7th International Conference on Web Research
May 19-20, 2021; Tehran, IRAN ICWR2021

Proposing a Meta-Heuristic Approach for the Long Tail Problem of Recommender Systems

Sayedeh Zahra Hosseini [1], Azadeh Mohammadi [2] *

[1] Bachelor Student of Information Technology Engineering, University of Isfahan, Isfahan, Iran
zahrahosseini.isf@gmail.com

[2] Assistant Professor, Computer Department, University of Isfahan, Isfahan, Iran
az.mohammadi@eng.ui.ac.ir

Abstract

Recommendation systems are a solution for providing appropriate suggestions to users and helping them in the decision-making process. In the most recommendation systems, the purpose is to offer items tailored to the user's interests based on the past scores. Focusing on the previous scores causes various problems including ignoring long tail items. Long tail items are those that are rated by a small number of users, therefor they are often not recommended by recommendation systems. This leads to the recommendation system bias towards offering previously popular items and ignoring the diversity and novelty of suggestions. To solve this problem, in this paper, a multi-objective optimization approach is used to increase the accuracy and chance of recommending long tail items. In the proposed method which is applied to the Movielens dataset, first, various groups of items are created by categorizing movies based on freshness (year of production). Then by applying the NSGAII algorithm, accuracy and diversity of suggested items are optimized. The results of applying the proposed method to the mentioned dataset indicate that the suggested method, while maintaining the accuracy of recommendation, has been able to increase the number of long tail items in the recommendation lists.

Keywords: Recommender System, Long Tail Items, Freshness, Multi Objective Optimization, Collaborative Filtering, NSGAII

7th International Conference on Web Research
May 19-20, 2021; Tehran, IRAN ICWR2021

ارائه یک روش فرا ابتکاری برای حل مشکل اقلام دنباله طولانی در سیستم‌های توصیه‌گر

سیده زهرا حسینی[۱]، آزاده محمدی*[۲]
[۱] کارشناسی مهندسی فناوری اطلاعات، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان
zahrahosseini.isf@gmail.com
[۲] استادیار، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان
az.mohammadi@eng.ui.ac.ir

چکیده

سیستم‌های توصیه‌گر راه‌حلی برای ارائه پیشنهادهای مناسب به کاربران و کمک به آن‌ها در فرایند تصمیم‌گیری هستند. در بیشتر سیستم‌های توصیه‌گر، هدف پیشنهاد اقلامی متناسب با علایق کاربر بر اساس امتیازدهی‌های گذشته می‌باشد. تمرکزِ صرف بر امتیازدهی-های پیشین، سبب مشکلات مختلفی از جمله نادیده گرفته شدن اقلام دنباله طولانی می‌گردد. اقلام دنباله طولانی، اقلامی هستند که توسط تعداد کمی از کاربران امتیازدهی شده‌اند و در نتیجه اغلب توسط سیستم‌های توصیه‌گر پیشنهاد نمی‌شوند. این امر سبب سوگیری سیستم توصیه به سمت پیشنهاد کالاهایی با محبوبیت بالای قبلی و نادیده گرفته شدن تنوع و تازگی در پیشنهادها می‌شود. برای حل این مشکل، در این مقاله از یک روش بهینه‌سازی چند هدفه به منظور افزایش دقت و افزایش شانسِ توصیه اقلام دنباله طولانی استفاده شده است. در روش پیشنهادی که بر روی مجموعه‌داده Movielens اعمال شده است ابتدا با دسته‌بندی فیلم‌ها بر اساس تازگی (سال تولید)، دسته‌های متنوعی از اقلام ایجاد می‌شوند. سپس با اعمال الگوریتم NSGAII، دو هدف افزایش دقت و تنوع در اقلامِ پیشنهادی، دنبال می‌شود. نتایج حاصل از اعمال روش پیشنهادی بر روی مجموعه‌داده ذکرشده بیانگر آن است که روش پیشنهادی در عین حفظ دقتِ توصیه توانسته است تعداد اقلام دنباله طولانی در لیست‌های توصیه را نیز افزایش بخشد.

کلمات کلیدی
سیستم توصیه‌گر، اقلام دنباله طولانی، تازگی، بهینه‌سازی چندهدفه، پالایش مشارکتی، ژنتیک مرتب‌سازی نامغلوب دو

۱- مقدمه

با افزایش حجم اطلاعات در اینترنت، علاقه کاربران به استفاده از اینترنت افزایش یافته است. مهم‌ترین چالش استفاده از اینترنت، پیچیده شدن فرایند تصمیم‌گیری و انتخاب برای کاربران است [۱]. سیستم‌های توصیه‌گر با جمع-آوری اطلاعات کاربران و اقلام و با استفاده از الگوریتم‌های مختلف، پیشنهادهای متناسبی به کاربران ارائه می‌دهند [۲]. سیستم‌های توصیه‌گر به چهار دسته کلی محتوا محور[۱]، پالایش مشارکتی[۲]، مبتنی بر دانش[۳] و روش-های ترکیبی[۴] تقسیم می‌شوند. در روش محتوا محور، تمرکز بر روی تجزیه و تحلیل ویژگی‌های اقلام برای پیش‌بینی است. در این روش با استفاده از ویژگی‌های استخراج شده از محتوای اقلامی که کاربر در گذشته ارزیابی کرده است، لیست اقلام به وی پیشنهاد می‌شود [۳]. در روش پالایش مشارکتی، با استفاده از امتیازدهی کاربران به اقلام، شباهت بین کاربران یا اقلام محاسبه و سپس بر اساس این شباهت، توصیه به کاربر انجام می‌شود [۴]. روش پالایش مشارکتی خود به سه دسته مبتنی بر کاربر[۵] [۵]، مبتنی بر اقلام[۶] [۶] و فاکتورگیری ماتریس[۷] [۷] تقسیم می‌شود. در روش مبتنی بر دانش، سیستم پیشنهادهای خود را بر پایه تفسیر و استنباط خود از سلایق و نیازهای کاربر ارائه می‌دهد. طبیعی است که برای پیاده‌سازی چنین سیستم‌هایی نیاز به یک بستر و ساختار مبتنی بر دانش وجود دارد [۸]. در روش ترکیبی، از روش‌های

گفته شده به طور همزمان استفاده می‌شود تا محدودیت‌های هر کدام با دیگری پوشش داده شود [۹].

اکثر سیستم‌های توصیه‌گر به دنبال پیشنهادهایی هستند که با احتمال بالا موردپسند کاربران باشند، درنتیجه در لیست پیشنهادی، از اقلامی استفاده می‌کنند که بیشترِ کاربران آن را پسندیده باشند [۱۰]. این امر سبب افزایش دقت توصیه می‌شود اما از طرف دیگر باعث می‌شود اقلامی که تا کنون چندان محبوب نبوده‌اند در پیشنهادهای بعدی نیز لحاظ نشده و به طور مکرر نادیده گرفته شوند، در حالی‌که ممکن است این اقلام در صورت توصیه شدن، مورد استقبال کاربران قرار گیرند. از این اقلام، تحت عنوان اقلام دنباله طولانی[۸] یاد می‌شود [۱۱–۱۳]. با توجه به آنچه ذکر شد یکی از چالش‌های سیستم‌های توصیه‌گر، عدم توصیه اقلام دنباله طولانی به دلیل تمرکز بر افزایش دقت است. طبق مطالعات پیشین، بین افزایش دقت و افزایش اقلام دنباله طولانی در لیست توصیه، رابطه عکس وجود دارد [۱۴]. بنابرین موازنه بین این دو هدف یک امر مهم در سیستم‌های توصیه‌گر است [۱۵].

