

Improving the Performance of The Recommender System by Solving the Cold Start Problem Using Frequent Pattern Extraction Based on Fuzzy Clustering

Samaneh Zirak¹, Mehdi Salkhordeh Haghighi^{2*}

1- Faculty of Computer Engineering and information technology, Sadjad University, Mashhad, Iran.

2*- Faculty of Computer Engineering and information technology, Sadjad University, Mashhad, Iran.

¹ samaneh.zirak574@sadjad.ac.ir, ^{2*}haghighi@sadjad.ac.ir

Corresponding author's address: Mehdi Salkhordeh Haghighi, Faculty of Computer Engineering and IT, Sadjad University, mashhad, Iran.

Abstract— The cold-start problem remains one of the major challenges in recommender system, as the lack of sufficient information about new users or items significantly reduces recommendation accuracy and reliability. This paper proposes a hybrid recommendation framework that combines Fuzzy C-Means (FCM) clustering, DBSCAN, and FP-Growth to alleviate the effects of the cold-start problem. After preprocessing the MovieLens datasets, DBSCAN is employed to determine the optimal number of user clusters, and FCM is subsequently applied to group users according to their behavioral similarities. Frequent behavioral patterns are then extracted from each cluster using the FP-Growth algorithm and utilized to generate personalized recommendations for users with similar preferences. The proposed approach is evaluated on the MovieLens 100K and MovieLens 1M datasets using Precision, Recall, and F-measure metrics. Experimental results demonstrate that the proposed method achieves Precision, Recall, and F-measure values of 0.78, 0.76, and 0.77 on the MovieLens 100K dataset, and 0.82, 0.77, and 0.795 on the MovieLens 1M dataset, respectively. To further address the new-user cold-start problem, three initialization strategies, including content-based, demographic-based, and questionnaire-based approaches, are investigated. Among them, the questionnaire-based approach yields the best performance, achieving a Precision value of 0.654. Comparative analyses indicate that the proposed framework effectively improves recommendation quality and reduces the impact of the cold-start problem compared with existing approaches.

Keywords— Cold-start problem, recommender system, fuzzy clustering, Fuzzy C-Means (FCM), DBSCAN, FP-Growth, collaborative filtering, frequent pattern mining.

بهبود عملکرد سیستم توصیه‌گر با حل مشکل شروع سرد با استفاده از استخراج الگوهای پرتکرار مبتنی بر خوشه‌بندی فازی

سمانه زیرک، مهدی سالخورده حقیقی^{۲*}

۱- دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه سجاد، مشهد، ایران.

۲*- دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه سجاد، مشهد، ایران.

¹ samaneh.zirak574@sadjad.ac.ir, ^{2*} haghghi@sadjad.ac.ir

* نشانی نویسنده مسئول: مهدی سالخورده حقیقی، مشهد، بلوار جلال آل احمد ۶۴، دانشگاه سجاد، دانشکده کامپیوتر و فناوری اطلاعات.

چکیده- یکی از چالش‌های مهم در سیستم‌های توصیه‌گر، مسئله شروع سرد است که به دلیل کمبود اطلاعات درباره کاربران یا آیتم‌های جدید، موجب کاهش دقت و کارایی توصیه‌ها می‌شود. در این پژوهش، رویکردی چند مرحله‌ای برای کاهش اثرات این مسئله در سیستم‌های توصیه‌گر ارائه شده است. در گام نخست، داده‌های مجموعه داده MovieLens پیش‌پردازش و برای فرآیند توصیه آماده‌سازی می‌شوند. سپس کاربران با استفاده از الگوریتم خوشه‌بندی فازی (FCM) گروه‌بندی شده و به منظور تعیین تعداد بهینه خوشه‌ها از الگوریتم DBSCAN بهره گرفته می‌شود. در ادامه، الگوریتم FP-Growth برای استخراج الگوهای پرتکرار رفتاری کاربران در هر خوشه به کار گرفته می‌شود. الگوهای استخراج شده مبنایی برای ارائه پیشنهاد های شخصی سازی شده فراهم می‌کنند و از طریق شناسایی کاربران دارای رفتار مشابه، دقت توصیه‌ها را افزایش می‌دهند. عملکرد روش پیشنهادی با استفاده از معیارهای Recall، Precision و F-measure ارزیابی شده است. نتایج حاصل بر روی مجموعه داده MovieLens-100K به ترتیب مقادیر ۰٫۷۸، ۰٫۷۶ و ۰٫۷۷ و بر روی مجموعه داده MovieLens-1M به ترتیب مقادیر ۰٫۸۲، ۰٫۷۷ و ۰٫۷۹۵ را برای معیارهای مذکور نشان می‌دهد. همچنین برای مواجهه با مسئله شروع سرد کاربران جدید، سه رویکرد مبتنی بر محتوا، مبتنی بر اطلاعات جمعیت‌شناختی و روش ترکیبی مبتنی بر پرسشنامه اولیه مورد بررسی قرار گرفت. نتایج نشان داد که رویکرد مبتنی بر پرسشنامه اولیه با مقدار دقت برابر ۰٫۶۵۴ عملکرد بهتری نسبت به سایر روش‌ها دارد. مقایسه نتایج با پژوهش‌های مرتبط نیز بیانگر کارایی و برتری نسبی روش پیشنهادی در کاهش اثرات مسئله شروع سرد در سیستم‌های توصیه‌گر است.

