته باشد و به عنوان کاربر شروع سرد شناخته نشود، ماتریسی بر اساس رتبه‌هایی که قبلا کاربر به آیتم ها داده است، به نام ماتریس کاربر- آیتم ایجاد می‌شود. سلیقه کاربر در لحظه زمانی صفر که به معنای زمان فعلی می‌باشد و در لحظه زمانی بعدی $(t+1)$بررسی می‌شود تا بتواند اعضای مناسب را جهت خوشه‌بندی در هر لحظه زمانی استخراج کند و ماتریس کاربر-آیتم را ایجاد کند. به صورت تصادفی عددی را سرخوشه K الگوریتم خوشه‌بندی K-means انتخاب می‌کند و عمل خوشه‌بندی را تا زمانی که جمعیت اولیه را برای ورودی الگوریتم ژنتیک ایجاد کند، تکرار می‌کند و بهترین خوشه‌بندی را براساس فاصله اقلیدسی تعیین می‌کند. در این مرحله برای یافتن بهینه‌ترین K از الگوریتم تبرید تدریجی بهره می‌برد. تابع هدف الگوریتم محاسبه می‌شود و حالت بهینه خوشه‌بندی را استخراج می‌کند، سپس در خوشه‌بندی بهینه در لحظه T را با لحظه اول که $t=0$ بود بررسی می‌کند که آیا سلیقه کاربر تغییر کرده است یا خیر. در فاز آخر، شباهت کاربران براساس فرمول پیرسون به‌دست می‌آید و با پیش‌بینی، پیشنهادات را با الگوریتم Top-N مرتب کرده و به کاربر پیشنهاد می‌دهد.

از آنجایی که در این مقاله، با مساله شروع سرد آیتم هم مقابله می‌شود؛ برای خوشه‌بندی آیتم شروع سرد، ویژگی‌های آیتم استخراج و بر اساس خوشه‌های موجود کاربری و ژانرهای مورد علاقه آن‌ها، آیتم شروع سرد به کاربران خوشه مربوطه پیشنهاد می‌شود. در خوشه‌بندی تکاملی، تمرکز بر کیفیت خوشه‌ها است.

[1] Simulated Annealing Genetic Algorithm

[DOI: 10.52547/jamlu.19.3.1]



شکل ۱. نمودار روش پیشنهادی

۳-۱ فرموله‌سازی مساله

برای پیاده‌سازی چنین سیستم پیشنهاددهنده‌ای، چارچوب خوشه‌بندی SAGA ارایه شده است. خوشه‌بندی تکاملی ارایه‌شده در [۳]، بعد زمان را استفاده نمی‌کند؛ که در این الگوریتم، خوشه‌ها، بر اساس بازه‌ی زمانی، داده‌ها را دسته‌بندی می‌کنند. در هر بازه زمانی، یک خوشه جدید، به‌وسیله بهینه‌سازی دو پارامتر جدید، به نام کیفیت خوشه‌بندی لحظه‌ای و هزینه پیشینه، تولید می‌شوند. کیفیت خوشه‌بندی لحظه‌ای، به صحت آن بستگی دارد. هزینه پیشینه، بیان می‌کند که خوشه جدید، با خوشه قبلی چقدر تفاوت دارد. مجموع این توالی که تابع هدف روش SAGA را بیان می‌کند توسط فرمول (۱) نشان داده می‌شود:

$$FitnessFunction = \sum_{t=1}^{T} SQ(C_t M_t) - \sum_{t=2}^{T} HC(C_{t-1}, C_t) \qquad (1)$$

که در آن $SQ(C_t M_t)$، کیفیت خوشه‌بندی لحظه‌ای، از خوشه C_t در زمان t با ورودی m است. $HC(C_{t-1}, C_t)$ هزینه پیشینه خوشه C_t در زمان $t-1$ است.

این تابع، دو هدف رقابتی کیفیت خوشه‌بندی لحظه‌ای SQ و هزینه پیشینه HC را بهینه می‌کند. معیار کیفیت خوشه‌بندی، چگونگی یک عمل خوب از خوشه‌ها را، در بازه زمانیt، بر روی داده‌ها نشان می‌دهد. این معیار، به کمک مقیاسی که واریانس نام دارد، تعریف می‌شود که این واریانس، تفاوت بین آیتم‌ها در هر خوشه را کمینه و شباهت آن‌ها را بیشینه می‌کند. پیاده‌سازی واریانس در [۱۴] و [۱۵] انجام‌شده و در [۳]، نیز، کیفیت لحظه‌ای مؤثر، ارایه گردیده است. تفاوت واریانس امتیاز، بین دسته‌بندی آیتم‌ها، در یک خوشه‌ی خاص، در یک نقطه از زمان، می‌باشد. برای محاسبه کیفیت خوشه‌بندی لحظه‌ای از فرمول (۲) و برای به‌دست آوردن میزان امتیاز واریانس از فرمول (۳) استفاده می‌شود [۴]:

$$SQ(C_t M_t) = \sum_{t=1}^{T} (1 - VScore(M_t, t)) \qquad (2)$$

[DOI: 10.52547/jamlu.19.3.1]

$$VScore(M_t,t)=\sum_{t=1}^{T}\frac{\sum_{u_k\in K(u_t)}(r(u_k,i)-\bar{r}(i))^2}{K} \quad (3)$$

که SQ، میزان کیفیت خوشه‌بندی لحظه‌ای داده‌ها در مرحله زمانی t است، که روش‌ها با واریانس امتیاز نشان داده می‌شوند و توسط اختلاف امتیازهای آیتم‌های یک خوشه‌ی خاص، در مرحله زمانی خاص محاسبه می‌شود.