همانطور که گفته شد تمرکز بر روی دقت لیست‌های توصیه، سبب افزایش اقلام دنباله طولانی می‌گردد. بنابراین استفاده از معیارهای دیگر مانند پوشش[۹] و تنوع[۱۰] در کنار دقت می‌تواند این مشکل را تا حدی بهبود بخشد [۱۶]. یکی از روش‌های مورد استفاده برای افزایش تنوع، رویکرد مبتنی بر نشست است [۱۲, ۱۷] که بر اساس رفتار کاربر در یک نشست سعی می‌کنند کالاهای محبوب و غیرمحبوب را تشخیص داده و ترکیبی از آن‌ها را به کاربر پیشنهاد دهند. از دیگر روش‌های افزایش تنوع در لیست‌های توصیه، استفاده از تنوع شخصی‌سازی شده است [۱۸, ۱۹]. تنوع شخصی‌سازی شده می‌تواند در کنار معیارهای دیگر برای افزایش تنوع در نظر گرفته شود. برخی پژوهشگران معیار محبوبیت را برای ایجاد تنوع انتخاب کرده‌اند [۱۶, ۲۰–۲۲]. این معیار می‌تواند برای دسته‌بندی اقلام یا به عنوان معیار ارزیابی در کنار معیارهای دیگر مانند دقت مورد استفاده قرار گیرد. در برخی پژوهش‌ها جذابیت اقلام به عنوان معیار ارزیابی برای افزایش اقلام دنباله طولانی استفاده شده است [۲۳]. رویکردهای گفته شده به صورت تک هدفه و یا به همراه اهداف دیگر مانند دقت به صورت بهینه‌سازی چند هدفه مورد استفاده قرار گرفته‌اند.

با بررسی‌های انجام‌شده، مشاهده گردید که در هیچ کدام از پژوهش‌های گذشته، معیار تازگی[۱۱] برای دسته‌بندی اقلام در نظر گرفته نشده‌است. بهینه‌سازی چند هدفه می‌تواند مبتنی بر دو هدف ایجاد تنوع از طریق تازگی و دقت برای تولید لیست‌های بهینه مورد استفاده قرار گیرد. دسته‌بندی اقلام بر اساس میزان تازگی می‌تواند سبب کاهش مشکل عدم توصیه اقلام طولانی گردد، زیرا بسیاری از اقلام در دنباله طولانی، اقلامی هستند که به دلیل تازه وارد بودنِ خود امتیازی از کاربران دریافت نکرده‌اند و در نتیجه در مقابل اقلامی با سابقه طولانی‌تر، شانس حضور در لیست‌های توصیه را از دست می‌دهند. بر این اساس در این مقاله، با تعریف مفهوم تازگی، ابتدا اقلام بر اساس میزان تازگی (زمان ورود به سیستم) دسته‌بندی شده و سپس از یک الگوریتم بهینه‌سازی چندهدفه به منظور افزایش دقت و افزایش تنوعِ انتخاب از دسته‌های زمانی مختلف استفاده می‌شود. در واقع روش پیشنهادی شامل سه گام اصلی می‌باشد. در گام اول، اقلام مشابه بر اساس امتیازدهی‌های گذشته و همچنین اطلاعات جانبی تعیین می‌شوند. در گام دوم اقلام بر اساس میزان تازگی دسته‌بندی می‌شوند و در گام سوم با استفاده از الگوریتم بهینه‌سازی چندهدفه NSGAII تلاش می‌شود تا با برقراری موازنه بین دو هدف دقت و تنوع تازگی، مشکل عدم توصیه اقلام دنباله طولانی تخفیف پیدا نماید.

در ادامه مقاله، ابتدا در بخش ۲ مروری بر پژوهش‌های انجام شده در حوزه مشکل اقلام طولانی در سیستم‌های توصیه شده است. سپس مفاهیم اولیه لازم در بخش ۳ به اختصار بیان شده‌اند. در بخش ۴ روش پیشنهادی و نحوه عملکرد آن توضیح داده شده و نتایج حاصل از اعمال روش پیشنهادی بر روی مجموعه داده‌های واقعی در بخش ۵ نمایش داده شده است. در نهایت در بخش ۶ نتیجه‌گیری و کارهای آینده بیان شده است.

۲- پیشینه پژوهش

همانطور که ذکر شد، در برخی پژوهش‌ها، پژوهشگران به دنبال افزایش تنوع در لیست‌های توصیه به منظور بهبود مشکل اقلام دنباله طولانی هستند. یکی از روش‌ها، استفاده از اطلاعات فعلی کاربران است. سی یی لو و همکارانش یک روش مبتنی بر نشست (جلسه) برای حل مشکل اقلام دنباله طولانی ارائه کرده‌اند [۱۲]. در این پژوهش، ابتدا بر اساس فرکانس کلیک کاربر، اقلام به دو دسته محبوب و غیر محبوب تقسیم شده و سپس با استفاده از یک شبکه عصبی، توصیه‌هایی در هر گروه از اقلام به کاربر پیشنهاد می‌شود. روش گفته شده ضمن حفظ دقت، مشکل اقلام دنباله طولانی را تا حدی بهبود می‌بخشد. قرهیقهی و همکاران نیز یک رویکرد مبتنی بر نشست برای موازنه بین دقت و تنوع در لیست توصیه ارائه کرده‌اند [۱۷]. در این روش با استفاده از اطلاعات کاربران و شباهت بین آن‌ها در نشست فعلی، امتیاز اقلام برای کاربران پیش‌بینی می‌شود. سپس لیست‌های توصیه با بالاترین امتیاز توصیه می‌شوند. نتایج این روش نشان می‌دهد که ضمن افت نسبی دقت، تنوع در لیست‌های توصیه افزایش یافته و مشکل اقلام دنباله طولانی تا حدی بهبود یافته‌است.

از دیگر روش‌های افزایش تنوع، استفاده از رویکرد تنوع شخصی‌سازی شده است. عبدلله‌پوری و همکاران یک رویکرد مبتنی بر تنوع شخصی‌سازی شده برای افزایش تنوع در لیست‌های توصیه ارائه کرده‌اند [۱۸]. در این پژوهش از تنوع شخصی‌سازی شده برای رتبه‌بندی اقلام استفاده شده است.