واژه‌های کلیدی: شروع سرد، سیستم‌های توصیه‌گر، خوشه‌بندی فازی، DBSCAN، FP-Growth، پالایش مشارکتی، استخراج الگوهای پرتکرار.

۱- مقدمه

طی سالهای گذشته پژوهشگران رویکردهای مختلفی برای ساخت سیستم‌های توصیه‌گر پیشنهاد کرده‌اند که از تاریخچه امتیازات کاربران و احتمالاً برخی اطلاعات دیگر مانند ویژگی‌های جمعیتی کاربران و مشخصات آیتم‌ها استفاده می‌کنند [۳]. اکثر این رویکردها به سه دسته اصلی تقسیم می‌شوند [۴].

سیستم‌های توصیه‌گر به یکی از ابزارهای کلیدی در بهبود تجربه کاربران در سرویس‌های پخش آنلاین، تجارت الکترونیک و رسانه‌های اجتماعی تبدیل شده‌اند [۱]. این سیستم‌ها با تحلیل رفتار و تعاملات کاربران، آیتم‌های مناسب را پیشنهاد می‌دهند [۲].

الگوهای پرتکرار رفتاری در هر خوشه به کار گرفته می شود تا توصیه های مرتبطتری برای کاربران تولید شود [۱۴]. رویکرد پیشنهادی با هدف بهبود دقت توصیه ها، افزایش توانایی سیستم در مواجهه با کاربران و آیتم های جدید و کاهش وابستگی به حجم زیاد داده های اولیه طراحی شده است. نتایج حاصل از ارزیابی بر روی مجموعه داده MovieLens نشان می دهد که چارچوب پیشنهادی می تواند عملکرد قابل قبولی در شرایط شروع سرد ارائه دهد.

۲- پیشینه تحقیق

سیستم های توصیه گر به عنوان ابزارهای قدرتمند ارائه پیشنهادات شخصی سازی شده، نقش مهمی در بهبود تجربه کاربران ایفا می کنند. اهمیت این حوزه در افزایش دقت و کارایی پیشنهادها نهفته است، زیرا ارائه توصیه های مرتبط تر منجر به افزایش رضایت کاربران و تعامل بیشتر آنان با سیستم می شود.

کنعوت و همکارانش [۱۵] کاهش اثر مشکل شروع سرد کاربران و آیتم های جدید را در سیستم های توصیه گر مبتنی بر پالایش مشارکتی بررسی کرده اند. در این مطالعه، به جای ارائه یک روش واحد، از چندین استراتژی برای تولید مجموعه آیتم های پرتکرار به منظور ارائه توصیه های معنادار استفاده شده است. برای ارزیابی، یک مدل پایه بدون استفاده از الگوریتم FP-Growth در نظر گرفته شده است.

گروزدکی و همکارانش [۱۶] با بهره گیری از خوشه بندی فازی، چارچوبی برای کاهش مشکل پراکندگی داده ها و بهبود عملکرد در سناریوی شروع سرد ارائه دادند. در این روش، پالایش مشارکتی مبتنی بر کاربر به صورت فازی پیاده سازی شده است.