VScore امتیاز واریانس، نشان‌دهنده‌ی تفاوت هر داده از میانگین، برای k همسایه است، که به آیتم i، در بازه زمانی t، در ماتریس M_t، امتیاز داده شده است. $r(u_k,i)$ امتیاز k کاربر همسایه nام، به آیتم i در مرحله زمانی t است که در ماتریس M_t قرار می‌گیرد؛ همچنین، $\bar{r}(i)$ میانگین امتیازات کاربران همسایه nام به آیتم i در مرحله زمانی t می‌باشد.

HC هزینه‌ی پیشینه می‌باشد و با استفاده از معیار آنتروپی نرمال شده، تعریف می شود. NMI اطلاعات متقابل نرمال‌شده می‌باشد، $NMI(t,t-1)$، که با استفاده از فرمول (۴) تعریف می‌شود [۳]:

$$NMI(t,t-1)=-2\frac{\sum_{i=1}^{C_t}\sum_{j=1}^{C_{t-1}}C_{ij}\log(C_{ij}N/C_iC_j)}{\sum_{i=1}^{C_t}C_i\log(C_i/N)+\sum_{j=1}^{C_{t-1}}C_j\log(C_j/N)} \quad (4)$$

C_t تعداد خوشه‌ها در مرحله زمانی t را در نظر می‌گیرد، C_{t-1} تعداد گروه ها در مرحله زمانی $t-1$ است؛ C_i مجموع عناصر C در سطر iام و C_j مجموع عناصر C در ستون jام می‌باشد و N تعداد نودها را نشان می‌دهد.

۳-۲ پیش‌بینی

به منظور محاسبه پیش‌بینی امتیاز P_{u_t,i_t}، برای کاربر هدف و آیتم هدف (u_t,i_t)، مراحل انجام‌شده به شرح ذیل می‌باشند: در مرحله اول، محاسبه ارزش کاربر مشابه مقصد، با هریک از کاربران مدل جانشین، که در استفاده شخصی امتیاز داده شده‌اند و با استفاده از ضریب همبستگی در فرمول (۵) به‌دست می‌آید، تا کاربران مشابه جانشین کاربر هدف پیدا شوند [۴]:

$$S_{u_t,u_k}=\frac{\sum_{i\in I}(r_{u_t,i}-\bar{r}_{u_t})(r_{u_k,i}-\bar{r}_{u_k})}{\sqrt{\sum_{i\in I}(r_{u_t,i}-\bar{r}_{u_t})^2\sum_{i\in I}(r_{u_k,i}-\bar{r}_{u_k})^2}} \quad (5)$$

I مجموعه آیتم‌های مربوط به کاربر هدف و iامین کاربر همسایه است، و $r_{u_t,i}$ امتیاز کاربر هدف u به آیتم i در مرحله زمانی t می‌باشد؛ همین طور که r_{u_t} میانگین امتیازات کاربر هدف u در مرحله زمانی t را در نظر می‌گیرد، $r_{u_k,i}$ امتیاز کاربران همسایه nام به آیتم i در مرحله زمانی t را نشان می‌دهد و $\bar{r}_{u_k}$ میانگین امتیازات کاربران همسایه nام در مرحله زمانی t را در خود جای می‌دهد.

در مرحله دو، پیش‌بینی با استفاده از میانگین وزنی تنظیم‌شده، به وسیله‌ی فرمول (۶) محاسبه می‌شود:

$$P_{u_t,i_t}=\bar{r}_{u_t}+\frac{\sum_{n=1}^{K}(r_{u_k,i}-\bar{r}_{u_k})*S_{u_t,u_k}}{\sum_{n=1}^{K}S_{u_t,u_k}} \quad (6)$$

[DOI: 10.52547/jamlu.19.3.1]

$r_{u_k,i}$ امتیاز کاربران همسایه nام به آیتم i در مرحله زمانی t را نشان می‌دهد، $\bar{r}_{u_t}$ میانگین امتیازات کاربر هدف در مرحله زمانی t می‌باشد، S_{u_t,u_k} میزان شباهت کاربر هدف و کاربران همسایه nام در مرحله زمانی t است و k تعداد همسایه‌ها در مرحله زمانی t را در بر دارد.

٤ ارزیابی روش پیشنهادی

یک سیستم پیشنهاددهنده، به ازای یک کاربر هدف، تمام آیتم‌هایی که امتیازی به آنها تخصیص داده نشده را پیش‌بینی کرده و آیتم‌هایی با بالاترین نرخ را پیشنهاد می‌کند. به منظور ارزیابی الگوریتم‌های پیشنهاددهنده، داده‌ها معمولاً به دو قسمت آموزش و تست تقسیم می‌شوند و از فرمول‌های صحت، فراخوانی، میانگین خطای مطلق و F-measure استفاده می‌شود. مجموعه داده مورد استفاده، مجموعه داده Movielens 100K، که یک سیستم پیشنهاددهنده‌ی تحقیقاتی بر مبنای وب است، استفاده شده است. این دیتاست شامل ۱۰۰۰۰۰ امتیاز است، که توسط ۹۴۳ کاربر به ۱۶۸۲ فیلم داده شده است (ماتریس آیتم-کاربر از ۹۴۳ سطر و ۱۶۸۲ ستون تشکیل می‌شود). امتیازاتی که کاربران به فیلم‌ها می‌دهند بین بازه ۱ تا ۵ می‌باشد و هر کاربر حداقل ۲۰ فیلم را امتیازدهی کرده است. ویژگی‌های موجود در دیتاست برای کاربران شامل آیدی کاربر، آیدی فیلم، امتیاز داده شده و مرحله زمانی می‌باشد.