از روش‌های دیگر برای بهبود تنوع، استفاده از محبوبیت اقلام است. یجین کین و همکاران یک رویکرد مبتنی بر شبکه عصبی برای تهیه لیست توصیه با حفظ دقت و افزایش تنوع ارائه کرده‌اند [۲۲]. در این پژوهش ابتدا اقلام بر اساس محبوبیت، خوشه‌بندی می‌شوند. در نهایت در لیست توصیه، اقلام محبوب با نزدیک‌ترین قلم در خوشه خود جایگزین می‌شوند. نتایج نشان داده‌اند که این روش ضمن حفظ نسبی دقت، تنوع را افزایش داده و تا حدی مشکل اقلام دنباله طولانی را رفع می‌کند.

در برخی از پژوهش‌های گذشته از مفهوم جذابیت برای ایجاد تنوع استفاده شده است. جورج دیاز و همکاران یک روش مبتنی بر جذابیت اقلام برای افزایش تنوع در لیست‌های توصیه ارائه کرده‌اند [۲۳]. هر لیست توصیه که شامل اقلام دنباله طولانی باشد پاداشی دریافت می‌کند. این پژوهش به دنبال افزایش دقت و افزایش پاداش در لیست‌های توصیه به منظور افزایش جذابیت لیست‌های توصیه است. با افزایش جذابیت در لیست‌های توصیه تنوع نیز افزایش می‌یابد. در [۲۴] به هر یک از اقلام، یک جذابیت اولیه[۱۲] نسبت داده شده که این مقدار با گذشت زمان کاهش می‌یابد. این امر سبب می‌شود احتمال توصیه آن قلم به تدریج کاهش یافته و از لیست توصیه حذف گردد. در نتیجه کالا پس از مدتی شانس انتخاب شدن را از دست خواهد داد. مشکل این روش عدم تطابق آن با دنیای واقعی است زیرا به عنوان مثال در یک سیستم پیشنهاد فیلم، هر فیلم همواره باید شانسی برای پیشنهاد شدن داشته باشد و عدم پیشنهاد یک فیلم تا زمانی مشخص نباید به منزله حذف آن باشد.

در تمامی روش‌های فوق‌الذکر، بهینه‌سازی به صورت تک هدفه و با ترکیب دو معیار دقت و تنوع در قالب یک هدف صورت پذیرفته است. در حالی که دو معیار تنوع و دقت، دو معیار ناسازگار بوده و افزایش تنوع می‌تواند سبب کاهش دقت شود. در نتیجه در این مسئله، استفاده از روش‌های بهینه‌سازی چند هدفه[۱۳] می‌تواند رویکرد مناسب‌تری باشد [۲۵]. در این راستا، جیاونا پانگ و همکارانش یک روش اندازه‌گیری شباهت وزنی بر اساس الگوریتم ژنتیک مرتب سازی غیر مغلوب[۱۴] (NSGA-II) ارائه کرده‌اند. در این کار شبیه‌ترین کاربران به یکدیگر پیدا می‌شوند و سپس لیست توصیه بر اساس این شباهت ارائه می‌گردد [۱۴]. در برخی پژوهش‌ها، اقلام به دو دسته محبوب و دنباله طولانی تقسیم شده و موازنه‌ای بین این دو دسته ایجاد می‌گردد [۲۲]. شانفنگ وانگ و همکاران یک روش بهینه‌سازی مبتنی بر دقت و اقلام دنباله طولانی براساس محبوبیت اقلام ارائه کرده‌اند. در این پژوهش اقلام بر اساس محبوبیت تقسیم و سپس با استفاده از بهینه‌سازی چند هدفه، بین اقلام دنباله طولانی و دقت، موازنه ایجاد شده است. برای ارزیابی محبوبیت اقلام از واریانس و میانگین رتبه‌های کاربر استفاده شده است که نتایج این روش نشان می‌دهد بیشتر لیست‌های توصیه شده کیفیت مطلوبی دارند [۲۶]. الهه ملک‌زاده و همکاران با استفاده از بهینه‌سازی چند هدفه، تنوع شخصی‌سازی شده و در نظر گرفتن همزمان دقت و اقلام دنباله طولانی، مشکل اقلام دنباله طولانی را تا حدی بهبود داده‌اند [۱۹]. لیانگ ژانگ و همکاران یک رویکرد بهینه‌سازی چند هدفه دو مرحله‌ای برای حل مشکل اقلام دنباله طولانی و ایجاد تنوع در لیست‌های توصیه ارائه کرده‌اند [۱۶]. در این روش، ابتدا بهترین همسایه‌ها برای هر کاربر پیدا شده و سپس با استفاده از روابط بین کاربران و پالایش مشارکتی، یک رتبه‌بندی جدید برای اقلام در نظر گرفته شده است. در نهایت با در نظر گرفتن دقت و محبوبیت، لیست‌های توصیه ارزیابی شده‌اند. روش ارائه شده با افت دقت، تنوع و محبوبیت را افزایش داده است. عبدلله‌پوری و همکاران یک روش بهینه‌سازی چند هدفه ارئه کرده‌اند که هدف آن متعادل‌ساختن تعداد اقلام محبوب و اقلام دنباله طولانی با توزیع یکنواخت این دو گروه اقلام در لیست‌های توصیه است [۲۰].

با توجه به بررسی‌های انجام شده، در هیچکدام از روش‌های چندهدفه ارائه شده برای سیستم‌های توصیه‌گر، بهینه‌سازی همزمان دو معیار دقت و تنوع بر اساس تازگی در نظر گرفته نشده است. از این رو، در این مقاله روشی جهت دسته‌بندی زمانی اقلام ارائه شده و سپس با توجه به دسته‌بندی انجام شده و با به‌کارگیری بهینه سازی چندهدفه تلاش می‌شود علاوه بر افزایش دقت، تنوع انتخاب از دسته‌های مختلف زمانی نیز افزایش یابد. جزییات روش پیشنهادی در بخش ۴ ارائه شده است. پیش از آن در بخش ۳ به بیان مفاهیم پایه پرداخته شده است.

۳- مفاهیم پایه

در این بخش مفاهیم اولیه موردنیاز به اختصار بیان می‌شوند:

۳-۱- سیستم‌های توصیه‌گر

در سیستم‌های توصیه‌گر مبتنی بر پالایش مشارکتی، مجموعه‌ای از کاربران و اقلام و یک ماتریس امتیاز (کاربر–اقلام) وجود دارد. امتیاز داده شده توسط کاربر u به آیتم i که با r_{ui} نمایش داده می‌شود میزان رضایت کاربر u از آیتم i را نشان می‌دهد. هدف از سیستم توصیه آن است که با توجه به امتیازهای داده شده توسط کاربران به اقلام مختلف، امتیاز یک کاربر خاص به سایر اقلام، پیش‌بینی شده و اقلام جدیدی به وی پیشنهاد گردد. در برخی سیستم‌های توصیه پس از تخمین امتیاز اقلام، یک آیتم با بالاترین امتیاز تخمینی به کاربر پیشنهاد می‌شود و در برخی دیگر یک لیست شامل k آیتم با بالاترین امتیاز تخمینی به کاربر پیشنهاد می شود که به این حالت top k items گفته می‌شود [۲۶]. در این مقاله از رویکرد دوم استفاده شده است.