ازهر و همکارانش [۱۷] عملکرد الگوریتم های FP-Growth و TPQ-Apriori را در استخراج الگوهای پرتکرار مقایسه کرده اند و نشان داده اند که FP-Growth از نظر زمان اجرا و مصرف حافظه کارایی بالاتری دارد. همچنین کنعوت و همکارانش [۱۸] چارچوب FPRS را به عنوان یک سیستم توصیه گر ترکیبی معرفی کردند که در آن از FP-Growth برای استخراج مجموعه آیتم های پرتکرار استفاده شده و با خوشه بندی تجمعی روی ژانرها، مشکل پراکندگی کاهش یافته است.

الرضایی و همکارانش [۱۹] با ترکیب مدل مبتنی بر محتوا و پالایش مشارکتی، روشی برای کاهش خطای MAE و RMSE در مسئله شروع سرد ارائه کردند. کاوایی و همکارانش [۲۰] نیز رویکردهای مبتنی بر مدل موضوع را برای حل مسئله کمبود اطلاعات اولیه بررسی کرده اند که نتایج آن ها بهبود عملکرد در شرایط داده های

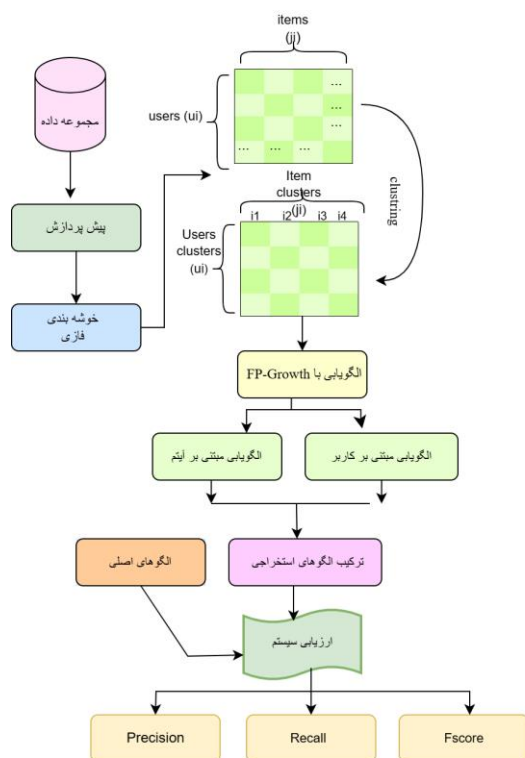
(۱) پالایش مبتنی بر همکاری (۲) پالایش مبتنی بر محتوا (۳) پالایش سازی ترکیبی. این تکنیک ها با وجود اینکه توسط بسیاری از متخصصین و کسب و کارها مورد توجه قرار گرفته اند اما با مشکلات قابل توجهی نیز روبه رو هستند [۵]. یکی از این چالش ها مشکل شروع سرد است. بنابراین در سیستم های توصیه گر و به ویژه روش های همکاری به دلیل اطلاعات ناکافی در مورد موجودیت های جدید هم از مسئله شروع سرد کاربر و هم از مسئله شروع سرد آیتم رنج می برند [۶]. این مشکل به ویژه هنگامی رخ می دهد که اطلاعات کافی درباره کاربران جدید یا آیتم های تازه وارد شده در دسترس نیست، که این باعث کاهش دقت و کارایی توصیه ها می شود [۷]. بیشتر تلاش ها برای حل چنین مسئله ای پیشنهاد ترکیب روش های پالایش مشارکتی با روش مبتنی بر محتوا است [۸]. برای حل این مشکل، روش های متعددی از جمله خوشه بندی و استخراج الگوهای پرتکرار نیز مورد استفاده قرار گرفته اند [۹]. روش های مبتنی بر خوشه بندی سلسله مراتبی در بسیاری از سیستم ها برای رفع این مشکل به کار گرفته شده اند [۱۰]. با این حال، این روش ها دارای محدودیتهایی هستند، از جمله عدم انعطاف پذیری در مواجهه با داده های ناقص یا عدم قطعیت های موجود در رفتار کاربران [۱۱]. از این رو، استفاده از روش های نوآورانه مانند خوشه بندی فازی که می تواند رفتارهای پیچیده و عدم قطعیت را بهتر مدل سازی کند، اهمیت زیادی دارد [۱۲].