٤-١ نتایج آزمایشات

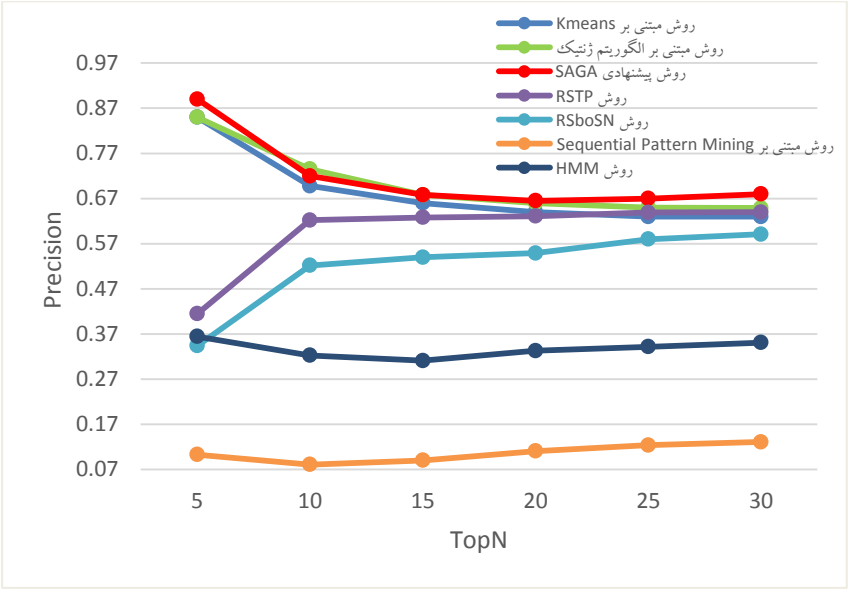
٤-١-١ ارزیابی روش پیشنهادی

برای ارزیابی سیستم پیشنهاددهنده SAGA، مقدار صحت، میانگین خطای مطلق، فراخوانی و F-measure روش مبتنی بر خوشه‌بندی K-means معمولی [۳]، روش خوشه‌بندی آگاه با زمان غنی‌شده با الگوریتم ژنتیک [۴]، روش خوشه‌بندی آگاه با زمان غنی‌شده با الگوریتم ژنتیک- تبرید تدریجی پیشنهادی SAGA، روش Sequential Pattern Mining [۹]، روش مدل مخفی مارکوف [۹]، روش RSTP [۷] و روش RSboSN [۷] در شکل‌های ۲، ۳، ۴ و ۵ به‌ترتیب نشان داده شده است. همان‌طور که مشاهده می‌شود، روش پیشنهادی SAGA دارای میانگین خطای پایین‌تری نسبت به دو روش دیگر می‌باشد. با توجه به شکل ۲ در پنج پیشنهاد اول ارایه‌شده به کاربر، صحت تا ۰/۸۹ افزایش‌یافته است و با افزایش تعداد پیشنهادات، صحت کمی کاهش می‌یابد؛ ولی با توجه به نتایج، باز هم صحت روش پیشنهادی از روش‌های دیگر بهتر می‌باشد. همچنین، هر چه تعداد پیشنهادها افزایش می‌یابد، خطای روش پیشنهادی نسبت به روش مبتنی بر K-means و روش مبتنی بر الگوریتم ژنتیک کاهش می‌یابد. در تمامی نمودارهای ارزیابی، محور افق، تعداد پیشنهادات ارایه شده به کاربران می‌باشد و محور عمود، یکی از پارامترهای صحت، میانگین خطای مطلق، فراخوانی و F-measure می‌باشد.

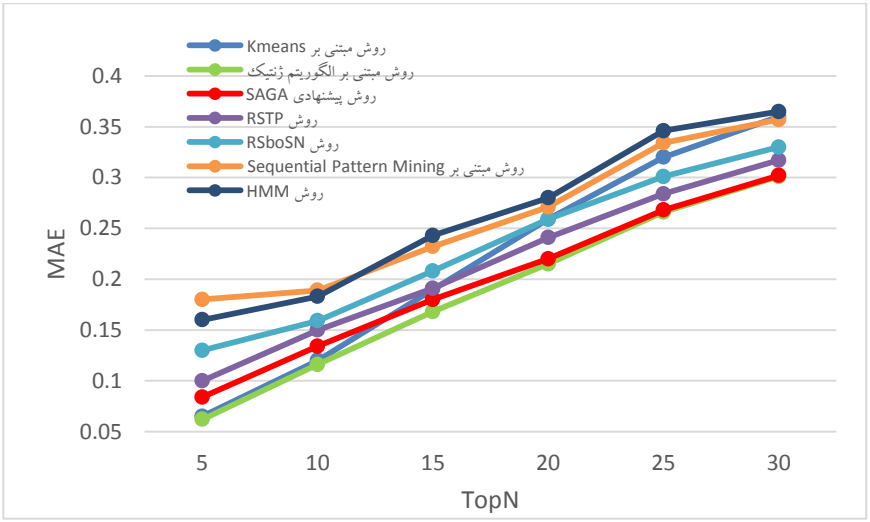
فراخوانی برای کل کاربران با در نظر گرفتن زمان در روش SAGA مانند روش‌های دیگر رو به رشد است که در شکل ۴ نمایش داده شده است. پارامتر F-measure عملکرد دسته‌بند را ارزیابی می کند که از ترکیب دو پارامتر صحت و فراخوانی به‌دست می‌آید. با توجه به نتایج نشان داده شده در جدول ۱، روش پیشنهادی SAGA

[DOI: 10.52547/jamlu.19.3.1]