۲-۳- اقلام دنباله طولانی

در شکل (۱) امتیازهای داده شده به اقلام در مجموعه داده Movielens نمایش داده شده است. محور عمودی در این شکل، تعداد دفعاتی است که اقلام (فیلم‌ها) امتیاز گرفته‌اند و محور افقی تعداد اقلامی را نشان می‌دهد که به آن تعداد، امتیاز گرفته‌اند. با توجه به این نمودار مشخص است تعداد بسیار کمی از اقلام در دفعات بالا توسط کاربران امتیازدهی شده‌اند. این اقلام فیلم-های محبوب را نشان می‌دهند. در مقابل تعداد زیادی از اقلام، تعداد دفعات کمی امتیاز گرفته اند که تشکیل یک دنباله طولانی می‌دهند. این اقلام در سیستم‌های توصیه مبتنی بر امتیازدهیِ قبلی کاربران، اغلب نادیده گرفته می-شوند در حالی‌که پیشنهاد این دسته اقلام می‌تواند برای کاربران تنوع و علاقه ایجاد نموده و رضایت آن‌ها را افزایش دهد [۱۵].

۳-۳- بهینه سازی چند هدفه

بهینه‌سازی چند هدفه، مسائلی را که در آن‌ها نیاز به بهینه‌سازی همزمان بیش از یک تابع هدف است را حل می‌کند [۲۷]. این نوع الگوریتم‌ها به جای ارائه یک پاسخ، مجموعه‌ای از پاسخ‌های بهینه را فراهم می‌کنند. این الگوریتم‌ها از مفهومی به نام بهینگی پارتو استفاده می‌کنند. بهینگی پارتو شامل پاسخ‌هایی است که هیچ کدام بر دیگری برتری ندارند و همزمان چندین هدف را بهینه می‌کنند. نقطه y در فضای مسئله یک بهینه پارتو است در صورتی که هیچ نقطه دیگری مانند x وجود نداشته باشد که برای همه نقاط در فرمول (۱) و حداقل یک نقطه در فرمول (۲) صدق کند [۲۶, ۲۸].

$$f_i(x) \lesssim f_i(y) \quad i = ۰,۱,...,n \qquad (۱)$$

$$f_i(x) < f_i(y) \quad i = ۰,۱,...,n \qquad (۲)$$

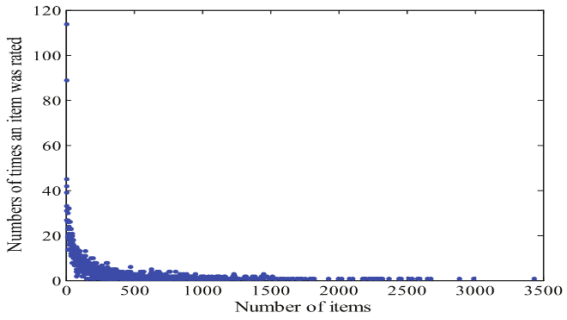
یکی از مطرح‌ترین الگوریتم‌های چندهدفه، الگوریتم ژنتیک مرتب‌سازی نامغلوب دو است. این الگوریتم توسط دپ و همکاران و بر پایه الگوریتم ژنتیک، ارائه گردید [۲۹, ۳۰]. در این الگوریتم، ابتدا یک جمعیت تصادفی از کروموزوم‌ها ایجاد شده، سپس عملگرهای انتخاب، تقاطع و جهش روی جمعیت اعمال شده و پس از آن بر اساس مرتب‌سازی نامغلوب و فاصله ازدحامی، جمعیت جدید انتخاب می‌شود. این فرایند تا رسیدن به شرط توقف، تکرار می‌شود [۲۹, ۳۰]. در بخش ۴ نحوه به‌کارگیری الگوریتم NSGAII در روش پیشنهادی و جزییات آن به طور کامل توضیح داده می‌شود.

٤- روش پیشنهادی

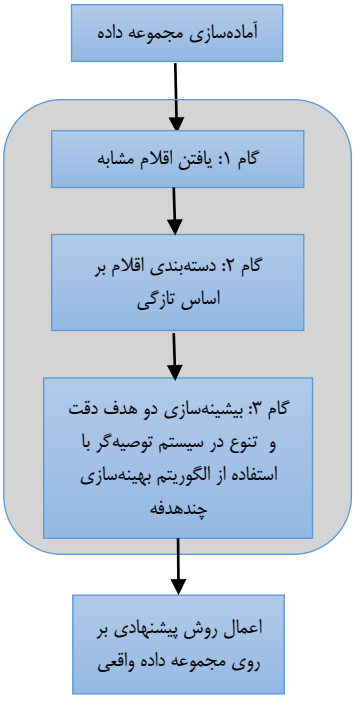
روش پیشنهادی در این مقاله از سه بخش اصلی تشکیل شده است. این مراحل که در شکل (۲) با گام ۱ تا ۳ نمایش داده شده‌اند، در بخش‌های ۴-۱ تا ۴-۳ به تفصیل بررسی می‌گردند:

٤-۱- گام اول: یافتن اقلام مشابه

از آنجایی که هدف اصلی در سیستم‌های توصیه‌گر، پیشنهاد اقلامی متناسب با سلیقه کاربران می‌باشد در گام اول از روش پیشنهادی، شباهت اقلام با یکدیگر پیدا می‌شود. یافتن اقلام مشابه کمک می‌کند تا مواردی مشابه با



شکل (۱) اقلام دنباله طولانی در مجموعه داده Movielens [۲۶]



شکل (۲): ساختار کلی روش پیشنهادی

علایق قبلی کاربران به آن‌ها نمایش داده شود. در این مقاله، مشابهت اقلام به دو شکل محاسبه می‌گردد:

در حالت اول، مشابهت دو آیتم i و j با توجه به امتیازهای داده شده قبلی توسط کاربران و بر اساس فرمول پیرسون مطابق با فرمول (۳) محاسبه می‌گردد [۳۱, ۳۲]:

$$similarity_r(i.j) = \frac{\sum_{u \in U}(r_{u.i} - \bar{r}_u)(r_{u.j} - \bar{r}_u)}{\sqrt{\sum_{u \in U}(r_{u.i} - \bar{r}_u)^2}\sqrt{\sum_{u \in U}(r_{u.j} - \bar{r}_u)^2}} \quad (۳)$$

در فرمول (۳)، $r_{u.i}$ امتیاز داده شده توسط کاربر u به مورد i است و $\bar{r}_u$ میانگین امتیازات داده شده توسط کاربر u است. بدین ترتیب، هر چه دو آیتم در گذشته توسط کاربران امتیازهای نزدیک‌تری گرفته باشند، مشابهت بالاتری می‌یابند.