با وجود پژوهش های متعدد در زمینه کاهش مشکل شروع سرد، بسیاری از روش های موجود یا بر خوشه بندی های قطعی متکی هستند و یا برای استخراج دانش از تعاملات کاربران از اطلاعات محدود استفاده می کنند. همچنین در بخش قابل توجهی از مطالعات، تعیین تعداد خوشه ها به صورت از پیش تعریف شده انجام می شود که می تواند بر کیفیت خوشه بندی و دقت توصیه ها تأثیرگذار باشد. از سوی دیگر، بهره گیری همزمان از خوشه بندی فازی و استخراج الگوهای پرتکرار برای مدل سازی رفتار کاربران در شرایط شروع سرد کمتر مورد توجه قرار گرفته است. این موضوع ضرورت ارائه رویکردی کارآمد برای شناسایی دقیق تر علایق کاربران و تولید توصیه های مؤثرتر را نشان می دهد. در این پژوهش، به منظور کاهش اثرات مشکل شروع سرد در سیستم های توصیه گر، یک چارچوب ترکیبی مبتنی بر خوشه بندی فازی و استخراج الگوهای پرتکرار ارائه شده است. در گام نخست، با بهره گیری از الگوریتم DBSCAN ساختار اولیه داده ها بررسی و تعداد مناسب خوشه ها تعیین می شود. سپس خوشه بندی فازی به منظور مدل سازی دقیق تر رفتار کاربران و امکان عضویت همزمان آن ها در چند خوشه مورد استفاده قرار می گیرد [۱۳]. در ادامه، الگوریتم FP-Growth برای استخراج

۳-۱- طرح پیشنهادی

محدود را نشان می‌دهد.

در شکل (۱) فرآیند پیشنهادی فیلم به کاربر جدید مبتنی بر ترکیب خوشه‌بندی فازی و الگوریتم FP-Growth آورده شده است. در شکل (۱) فرآیند پیش‌بینی فیلم مبتنی بر ترکیب الگوریتم خوشه‌بندی فازی و الگوریتم FP-Growth آورده شده است که این فرآیند از بخش‌های پیش‌پردازش، خوشه‌بندی فازی و الگویی تشکیل گردیده است. در ادامه هر یک از بخش‌های مورد نظر شرح داده شده است.



شکل ۱. فرآیند پیشنهادی فیلم مبتنی بر ترکیب خوشه‌بندی فازی و الگوریتم FP-Growth

۳-۲- پیش‌پردازش

هدف از این بخش بهبود داده‌ها قبل از عملیات خوشه‌بندی می‌باشد. در شکل (۲) عملیات پیش‌پردازش شرح داده شده است. همانطور که در شکل (۲) نشان داده شده است، ابتدا داده‌ها به دو بخش آموزش و آزمون تقسیم می‌شوند. سپس عملیات اصلاح داده‌های گمشده و نرمال‌سازی تنها با استفاده از داده‌های آموزش انجام شده و پارامترهای به‌دست‌آمده، بدون استفاده از اطلاعات آزمون، روی داده‌های آزمون اعمال می‌گردد. برای جلوگیری از نشت داده و افزایش غیرواقعی معیارهای ارزیابی، مراحل پیش‌پردازش به صورت استاندارد انجام شده است.

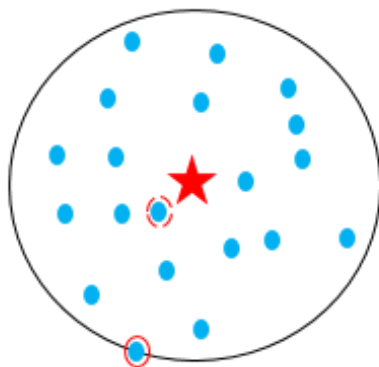
پولیس و همکارانش [۲۱] از شبکه‌های عصبی سیامی برای بهبود شباهت‌سنجی در سیستم‌های توصیه‌گر مبتنی بر محتوا استفاده کرده‌اند. نتایج آن‌ها نشان می‌دهد مدل‌های مبتنی بر یادگیری عمیق می‌توانند در شرایط شروع سرد عملکرد بهتری نسبت به روش‌های سنتی داشته باشند.

در سال‌های اخیر، رویکردهای مبتنی بر یادگیری تضادی (Contrastive Learning) و مدل‌های گرافی نیز برای حل مسئله شروع سرد مورد توجه قرار گرفته‌اند. پژوهش‌های جدید نشان داده‌اند که استفاده از یادگیری تضادی برای تقویت نمایش‌های نهفته کاربران و آیتم‌ها می‌تواند عملکرد سیستم را در شرایط داده‌های پراکنده بهبود دهد [۲۲]. همچنین مدل‌های مبتنی بر گراف و هیپرگراف با استخراج ساختار پیچیده روابط کاربران و آیتم‌ها، توانسته‌اند دقت توصیه‌ها را در سناریوهای cold-start افزایش دهند [۲۳]، [۲۴].