نسبت به روش Sequential Pattern Mining که به صورت متوالی خوشه‌بندی را انجام می‌دهد [۹] بسیار دقیق‌تر عمل می‌کند؛ همچنین از مدل مخفی مارکوف [۹]، روش RSboSN [۷] و روش RSTP [۷] عملکرد بهتری دارد. با این‌که روش مبتنی بر K-means [۳] و روش مبتنی بر الگوریتم ژنتیک [۴] عملکرد خوبی را ارایه داده‌اند ولی روش پیشنهادی این مقاله، عمل پیش‌بینی را نسبت به روش‌های موجود، دقیق‌تر انجام می‌دهد که صحت به‌دست آمده ۰/۸۹ می‌باشد. همان‌طور که صحت روش پیشنهادی در جدول ۱ نشان داده شده است، این سیستم قابلیت ارایه پیشنهادات مناسب و دقیق را نسبت به سیستم‌های موجود به کاربر دارند که اولویت کاربران را با صحت خوبی در نظر می‌گیرد.

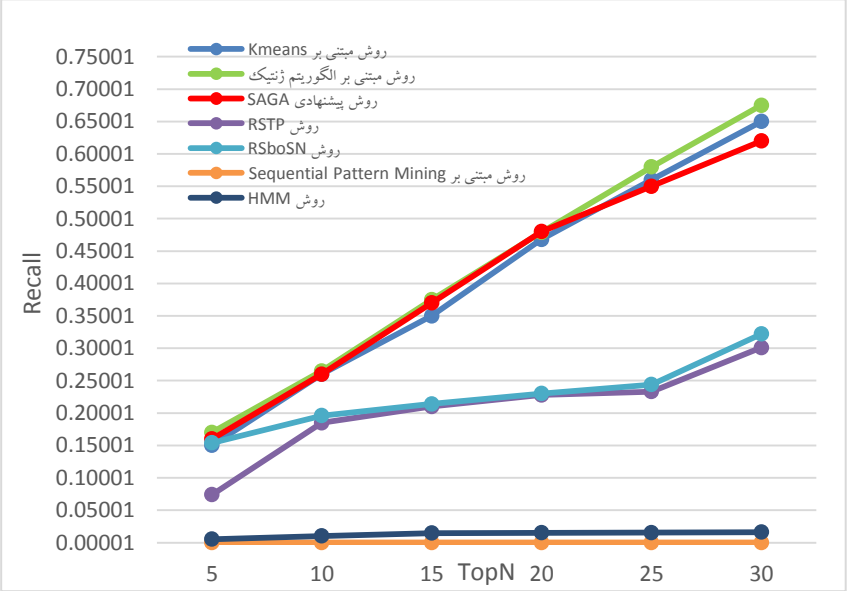


شکل ۲. مقدار صحت روش پیشنهادی در مقایسه با روش‌های دیگر

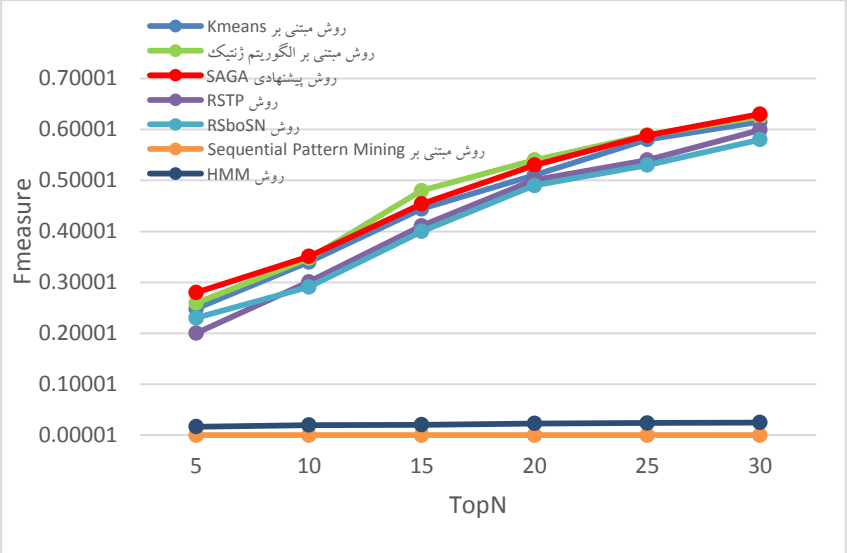


شکل ۳. مقدار MAE روش پیشنهادی در مقایسه با روش‌های دیگر

[DOI: 10.52547/jamlu.19.3.1]



شکل ٤. مقدار فراخوانی روش پیشنهادی در مقایسه با روش‌های دیگر



شکل ٥. مقدار F-measure روش پیشنهادی در مقایسه با روش‌های دیگر

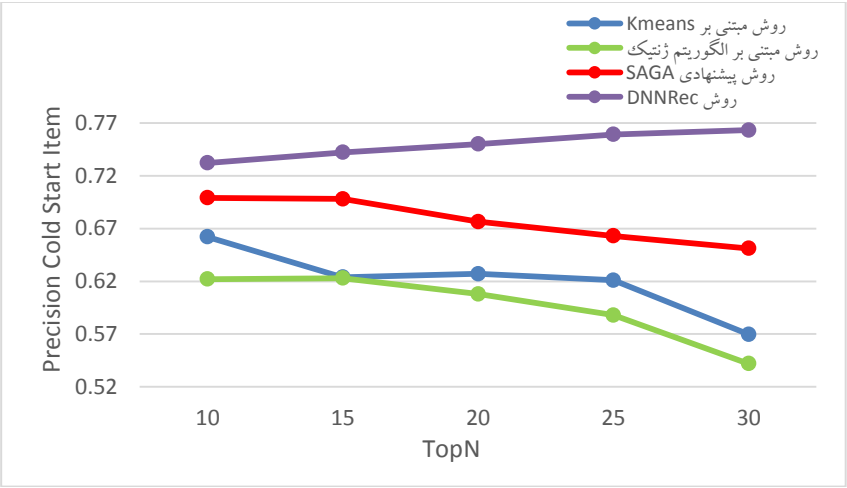
جدول ۱. مقایسه صحت و فراخوانی روش پیشنهادی با روش‌های دیگر