از طرف دیگر، مشابهت اقلام را می‌توان مستقل از امتیازهای داده شده و بر اساس اطلاعات جانبی مانند ویژگی‌های اقلام محاسبه نمود. به‌عنوان مثال، یک فیلم با توجه به موضوع خود ممکن است در ژانرهای مختلف مانند ژانر کمدی و علمی تخیلی قرار گیرد. مجموعه ژانرها، بردار ویژگی هر فیلم را تشکیل می‌دهند. در این بردار، مولفه‌های متناظر با ژانرهایی که فیلم به آن‌ها تعلق دارد برابر ۱ و سایر مولفه‌ها برابر ۰ خواهند بود. بنابراین مشابهت دو فیلم مستقل از نحوه امتیازدهی آن‌ها توسط کاربران، می‌تواند با توجه به نوع ژانرها محاسبه گردد. برای محاسبه مشابهت دو آیتم بر اساس ویژگی آن‌ها، از شباهت کسینوسی مطابق با فرمول (۴) استفاده می‌شود [۱۹].

$$similarity_f(i.j) = \frac{\vec{D}_i \cdot \vec{D}_j}{|\vec{D}_i||\vec{D}_j|} \quad (۴)$$

در فرمول (۴)، D_i و D_j بردار ویژگی اقلام i و j هستند. صورت کسر، ضرب داخلی بردارها و مخرج کسر، ضرب اندازه بردارها را نشان می‌دهد.

پس از محاسبه شباهت اقلام بر اساس دو روش گفته شده، شباهت کلی از طریق فرمول (۵) محاسبه می‌شود:

$$similarity_t = \alpha\, similarity_f + (1-\alpha) similarity_r \quad (۵)$$

در فرمول (۵)، ضریب α وزن تاثیر هر نوع شباهت در محاسبه شباهت کلی را نشان می‌دهد. در این مقاله هر دو نوع شباهت دارای اهمیت یکسانی درنظر گرفته شده‌اند و در نتیجه مقدار α در فرمول (۵) برابر ۰.۵ در نظر گرفته شده است.

پس از محاسبه شباهت بین هر دو قلم، امتیاز کاربر u به قلم i، بر اساس امتیاز داده شده کاربر به اقلام مشابه و بر اساس فرمول (۶) تخمین زده می‌شود:

$$\hat{r}_{u.i} = \sum_{all\ simillar\ items}(S_{i.n} * r_{u.n}) / \sum_{all\ similar\ items} |S_{i.n}| \quad (۶)$$

در این فرمول، $\hat{r}_{u.i}$ تخمینِ امتیاز داده شده توسط کاربر u به قلم i است. $S_{i.n}$ شباهت بین قلم i و قلم n از مجموعه اقلام مشابه است و $r_{u.n}$ نیز امتیاز داده شده توسط کاربر u به قلم مشابه n می‌باشد [۱۹].

۴-۲- گام دوم: دسته‌بندی اقلام بر اساس تازگی

همان‌طور که پیش‌تر ذکر شد، پیشنهاد اقلام بر اساس امتیازهای گذشته سبب سوگیری سیستم توصیه‌گر به سمت پیشنهاد کالاهای محبوب و نادیده گرفتن اقلام دنباله طولانی می‌شود. به منظور حل مشکل اقلام دنباله طولانی و ایجاد تنوع در اقلام پیشنهادی، در گام دوم از روش ارائه شده در این مقاله، اقلام بر اساس میزان تازگی دسته‌بندی می‌شوند. دسته‌بندی بر اساس میزان تازگی به اقلام جدید که هنوز امتیاز زیادی کسب نکرده‌اند نیز فرصت انتخاب می‌دهد و از طرف دیگر، انتخاب اقلام از دسته‌های مختلف سبب افزایش تنوع و کاهش مشکل اقلام دنباله طولانی می‌شود.

با توجه به آنکه مجموعه داده مورداستفاده در این مقاله مربوط به اطلاعات تعدادی فیلم در فاصله سال‌های ۱۹۸۰ تا ۲۰۱۹ می‌باشد، برای دسته‌بندی بر حسب تازگی، سال تولید فیلم در نظر گرفته شده است. این دسته‌بندی به دو شکل صورت گرفته است: در حالت اول فیلم‌های هر سال در یک دسته قرار داده شده‌اند و در حالت دوم فیلم‌ها بر حسب دهه تقسیم‌بندی شده‌اند. نتایج حاصل از هر حالت در بخش ۵ ارائه شده است.

لازم به ذکر است روش پیشنهادی به‌طور مشابه می‌تواند بر روی هر مجموعه‌داده دیگری اعمال شود. به‌عنوان مثال در پیشنهاد کالا، زمان وارد شدن کالا به بازار می‌تواند مبنای دسته‌بندی قرار گیرد.

۴-۳- گام سوم: بهینه‌سازی چندهدفه توصیه اقلام

یافتن اقلام مشابه با سلایق قبلی کاربر، دقت پیشنهادهای ارائه شده توسط سیستمهای توصیه‌گر را افزایش می‌دهد. از طرف دیگر دسته‌بندی اقلام بر اساس میزان تازگی و افزایش انتخاب از دسته‌های مختلف زمانی، سبب افزایش تنوع در کالاهای انتخابی و کاهش مشکل اقلام دنباله طولانی در سیستم‌های توصیه‌گر می‌گردد. به منظور برآورده کردن همزمان دو هدفِ افزایش دقت و افزایشِ تنوع انتخاب، در گام سوم از روش پیشنهادی از روش

بهینه‌سازی چندهدفه NSGAII استفاده شده است. در ادامه، نحوه استفاده از این الگوریتم در مسئله بهینه‌سازی چندهدفهٔ توصیه اقلام بیان می‌شود.

٤-٣-١- جمعیت اولیه

جمعیت اولیه شامل مجموعه‌ای از کروموزوم‌ها است. هر کروموزوم یک جواب ممکن یعنی یک لیست توصیه را مشخص می‌کند. از آنجایی که در این مقاله توصیه به روش top k items بوده و لیست اقلام توصیه برابر ۱۰ در نظر گرفته شده است، هر کروموزوم از ۱۰ ژن تشکیل می‌شود. هر ژن در کروموزوم متناظر با شماره یک قلم می‌باشد. شکل (۳) یک نمونه کروموزوم را نشان می‌دهد.

٤-٣-٢- توابع برازندگی

همان‌طور که در بخش ۴-۳ ذکر شد، در این مقاله، بیشینه‌سازی هر دو هدف دقت و تنوع در انتخاب از دسته‌های زمانی مدنظر قرار گرفته است. در نتیجه دو تابع هدف دقت و تنوع در روش بهینه‌سازی چندهدفه مورد استفاده قرار می‌گیرند.

برای بهینه‌سازی دقت در لیست‌های توصیه، از مجموع امتیازات پیش‌بینی شده برای هر کاربر استفاده می‌شود [۲۶]. فرمول (۷) تابع برازندگی برای دقت را نشان می‌دهد.

$$f_1 = \sum_{i=1}^{k} \hat{r}_{u.i} \qquad (۷)$$

در این فرمول، k طول لیست توصیه و $\hat{r}_{u.i}$ تخمین امتیاز کاربر u به آیتم i در لیست می‌باشد.