با وجود پیشرفت‌های قابل توجه روش‌های مبتنی بر یادگیری عمیق، این مدل‌ها معمولاً نیازمند تنظیم پارامترهای گسترده و هزینه محاسباتی بالا هستند. در مقابل، پژوهش حاضر با تمرکز بر ترکیب خوشه‌بندی فازی و استخراج الگوهای پرتکرار، تلاش کرده است رویکردی با پیچیدگی کمتر و قابلیت پیاده‌سازی ساده‌تر ارائه دهد که همچنان در مواجهه با مسئله شروع سرد عملکرد رقابتی داشته باشد.

۳- روش پیشنهادی

در این پژوهش جهت رفع مشکل شروع سرد از یک فرآیند چند مرحله‌ای استفاده می‌شود. در مرحله‌ی نخست عملیات دریافت داده انجام می‌شود. در این بخش داده‌ها موجود در دیتاست استاندارد MovieLens به صورت قابل فهم برای سیستم پیشنهادی تبدیل می‌گردد. در بخش دوم عملیات اصلاح داده صورت می‌گیرد که این عملیات جز عملیات پیش‌پردازش می‌باشد. در بخش سوم، عملیات خوشه‌بندی براساس الگوریتم فازی گروه‌بندی می‌شود. در این بخش سعی می‌گردد کاربران را از نظر رفتاری به خوشه‌های متعدد گروه‌بندی نمود به صورتی که در هر گروه کاربران هم سلیقه قرار گیرد. در بخش نهایی از الگوریتم FP-Growth برای یافتن الگوهای پرتکرار استفاده می‌گردد. هدف از این کار یافتن الگوهایی است که بتواند براساس آن‌ها، فیلم‌هایی به کاربر پیشنهاد دهد. در ادامه هر یک از بخش‌های روش پیشنهادی شرح داده شده است.



شکل ۳: درجه عضویت در خوشه‌بندی فازی.

بخش اصلی الگوریتم فازی‌سازی مربوط به محاسبه درجات عضویت و به‌روزرسانی مراکز است. مقدار عضویت نشان‌دهنده میزان تعلق نمونه به هر خوشه است و برای نمایش این ارتباط از یک رابطه مشخص استفاده می‌شود که در ادامه آورده شده است.

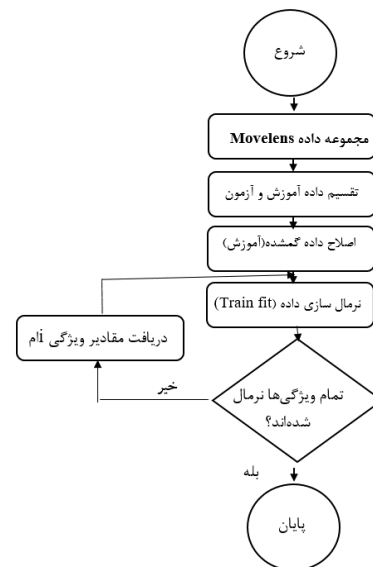
$$J_m(U, V) = \sum_{k=1}^M \sum_{i=1}^C u_{ik}^m \|x_k - v_i\|^2 \quad (1)$$

در رابطه (۱)، x_k نشان دهنده نمونه i ام، v_i مرکز خوشه i -ی، u_{ik}^m درجه عضویت نمونه k ام در خوشه i ام، تعداد خوشه‌ها، M ، تعداد نمونه‌ها و $m > 1$ پارامتر فازی‌سازی است. مقدار u_{ik}^m عددی بین صفر و یک است و میزان تعلق هر نمونه به هر خوشه را نشان می‌دهد هرچی مقدار تابع هدف J_m کمتر باشد، ساختار خوشه‌بندی حاصل، مناسب‌تر خواهد بود. الگوریتم با به‌روزرسانی متناوب ماتریس عضویت U و مراکز خوشه V ، مقدار J_m را در هر تکرار کاهش می‌دهد تا به همگرایی برسد. فرآیند تکرار تازمانی ادامه می‌یابد که یکی از دو شرط زیر برقرار شود:

$$u_{ik} = \left[\sum_{j=1}^C \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (2)$$