روش مبتنی بر	صحت	فراخوانی
Sequential Pattern Mining	۰/۱۰۳	۳/۵۶
HMM	۰/۳۲	۰/۰۰۶
RSboSN	۰/۵۲۲	۰/۱۹۶۳
RSTP	۰/۶۲۲۲	۰/۱۸۵
K-means	۰/۸۵	۰/۱۵
GA	۰/۸۴۳	۰/۱۸
Proposed SAGA	**۰/۸۹**	**۰/۱۷**

[DOI: 10.52547/jamlu.19.3.1]

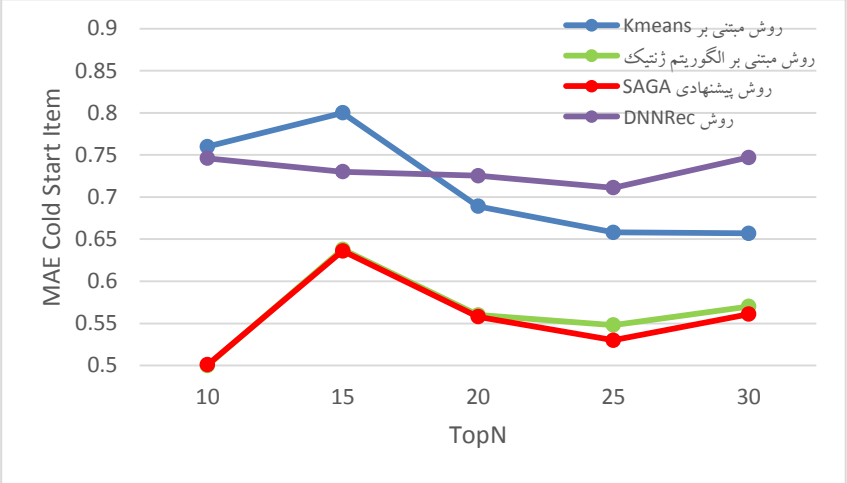
٤-١-٢ ارزیابی نتایج تاثیر عملکرد SAGA بر روی مساله آیتم‌های جدید

تا اینجا، روش پیشنهادی SAGA برای کاربران و آیتم‌های موجود ارزیابی شده است. حال به ارزیابی روش SAGA با رویکرد شروع سرد آیتمی پرداخته می‌شود و نتایج آزمایشاتی که برای تأثیر رویکرد SAGA به منظور پرداختن به مساله‌ی آیتم‌های شروع سرد انجام شده است. البته باید خاطر نشان کرد که مجموعه داده‌ی MovieLens نسبت به مجموعه داده‌های دیگری که برای ارزیابی سیستم‌های پیشنهاددهنده به کار می‌رود، نیز متداول‌تر است و در مجموعه داده‌ی MovieLens، آیتم‌های غیرشروع سرد باید حداقل بیست امتیازدهی داشته باشند. محاسبه معیارهای ارزیابی روش مبتنی بر خوشه‌بندی K-means معمولی [۳]، روش مبتنی بر خوشه‌بندی آگاه با زمان غنی‌شده با الگوریتم ژنتیک [۴]، روش پیشنهادی مبتنی بر خوشه‌بندی آگاه با زمان غنی‌شده با SAGA با رویکرد شروع سرد آیتم و روش DNNRec [۱۵] در شکل‌های ۶، ۷، ۸ و ۹ به ترتیب نشان داده شده است. همان طور که در شکل ۶ مشاهده می‌شود روش SAGA در مساله‌ی شروع سرد، صحت پیش‌بینی پایینی نسبت به روش مبتنی بر k-means معمولی دارد و دارای خطای میانگین بسیار پایینی می‌باشد که بدان معنا است که صحت حدود ۰/۶ درشروع سرد آیتم، با خطای بسیار کمی عمل پیش‌بینی را انجام می‌دهد. نتایج ثابت می‌کند، که رویکرد SAGA و روش الگوریتم ژنتیک کم‌ترین خطا را در شروع سرد آیتم، دارا می‌باشند، و روش SAGA دارای عملکرد فراخوانی بسیار خوبی نسبت به دو روش دیگر است. با در نظر گرفتن این‌که صحت روش SAGA در سه روش مبتنی بر K-means [۳]، مبتنی بر الگوریتم ژنتیک [۴] بالاتر بوده، در نتیجه برای ارزیابی رویکرد شروع سرد آیتم و شروع سرد کاربر، این دو روش با روش پیشنهادی در شکل ۶ مقایسه شده‌اند. با توجه به مقایسه مساله شروع سرد آیتم روش پیشنهادی با روش DNNRec [۱۵] نشان داده شده است که با توجه به شکل ۶ صحت روش پیشنهادی پایین‌تر از روش DNNRec می‌باشد و ولی میانگین خطای مطلق روش پیشنهادی SAGA که در شکل ۷ نشان داده شده است، از روش‌های دیگر پایین‌تر است.