برای بالابردن تنوع اقلام در لیست توصیه، فرمول (۸) به عنوان تابع هدف دوم تعریف می‌شود:

$$f_2 = c_k \qquad (۸)$$

در فرمول (۸)، c_K تعداد دسته‌های مختلف در لیست توصیه شده به هر کاربر را نشان می‌دهد. به عنوان مثال، در دسته‌بندی بر اساس سال در یک لیست توصیه به طول ۱۰، مقدار f_2 می‌تواند بین ۱ تا ۱۰ باشد، مقدار ۱ بدان معنی است که همه اقلام لیست توصیه از یک سال انتخاب شده‌اند و مقدار ۱۰ یعنی اقلام از ۱۰ سال متفاوت توصیه شده‌اند.

۲۵	۳۷	۵۶	۶۱	۲	۱۰۵	۲۸	۶۹	۳۴۳	۱۱

شکل (۳) نمایش کروموزوم

٤-٣-٣- نحوه اعمال الگوریتم NSGAII

در این بخش عملگرها و پارامترهای مورداستفاده در این الگوریتم توضیح داده می‌شوند.

- عملگر انتخاب: این عملگر کروموزوم‌هایی که از لحاظ شایستگی بهتر هستند را برای ادامه نسل انتخاب و به نسل بعدی اضافه می‌کند. این عملگر به این معنی است که کروموزوم‌های بهتر باقی مانده و ضعیف‌ها می‌میرند. در این مقاله از روش یکنواخت تصادفی برای انتخاب کروموزوم-ها استفاده شده‌است [۱].
- عملگر تقاطع: از این عملگر برای تولید فرزندان جدید استفاده می‌شود. در این مقاله، از ترکیب تک نقطه‌ای و نرخ تقاطع ۰.۹ استفاده شده است. بدین ترتیب که ابتدا یک عدد تصادفی از بازه ای برابر با طول کروموزوم به عنوان نقطه تقاطع انتخاب شده و سپس دو کروموزوم به عنوان والد با هم ترکیب و دو فرزند تولید می‌شوند [۲۶].
- عملگر جهش: از عملگر جهش برای ایجاد تنوع در نسل استفاده می‌شود. . با توجه به نرخ جهش که در این مقاله ۰.۱ در نظر گرفته شده است، یک ژن به صورت تصادفی انتخاب می‌شود و مقدار یک ژن آن تغییر می‌کند. در صورت عدم وجود شماره جدید در لیست توصیه، این شماره به جای شماره قلم قبلی قرار می‌گیرد.
- مرتب‌سازی نامغلوب: از مرتب‌سازی نامغلوب برای محاسبه جبهه‌های پارتو استفاده می‌شود. روش کار به شرح زیر است.

۱. به ازای هر عضو در جمعیت، یک مجموعه S شامل اعضایی که توسط آن فرد مغلوب می‌شوند، در نظر گرفته می‌شود.

۲. به ازای هر عضو در جمعیت یک شمارنده c در نظر گرفته می‌شود. این شمارنده تعداد اعضایی که آن عضو را مغلوب می‌کنند نشان می‌دهد.

۳. به ازای هر عضو مانند p در صورتی که p، عضو دیگری مانند p' را مغلوب کند، p' به مجموعه S از p اضافه می‌شود در غیر این صورت یکی به شمارنده p اضافه می‌گردد.

۴. در نهایت برای جبهه فعلی اعضایی که شمارنده صفر دارند، به آن اضافه می‌شوند و از شمارنده تمام اعضایی که در مجموعه مغلوب آن اعضا بوده‌اند یک واحد کم می‌شود؛ یعنی تاثیر غلبه اعضای جبهه فعلی از آن‌ها برداشته می‌شود. برای ساخت جبهه بعدی، اعضایی که بدون تاثیر غلبه اعضای جبهه فعلی هستند و شمارنده آن‌ها صفر است، به جبهه بعدی اضافه می‌شوند و این فرایند برای ساخت تمامی جبهه‌ها تکرار می‌شود تا زمانی که عضوی بدون جبهه باقی نمانده باشد.

- فاصله ازدحامی: رتبه‌بندی کروموزوم‌ها توسط فاصله ازدحامی و مرتب-سازی نامغلوب انجام می‌شود. فاصله ازدحامی تفاوت هر عضو در هر جبهه را مشخص می‌کند. مراحل کار به صورت زیر است:

۱. ابتدا برای هر عضو در هر جبهه فاصله صفر در نظر گرفته می‌شود.

۲. اعضا در هر جبهه بر اساس برازندگی توابع هدف مرتب می‌شوند.

۳. برای اعضا در مرز جبهه‌ها فاصله ازدحامی بی‌نهایت در نظر گرفته می‌شود.

۴. برای سایر اعضا در جبهه‌ها فاصله ازدحامی از فرمول (۹) محاسبه می‌شود. در این فرمول n تعداد توابع هدف مورد ارزیابی است و d_i^j در هر تابع هدف از طریق فرمول (۱۰) بدست می‌آید. در فرمول (۱۰) $f_{j=i-1}$ برازندگی مورد قبل از i و $f_{j=i+1}$ برازندگی مورد بعد از i است. $f_{j\,max}$ و $f_{j\,min}$ به ترتیب مقدار بیشینه و کمینه برازندگی در آن تابع است [۳۰].

$$d_i = \sum_{j=1}^{n} d_i^j \qquad (۹)$$

$$d_i^j = |f_{j=i-1} - f_{j=i+1}| / f_{j\,max} - f_{j\,min} \qquad (۱۰)$$

۵- نتایج

در این بخش، نتایج حاصل از اعمال روش پیشنهادی بر روی مجموعه داده واقعی ارائه و تحلیل شده است.

۵-۱- مجموعه داده‌ها

برای ارزیابی روش پیشنهادی در این مقاله، از مجموعه داده‌های Movielens استفاده شده است. این مجموعه‌داده شامل ۲۷۷۰ فیلم و ۲۴۳۷ کاربر و ۶۱۰۴۰۲ امتیاز داده شده توسط کاربران به فیلم‌ها است. امتیاز فیلم‌ها عددی بین ۰ تا ۵ است [۳۳].

فیلم‌ها از سال‌های ۱۹۸۰ تا ۲۰۱۹ انتخاب شده‌اند. همان‌طور که در بخش ۴-۲ ذکر گردید برای دسته‌بندی فیلم‌ها بر اساس تازگی، دو حالت بررسی شده است. در حالت اول فیلم‌ها بر حسب سال دسته بندی شده‌اند که به این ترتیب ۴۰ دسته مختلف شکل می‌گیرد و یکبار نیز فیلم‌ها بر حسب دهه تقسیم شده-اند که در این حالت ۴ دسته تشکیل می‌شود. ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد داده‌ها برای آزمایش مورد استفاده قرار گرفته‌اند.

۵-۲- معیارهای ارزیابی

برای ارزیابی روش پیشنهادی، باید معیارهایی برای سنجش دقت و تنوع در لیست توصیه به منظور افزایش تعداد اقلام دنباله طولانی استفاده کرد. برای سنجش دقت از فرمول (۱۱) استفاده می‌شود:

$$Precision = \frac{1}{m}\sum_{u\in U}\frac{N_{rs}}{K} \qquad (۱۱)$$

در فرمول (۱۱) K تعداد کل اقلام توصیه شده، m تعداد کل کاربران و N_{rs} تعداد اقلام مرتبط توصیه شده است. اقلام مرتبط اقلامی هستند که امتیازشان از میانگین امتیازات کاربر بیشتر باشند. هر چه مقدار دقت الگوریتم بیشتر باشد مطلوب‌تر است.