در رابطه (۲)، u_{ik} درجه عضویت نمونه x_k در خوشه i ام را نشان می‌دهد همچنین v_i ، v_j به ترتیب مراکز خوشه‌های i و j هستند. عبارت $\|x_k - v_i\|$ فاصله نمونه x_k از مرکز خوشه i را نشان می‌دهد. پارامتر m میزان فازی بودن خوشه‌بندی را کنترل می‌کند؛ هرچه مقدار m بزرگ‌تر باشد، مرز بین خوشه‌ها نرم‌تر شده و عضویت نمونه‌ها در خوشه‌های مختلف پراکنده‌تر خواهد شد.

پس از محاسبه درجه عضویت نمونه‌ها، مراکز خوشه‌ها براساس میانگین وزن‌دار نمونه‌ها به‌روزرسانی می‌شوند. در این مرحله، نمونه‌هایی که درجه عضویت بیشتری در یک خوشه دارند، تأثیر



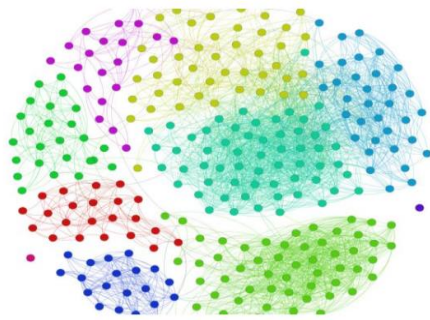
شکل ۴. مراحل پیش‌پردازش داده‌ها.

۳-۳- خوشه‌بندی

الگوریتم فازی FCM، نسخه‌ای فازی از الگوریتم خوشه‌بندی K-means است که به نام FCM شناخته می‌شود. این روش، یکی از متداول‌ترین ابزارها در کار خوشه‌بندی به شمار می‌رود. در این رویکرد، داده‌ها به جای اختصاص به یک خوشه مشخص، با درجات عضویت متفاوت به خوشه‌های مختلف تعلق می‌گیرند. مقدار این درجه عضویت بین ۰ و ۱ است و نشان‌دهنده فاصله نمونه از مراکز خوشه‌هاست. به عبارت دیگر، نمونه‌ای که به مرکز خوشه نزدیک‌تر باشد، مقدار عضویت بالاتری از آن خوشه دریافت می‌کند؛ اگر هم در ناحیه مرزی بین دو خوشه باشد، مقدار عضویت می‌تواند کاهش یابد و به صفر میل کند. شکل (۳) نشان‌دهنده این فرایند است.

مراحل اجرای الگوریتم خوشه‌بندی فازی را می‌توان به صورت شبه‌کد زیر تصور کرد:

- ۱- تعیین تعداد خوشه‌ها برای عملیات خوشه‌بندی (تعیین پارامتر k توسط الگوریتم DBSCAN)
- ۲- انتخاب تصادفی مراکز خوشه
- ۳- محاسبه درجه عضویت هر نمونه نسبت به خوشه‌ها
- ۴- به‌روزرسانی مراکز خوشه بر اساس عضویت‌های محاسبه‌شده
- ۵- ارزیابی تابع هدف
- ۶- تکرار مراحل ۳ تا ۵ تا رسیدن به شرط توقف



شکل ۵: خوشه‌بندی کاربران براساس FCM.

۳-۴- تعیین تعداد خوشه با DBSCAN

یکی از بخش‌های مهم الگوریتم خوشه‌بندی فازی، تعداد خوشه‌ها می‌باشد. این الگوریتم فاقد روش‌های بهینه‌سازی برای یافتن تعداد خوشه‌ی بهینه می‌باشد. از همین رو از الگوریتم DBSCAN برای یافتن تعداد خوشه‌ی بهینه استفاده می‌گردد. الگوریتم DBSCAN یکی از پرکاربردترین روش‌های خوشه‌بندی مبتنی بر چگالی است که در سال ۱۹۹۶ توسط استر معرفی شد. برخلاف الگوریتم‌هایی مانند K-Means که فرض می‌کنند خوشه‌ها به شکل کروی و با اندازه‌های تقریباً برابر هستند، DBSCAN می‌تواند خوشه‌هایی با شکل‌های دلخواه را شناسایی کند و در عین حال نقاط نویز را از داده‌ها جدا نماید. به بیان ساده، DBSCAN نقاطی را که در ناحیه‌ای متراکم از داده‌ها قرار دارند به عنوان خوشه تشخیص می‌دهد و نقاطی را که تنها یا دور از بقیه هستند، به عنوان نویز یا داده پرت مشخص می‌کند.