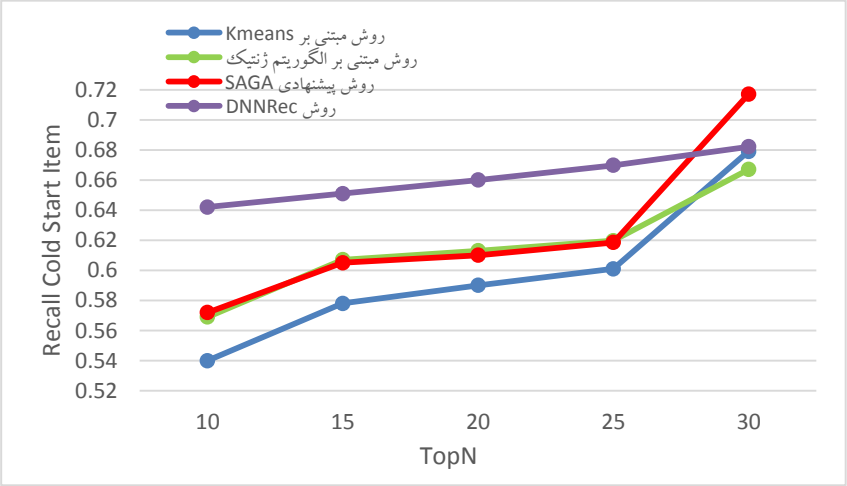


شکل ٦. مقدار صحت روش پیشنهادی با رویکرد شروع سرد آیتم در مقایسه با روش‌های دیگر

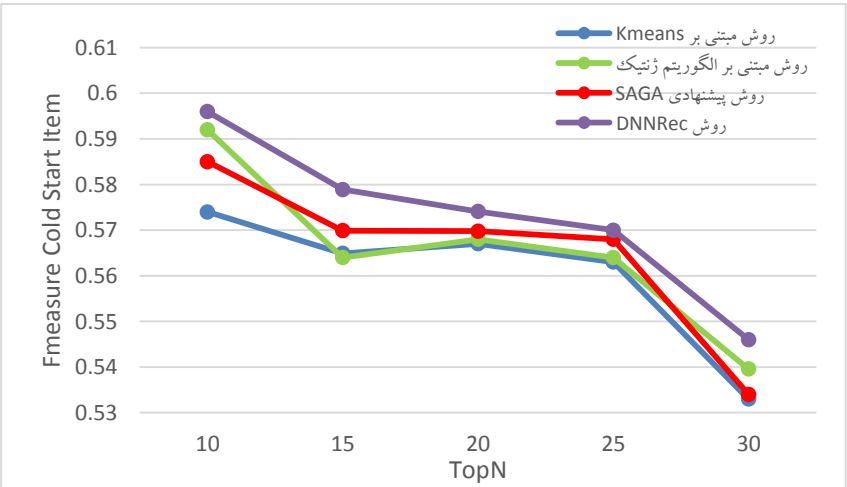
[DOI: 10.52547/jamlu.19.3.1]



شکل ۷. مقدار MAE روش پیشنهادی با رویکرد شروع سرد آیتم در مقایسه با روش‌های دیگر



شکل ۸. مقدار فراخوانی روش پیشنهادی با رویکرد شروع سرد آیتم در مقایسه با روش‌های دیگر

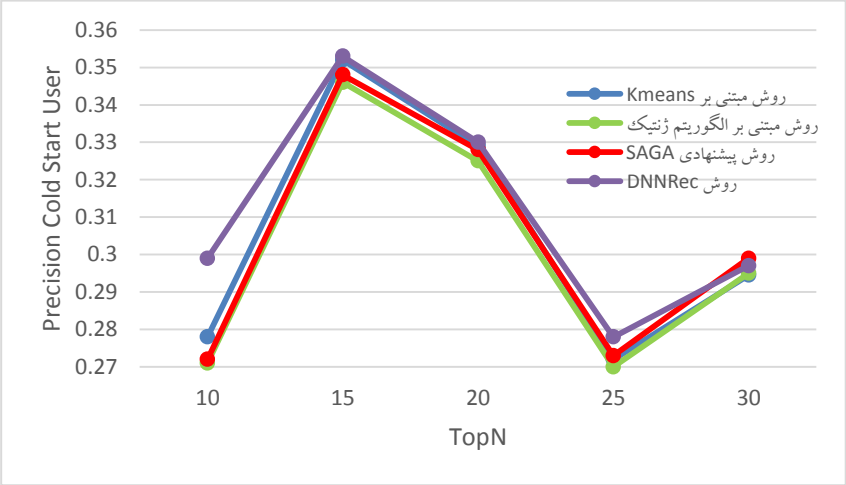


شکل ۹. مقدار F-measure روش پیشنهادی با رویکرد شروع سرد آیتم در مقایسه با روش‌های دیگر

[DOI: 10.52547/jamlu.19.3.1]