برای سنجش موفقیت الگوریتم در توصیه اقلام دنباله طولانی از فرمول (۱۲) استفاده می‌شود. در این فرمول di تعداد دفعاتی است که مورد i توسط کاربران امتیاز گرفته است. به بیان دیگر هر چه di بیشتر باشد نشان‌دهنده محبوبیت بیشتر قلم i است. همچنین m تعداد کل کاربران، k طول لیست توصیه و L_u لیست توصیه به کاربر u است [۱۹].

$$Popularity = \frac{1}{mk}\sum_{u=1}^{m}\sum_{i\in L_u} di \qquad (۱۲)$$

هر چه مقدار popularity (عمومیت) بیشتر باشد بدان معنی است که سیستم توصیه‌گر در توصیه‌های خود اقلام محبوب بیشتری را توصیه نموده است و در مقابل هر چه این مقدار کمتر باشد یعنی سیستم توصیه‌گر در عملکرد کلی خود توانسته است تعداد اقلام دنباله طولانی بیشتری را توصیه نماید.

۵-۳- ارزیابی نتایج

برای ارزیابی روش پیشنهادی، از آنجایی که الگوریتم NSGAII مجموعه‌ای از جواب‌ها را برمی‌گرداند از سه رویکرد زیر به منظور انتخاب لیست‌های توصیه به کاربران استفاده می‌شود [۲۶]:

رویکرد اول: در مجموعه جواب بدست‌آمده از الگوریتم NSGAII از میانگین جواب‌ها استفاده می‌شود (mean).

رویکرد دوم: در مجموعه جواب بدست آمده از الگوریتم NSGAII، کروموزومی که بیشترین دقت را دارد انتخاب شود (most-accuracy).

رویکرد سوم: در مجموعه جواب بدست آمده از الگوریتم NSGAII، کروموزومی که بیشترین تنوع دسته‌بندی در لیست را دارد انتخاب شود (most-diversity).

عملکرد روش پیشنهادی در هر یک از رویکردهای فوق در جداول (۱) و (۲) نمایش داده شده و با روش پالایش مشارکتی نیز مورد مقایسه قرار گرفته است.

معیارهای ارزیابی همان‌طور که در بخش ۵-۲ ذکر شد دقت و عمومیت می‌باشند. هر چه مقدار عمومیت بدست آمده کمتر باشد بدان معنی است که سیستم توصیه‌گر توانسته است در توصیه‌های خود اقلام دنباله طولانی بیشتری را توصیه نماید. میزان دقت بالا نیز بیانگر آن است که سیستم توانسته است با توجه به امتیازات قبلی کاربران، اقلامی متناسب با امتیازهای قبلی را توصیه نماید.

جدول (۱) نتایج بدست‌آمده در حالت دسته‌بندی بر اساس سال و جدول (۲) نتایج بدست‌آمده در حالت دسته‌بندی بر اساس دهه را نشان می‌دهند. همان‌طور که از نتایج حاصل قابل مشاهده است، در هر سه رویکرد از روش پیشنهادی، معیار عمومیت به شکل قابل ملاحظه‌ای نسبت به روش پالایش مشارکتی کاهش یافته است در حالیکه میزان دقت تفاوت چشمگیری نداشته است. طبق نتایج بدست‌آمده در دسته‌بندی بر اساس سال افت دقت بیشتری نسبت به دسته‌بندی بر اساس دهه وجود دارد اما تعداد اقلام طولانی به طور قابل توجهی بیشتر است که علت این امر افزایش تعداد دسته‌ها در دسته‌بندی بر اساس سال است.

جدول (۱) نتایج بدست آمده برای دسته‌بندی بر اساس سال روی مجموعه داده Movielens

معیار ارزیابی	روش			
	CF	mean	most-accuracy	most-diversity
دقت	۰.۹۹۹	۰.۸۸۵	۰.۹۵۷	۰.۸۷۸
عمومیت	۰.۰۵۲۵	۰.۰۲۳۰	۰.۰۳۵۱	۰.۰۰۳۷

جدول (۲) نتایج بدست آمده برای دسته‌بندی بر اساس دهه روی مجموعه داده Movielens

معیار ارزیابی	روش			
	CF	mean	most-accuracy	most-diversity
دقت	۰.۹۹۹	۰.۹۱۱	۰.۹۶۲	۰.۸۷۱
عمومیت	۰.۰۵۲۵	۰.۰۲۳۹	۰.۰۲۴۹	۰.۰۲۴۰

این نتایج نشان می‌دهد دسته‌بندی اقلام بر اساس تازگی و بهینه‌سازی همزمان دو هدف دقت و تنوعِ انتخاب از دسته‌های زمانی توانسته است مشکل عدم توصیه اقلام طولانی در سیستم‌های توصیه‌گر را تخفیف بخشد.

٦- نتیجه‌گیری

در این مقاله، برای حل مشکل اقلام دنباله طولانی در سیستم‌های توصیه‌گر، یک روش بهینه‌سازی چندهدفه مبتنی بر دو هدف افزایش دقت و افزایش تنوع بر اساس تازگی ارائه شده است. در روش پیشنهادی، ابتدا اقلام بر اساس تازگی دسته‌بندی می‌شوند. در مجموعه‌داده مورد استفاده که مرتبط با اطلاعات فیلم‌ها در بازه زمانی ۱۹۸۰ تا ۲۰۱۹ میلادی است دسته بندی در دو حالت تقسیم بر اساس دهه و تقسیم بر حسب سال صورت گرفته است. پس از دسته‌بندی اقلام بر اساس میزان تازگی، مسئله پیشنهاد اقلام، به فرم یک مساله بهینه‌سازی چندهدفه با هدف افزایش دقت توصیه و افزایش تنوع انتخاب از دسته‌ها تبدیل و الگوریتم NSGAII بر روی آن اعمال گردیده است. نتایج حاصل از اعمال روش پیشنهادی بر روی مجموعه‌داده Movielens نشان می‌دهد روش پیشنهادی ضمن افت کمی از دقت نسبت به روش پالایش مشارکتی مبتنی بر اقلام، تعداد اقلام دنباله طولانی در لیست‌های توصیه را به طور قابل توجه افزایش داده است. به عنوان کارهای آینده می‌توان معیارهای دیگری جز تازگی جهت ایجاد تنوع در اقلام پیشنهادی و کاهش مشکل اقلام دنباله طولانی ارائه نمود.

مراجع

[۱] ا. سمیعی، ک. میرزائی، **ارزیابی یک سیستم توصیه گر با استفاده از الگوریتم تکاملی ترکیبی GAPSO**، اولین کنفرانس بین المللی وب پژوهی، تهران، ۱۳۹۴.

[۲] م. حسین زاده اقدم، ح. امامی، ر. دهخوارقانی، **ارائه روشی نوین برای خوشه بندی در سیستم های توصیه گر با استفاده ازخود تشابهی**، دومین کنفرانس ملی فناوری های نوین در مهندسی برق و کامپیوتر، اصفهان، ۱۳۹۷.