مبنای اصلی DBSCAN مبتنی بر تراکم هست زیرا خوشه‌ها، نواحی متراکمی از داده‌ها هستند که توسط نواحی کم‌تراکم از یکدیگر جدا می‌شوند. در این روش، تراکم هر ناحیه با شمارش تعداد نقاط موجود در شعاع معینی از هر نقطه مشخص می‌گردد. دو پارامتر اصلی الگوریتم عبارتند از ε (اپسیلون) که شعاع همسایگی را نشان می‌دهد، و MinPts که حداقل تعداد نقاط لازم در این همسایگی برای تشکیل یک ناحیه متراکم را تعیین می‌کند.

فرآیند اجرای DBSCAN به این صورت است که برای هر نقطه x_i در مجموعه داده‌ی $D = \{x_1, x_2, \dots, x_n\}$ ، مجموعه همسایگی ε آن به صورت زیر ارائه می‌شود.

رابطه همسایگی اپسیلون:

$$N_{\varepsilon}(x_i) = \{x_j \in D \mid \text{dist}(x_i, x_j) \leq \varepsilon\} \quad (5)$$

بیشتری بر مرکز آن خوشه خواهند داشت. رابطه به‌روزرسانی مرکز خوشه در الگوریتم FCM به صورت رابطه ۳ است:

$$v_i = \frac{\sum_{k=1}^M u_{ik}^m x_k}{\sum_{k=1}^M u_{ik}^m}, 1 \leq i \leq C \quad (3)$$

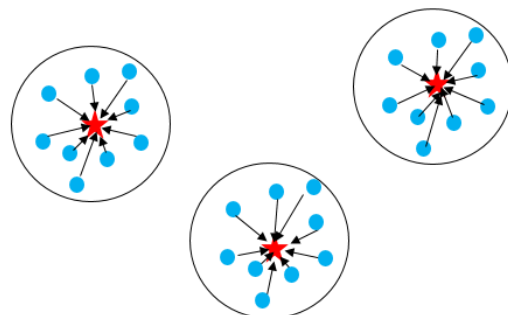
در رابطه (۳)، v_i مرکز خوشه i ام است که با استفاده از تمام نمونه‌ها و براساس درجه عضویت آن‌ها محاسبه می‌شود. مقدار u_{ik}^m نقش وزن هر نمونه را در محاسبه مرکز خوشه مشخص می‌کند. بنابراین، نمونه‌هایی که عضویت بیشتری در خوشه i ام دارند، سهم بیشتری در تعیین مرکز آن خوشه خواهند داشت.

فرآیند محاسبه درجات عضویت و به‌روزرسانی مراکز خوشه‌ها به صورت تکراری انجام می‌شود. این فرآیند تا زمانی ادامه می‌یابد که تغییرات ماتریس عضویت در دو تکرار متوالی کمتر از یک آستانه مشخص شود یا تعداد تکرارها به مقدار بیشینه برسد. شرط توقف الگوریتم در رابطه ۴ نشان داده شده است:

$$\|U^{(t+1)} - U^{(t)}\| < \tau \text{ or } t = \text{MaxIter} \quad (4)$$

در رابطه (۴)، $U^{(t)}$ ماتریس عضویت در تکرار t ، $U^{(t+1)}$ ماتریس عضویت در تکرار بعدی، τ آستانه توقف و MaxIter بیشینه تعداد تکرارها است. اگر تغییرات ماتریس عضویت از مقدار τ کمتر شود یا تعداد تکرارها به MaxIter برسد، الگوریتم به همگرایی رسیده و فرآیند خوشه‌بندی متوقف می‌شود.

مهم‌ترین فاکتور در خوشه‌بندی فازی، فاصله‌ی درون خوشه‌ای می‌باشد. در این حالت خوشه‌ها نسبت به یکدیگر بیشترین فاصله ولی اعضا نسبت به مرکز خوشه در کمترین فاصله قرار دارد. در شکل (۴) این عملیات نشان داده شده است.