٤-١-٣ ارزیابی نتایج تاثیر رویکرد SAGA بر روی مساله کاربران جدید

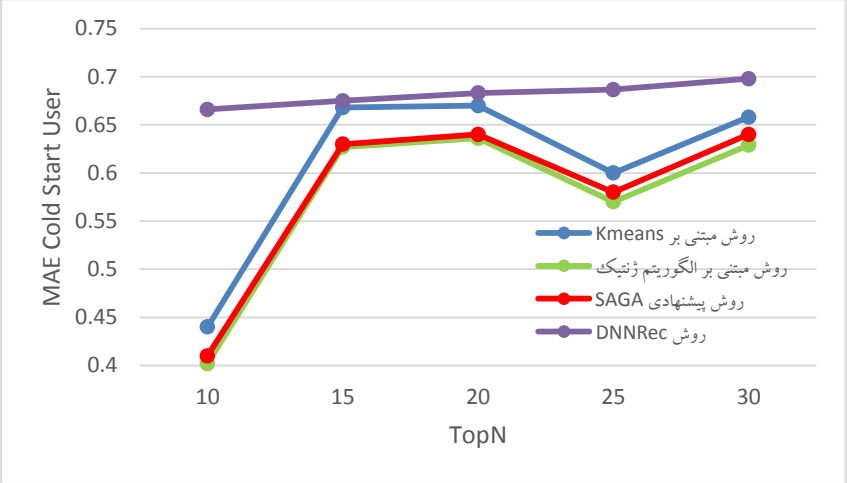
همان‌طور که در قسمت۴-۱-۲ بیان شده است، تمرکز اصلی این مطالعه بر روی خوشه‌بندی بر اساس سلیقه کاربران بوده است و برای مقابله با رویکرد شروع سرد دقت سیستم کاهش می‌یابد. این قسمت، نتایج آزمایشاتی که برای تأثیر رویکرد SAGA برای مساله‌ی کاربران جدید انجام شده، را نشان می‌دهد. مقادیر ارزیابی در روش‌های مبتنی بر خوشه بندی K-means معمولی [۳]، مبتنی بر خوشه‌بندی آگاه با زمان غنی‌شده با الگوریتم ژنتیک [۴]، مبتنی بر خوشه‌بندی آگاه با زمان غنی‌شده با SAGA با رویکرد شروع سرد کاربر و روش DNNRec [۱۵] در شکل‌های ۱۰، ۱۱، ۱۲ و ۱۳ به ترتیب نشان داده شده است.



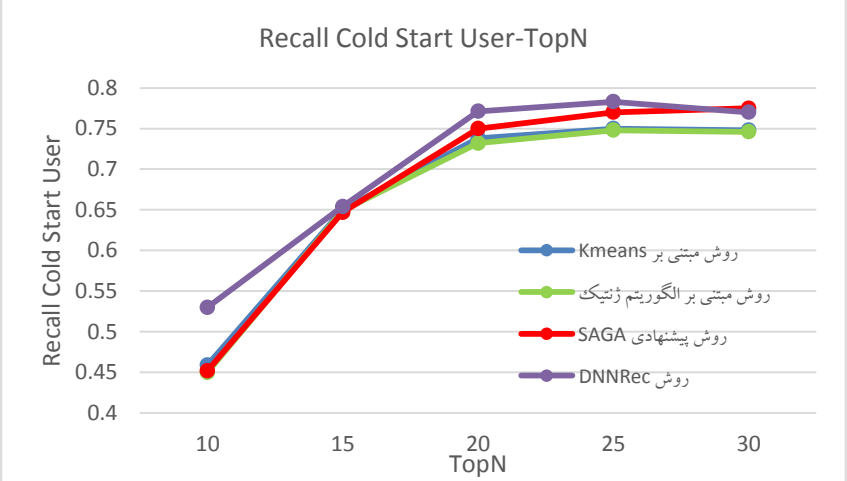
شکل ۱۰. مقدار صحت روش پیشنهادی با رویکرد شروع سرد کاربر در مقایسه با روش‌های دیگر

آزمایشات ثابت می‌کند که رویکرد SAGA و همچنین روش الگوریتم ژنتیک از خطای پایینی برخوردارند. با توجه به نتایج به‌دست آمده در شکل ۱۰، در مساله‌ی شروع سرد کاربر، روش SAGA در تعداد پیشنهادات پایین‌تر از ۳۰، دارای صحت‌های نزدیکی به سایر روش‌ها می‌باشد ولی با افزایش تعداد پیشنهادات، صحت از سایر روش‌ها بهتر عمل می‌کند. همچنین علت افزایش ناگهانی صحت در ۱۵ پیشنهاد به دام افتادن الگوریتم در بهینه محلی را توصیف می کند که پس از رسیدن به کم‌ترین مقدار دقت سیستم در ۲۵ پیشنهاد و اعمال الگوریتم تبرید تدریجی شروع به افزایش صحت می‌شود که هرچه تعداد پیشنهادات بالا می‌رود، صحت سیستم نیز بهبود می‌یابد.

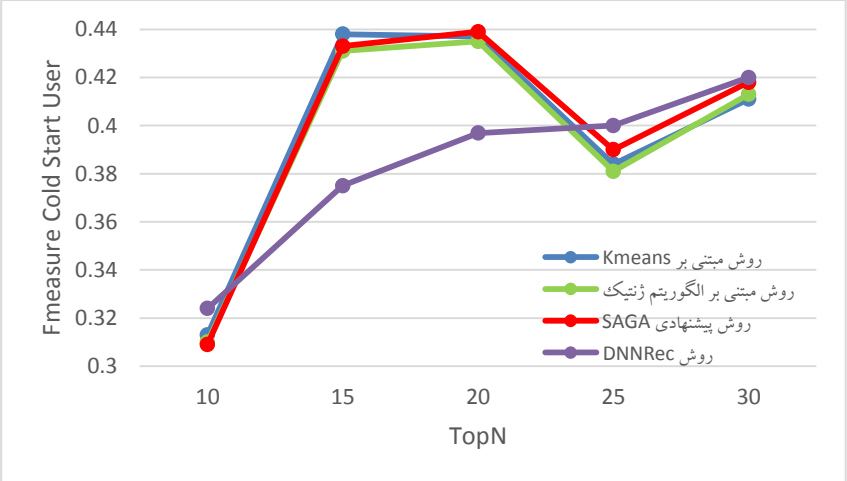
[DOI: 10.52547/jamlu.19.3.1]



شکل ۱۱. مقدار MAE روش پیشنهادی با رویکرد شروع سرد کاربر در مقایسه با روش‌های دیگر



شکل ۱۲. مقدار فراخوانی روش پیشنهادی با رویکرد شروع سرد کاربر در مقایسه با روش‌های دیگر



شکل ۱۳. مقدار F-measure روش پیشنهادی با رویکرد شروع سرد کاربر در مقایسه با روش‌های دیگر

در روش SAGA، با تعداد پیشنهادات بیست عدد یا بیشتر دارای فراخوانی بالاتری نسبت به سایر روش ها می‌باشد که با افزایش تعداد پیشنهادات، افزایش فراخوانی بیشتری اتفاق می‌افتد که در شکل ۱۲ نشان داده شده است.