[۳] س. موسوی فخر، خ. رهسپارفرد، **ارائه یک الگوریتم جهت گروه بندی محصولات و مشتریان در سیستم توصیه گر محصول با خوشه بندی مبتنی بر فیلتر مشارکتی**، اولین کنفرانس بین المللی چالش ها و راهکارهای نوین در مهندسی صنایع و مدیریت و حسابداری، ساری، ۱۳۹۹.

[۴] ز. گل نتاج، م. عادل خواه، **سیستم های توصیه گر مبتنی بر پالایش مشارکتی**، دومین کنفرانس ملی رویکردهای نوین در مهندسی کامپیوتر و برق، رودسر، ۱۳۹۵.

Social Data, IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 1, pp. 194-208, 2021.

[16] L. Zhang, Q. Wei, L. Zhang, B. Wang, and W.-H. Ho, *Diversity Balancing for Two-Stage Collaborative Filtering in Recommender Systems*, Applied Sciences ,vol. 10, no. 4, 2020.

[17] A. Gharahighehi and C. Vens, *Personalizing Diversity Versus Accuracy in Session-Based Recommender Systems*, SN Computer Science, vol. 2, no. 1, p. 39, 2021.

[18] H. Abdollahpouri, R. Burke, and B. Mobasher, *Managing Popularity Bias in Recommender Systems with Personalized Re-ranking*, 2019.

[19] E. Malekzadeh Hamedani and M. Kaedi, *Recommending the long tail items through personalized diversification*, Knowledge-Based Systems, vol. 164, pp. 348-357, 2019.

[20] H. Abdollahpouri, R. Burke, and B. Mobasher, *Controlling Popularity Bias in Learning-to-Rank Recommendation*, 2017.

[21] A. Jain, P. K. Singh, and J. Dhar, *Multi-objective item evaluation for diverse as well as novel item recommendations*, Expert Systems with Applications, vol. 139, p. 112857, 2020.

[22] K. Kim, C. Park, and H. Yu, *Sequential and Diverse Recommendation with Long Tail*, pp. 2740-2746, 2019.

[23] J. Díez, D. Martinez, A. Alonso-Betanzos, O. Luaces, and A. Bahamonde, *Optimizing novelty and diversity in recommendations*, Progress in Artificial Intelligence, vol. 8, 2018.

[24] M. B. D. õaz, *Analysis of a Bipartite Network of Movie Ratings and Catalogue Network Growth Models*, 2008.

[25] X. Cai, Z. Hu, P. Zhao, W. Zhang, and J. Chen, *A hybrid recommendation system with many-objective evolutionary algorithm*, Expert Systems with Applications, vol. 159, p.1, 2020.

[26] S. Wang, M. Gong, H. Li, and J. Yang, *Multi-objective optimization for long tail recommendation*, Knowledge-Based Systems, vol. 104, pp. 145-155, 2016.

[27] S. Huang, J. Liu, D. Zhu, and D. Cai, *Application of Multi-objective Evolutionary Algorithm in E-Commerce Recommendation System*, in 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering *(*AEMCSE*)*,Shenzhen, pp. 129-132, 2020.

[5] V. Jain and K. Bansal, *Novel Approach using user based Similarity for Recommendation Systems*, in 2020 Fourth International Conference on Inventive Systems and Control (ICISC) ,pp. 522-526, 2020.

[6] C. Su, T. Zhang, and X. Xie, *I_ConvCF: Item-based Convolution Collaborative Filtering Recommendation*, presented at the Proceedings of the 2020 4th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, Thimphu, Bhutan, 2020.

[7] N. Nassar, A. Jafar, and Y. Rahhal, *Multi-criteria collaborative filtering recommender by fusing deep neural network and matrix factorization*, Journal of Big Data, vol. 7, pp. 1-12, 2020.

[8] W. Meng, D. Yang, and Y. Xiao, *Incorporating user micro-behaviors and item knowledge into multi-task learning for session-based recommendation*, in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1091-1100, 2020.

[9] J. Wang, A. K. Sangaiah, and W. Liu, *A hybrid collaborative filtering recommendation algorithm: integrating content information and matrix factorisation*, International Journal of Grid and Utility Computing, vol. 11, no. 3, pp. 367-377, 2020.

[10] B. Alhijawi and Y. Kilani, *A collaborative filtering recommender system using genetic algorithm*, Information Processing & Management, vol. 57, no. 6, p. 102310, 2020.

[11] N. S. Achary and B. K. Patra, *Graph Based Hybrid Approach for Long-Tail Item Recommendation in Collaborative Filtering*, in 8th ACM IKDD CODS and 26th COMAD, pp. 426-426, 2021.

[12] S. Liu and Y. Zheng, *Long-tail Session-based Recommendation*, arXiv e-prints, p. arXiv:2007.12329, 2020.

[13] Y. Wang, J. Wu, Z. Wu, and G. Adomavicius, *From Head to Long Tail: Efficient and Flexible Recommendation Using Cosine Patterns*, Available at SSRN 3762687, 2020.

[14] J. Pang, J. Guo, and W. Zhang, *Using Multi-objective Optimization to Solve the Long Tail Problem in Recommender System*, in Advances in Knowledge Discovery and Data Mining, pp. 302-313, 2019.

[15] J. Li, K. Lu, Z. Huang, and H. T. Shen, *On Both Cold-Start and Long-Tail Recommendation with*

[28] N. E. I. Karabadji, S. Beldjoudi, H. Seridi, S. Aridhi, and W. Dhifli, *Improving memory-based user collaborative filtering with evolutionary multi-objective optimization*, Expert Systems with Applications, vol. 98, pp. 153-165, 2018.
[29] K. Deb, *Multi-objective Genetic Algorithms: Problem Difficulties and Construction of Test Problems*, Evolutionary Computation, vol. 7, no. 3, pp. 205-230, 1999.
[30] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, *A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II*, in Parallel Problem Solving from Nature PPSN VI, Berlin, Heidelberg, M. Schoenauer et al, pp. 849-858, 2000.
[31] L. A. Hassanieh, C. A. Jaoudeh, J. B. Abdo, and J. Demerjian, *Similarity measures for collaborative filtering recommender systems*, in 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), Lebanon, pp. 1-5, 2018.
[32] M. Rahul, P. Pal, V. Yadav, D. K. Dellwar, and S. Singh, *Impact of Similarity Measures in K-means Clustering Method used in Movie Recommender Systems*, IOP Conference Series: Materials Science and Engineering, Kolaghat, vol. 1022, p. 012101, 2021.
[33] https://grouplens.org/datasets/movielens/1m/

زیرنویس

[1] Content-based filtering
[2] Collaborative filtering
[3] Knowledge-based
[4] Hybrid
[5] User-based
[6] Item-based
[7] Matrix factorization
[8] Long-tail items
[9] Coverage
[10] Diversity
[11] Freshness
[12] Initial attractiveness
[13] Multi objective optimization
[14] Non-dominated Sorting Genetic Algorithm