[DOI: 10.52547/jamlu.19.3.1]

۵ نتیجه‌گیری

امروزه در دنیای تجارت الکترونیک، سیستم‌های پیشنهاددهنده قوی‌ترین ابزار برای شخصی‌سازی اطلاعات مشتریان محسوب می‌شوند. سهم روش پیشنهادی در سیستم‌های پیشنهاددهنده به این معناست که یک الگوریتم تکاملی به منظور بررسی فرآیند تکامل یک سیستم استفاده می‌شود. این روش می‌تواند در پالایش مشارکتی سیستم‌های پیشنهاددهنده برای ادغام داده‌های جدید در طی یک دوره از زمان و به‌روزرسانی مشخصات کاربر با شرایط فعلی آن به‌کار گرفته شود. این مقاله، کارایی الگوریتم خوشه‌بندی تکاملی و تولید خوشه‌های با کیفیت و پیش‌بینی توصیه، را بر اساس مجموعه داده واقعی مورد بررسی قرار داده است. نتایج تجربی در مجموعه داده بزرگ دنیای واقعی نشان می‌دهد که این الگوریتم می‌تواند پیش‌بینی‌های با کیفیت بالا را نسبت به الگوریتم‌های خوشه‌بندی سنتی و دیگر روش‌های مبتنی بر مدل ارایه دهد. بزرگ‌ترین عیب روش‌های مبتنی بر مدل، آموزش نسبتاً آهسته آن‌ها است که در روش پیشنهادی به‌طور قابل توجهی بهبودیافته است. روش پیشنهادی می‌تواند به راحتی در حوزه‌های دیگر نیز استفاده شود. یکی از چالش‌های اصلی در پیاده‌سازی یک الگوریتم مبتنی بر کاربر، موضوع شروع سرد آیتمی و کاربری در ماتریس امتیازها است و با استفاده از خوشه‌بندی کاربران و آیتم‌ها منجر به بهبود پیش‌بینی و در نتیجه افزایش صحت پیشنهاددهی می‌شود.

منابع

[1] Katarya, R., Verma, O. P. (2016). A collaborative recommender system enhanced with particle swarm optimization technique, Multimedia Tools and Applications, 75(15), 9225-9239.

[2] Esmaeili, L., Nasiri, M., Minaei-Bidgoli, B., (2011). Personalizing group recommendation to social network users, in proceeding of Web Information systems and mining, 124-133.

[3] D. Chakrabarti, R. Kumar, A. Tomkins. (2006). in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, PA, USA. Evolutionary clustering. 554-560.

[4] C. Rana, S. K. Jain. (2014). Swarm and Evolutionary Computation. An evolutionary clustering algorithm based on temporal features for dynamic recommender systems.14. 21-30.

[5] N. Lathia, S. Hailes, L. Capra. (2009). Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, Boston, MA, USA. Temporal collaborative filtering with adaptive neighbourhoods. 796-797. 10.1145/1571941.1572133.

[6] Selvarajah K., Kobti Z., Kargar M. (2019). Applications of Evolutionary Computation. A Cultural Algorithm for Determining Similarity Values Between Users in Recommender Systems. 11454. 270-283.

[7] Imane Belkhadir, Elamine Didi Omar, Jaouad Boumhidi. (2019). Procedia Computer Science. An intelligent recommender system using social trust path for recommendations in web-based social networks. 148. 181-190.

[8] Jianrui Chen · Chunxia Zhao · Uliji · Lifang Chen . (2019). Complex & Intelligent Systems. Collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering.

[9] Mehdi Hosseinzadeh Aghdam. (2019). Physica A: Statistical Mechanics and its Applications. -۱ Context-aware recommender systems using hierarchical hidden Markov model. 518. 89-98.

[10] Ma Huifang, Jia Meihuizi, Zhang Di, Lin Xianghong. (2017). Information Sciences. Combining tag correlation and user social relation for microblog recommendation. 385. 325-337.

[11] Tomas Horvath, André C. P. L. F. de Carvalho. (2017). Natural Computing. Evolutionary computing in recommender systems: a review of recent research. 16. 441-462.

[12] Shanfeng Wang, Maoguo Gong, Haoliang Li, Junwei Yang, Yue Wu. (2017). Knowledge-Based Systems. Memetic algorithm based location and topic aware recommender system. 131. 125-134.

[DOI: 10.52547/jamlu.19.3.1]

[13] Kian R, Pradeep Kumar, Bharat Bhasker. (2020). Expert Systems with Applications. DNNRec: A Nover Deep Learning based Hybrid Recommender System. 144. 113054.
[14] Gediminas Adomavicius, Young Ok Kwon. (2008). in Proceedings of the 18th Workshop on Information Technology and Systems (WITS'08), Paris, France. Overcoming accuracy-diversity tradeoff in recommender systems: A variance-based approach.
[15] Swagatam Das, Ajith Abraham, Amit Konar. (2008). in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*. Automatic Clustering Using an Improved Differential Evolution Algorithm. 38. 218-237.

[DOI: 10.52547/jamlu.19.3.1]