شکل ۴: نمودار توزیع کاربران در خوشه‌های استخراج شده

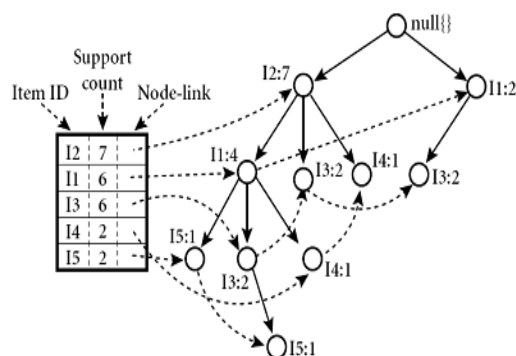
در این بخش کاربران براساس سلایقی که در تماشای فیلم‌ها دارند در خوشه‌های مشابه قرار می‌گیرند که این عملیات در شکل (۵) آورده شده است.

الگوهای موجود بالاست، کارایی Apriori به شدت کاهش می‌یابد. همچنین در برخی کاربردها، این الگوریتم ناگزیر است به دفعات پایگاه داده را پیمایش کند که این کار زمان‌بر است. برای رفع این چالش، FP-Growth معرفی شد. این الگوریتم بدون تولید لیست کاندیدها، با بهره‌گیری از ساختار درختی، الگوهای پرتکرار را استخراج می‌کند. اصل کار FP-Growth بر اساس استراتژی تقسیم و غلبه است. ابتدا نمونه‌ها به صورت درختی نمایش داده می‌شود؛ سپس از این درخت، الگوهای پرتکرار استخراج می‌شوند. درخت نمایش داده‌شده، یک فشرده‌سازی از پایگاه داده را ارائه می‌دهد، به طوری که هر نمونه به عنوان یک مسیر جدید در درخت ثبت می‌شود و با وجود مسیرهای مشابه، همپوشانی بین آن‌ها وجود دارد. در ادامه، مثالی برای ساخت درخت از مجموعه نمونه‌ها ارائه می‌شود؛ در جدول (۱)، کاربرها و فیلم‌های مشاهده شده آورده شده است.

جدول ۱: لیست فیلم‌های مشاهده شده توسط هر کاربر.

User ID	Film ID
1	I1,I2,I3
2	I1,I2,I4
3	I2,I3
4	I1,I3
5	I1,I2,I5
6	I2,I4
7	I2,I3
8	I1,I3
9	I1,I2,I3,I5

همانطور که در جدول (۱) بیان گردیده است، تعداد کاربران ۹ و تعداد فیلم‌ها برابر ۵ می‌باشد. شکل (۶) درخت حاصل از فیلم‌ها را نشان می‌دهد.



شکل ۶: نمونه ساختار FP-tree بر اساس تراکنش‌های کاربر-

در رابطه (۵)، $N_{\epsilon}(x_i)$ مجموعه همسایگان نمونه x_i در شعاع ϵ است. همچنین D مجموعه کل داده‌ها و $dist(x_i, x_j)$ فاصله بین نمونه x_i و x_j را نشان می‌دهد. هر نمونه‌ای که فاصله آن از x_i کمتر یا مساوی ϵ باشد، در همسایگی آن نمونه قرار دارد اگر اندازه‌ی این همسایگی برابر یا بیشتر از مقدار $MinPts$ باشد، یعنی:

$$MinPts \leq |N_{\epsilon}(x_i)| \quad (۶)$$

در رابطه (۶)، $N_{\epsilon}(x_i)$ تعداد نقاط موجود در همسایگی نمونه x_i را نشان می‌دهد. و $MinPts$ حداقل تعداد نقاط لازم برای تشکیل یک ناحیه متراکم است. اگر این شرط برقرار باشد، نمونه x_i یک نقطه مرکزی محسوب می‌شود. نقاطی که در همسایگی یک نقطه مرکزی قرار دارند اما خودشان شرط رابطه ۶ را ندارند، به عنوان نقاط مرزی شناخته می‌شوند.

پس از اجرای DBSCAN، به هر نمونه یک برچسب خوشه اختصاص داده می‌شود. برچسب‌های غیرمنفی نشان‌دهنده خوشه‌های شناسایی شده هستند و برچسب -1 برای نقاط نویز در نظر گرفته می‌شود. در این پژوهش، تعداد خوشه‌های خروجی DBSCAN پس از حذف نقاط نویز به عنوان تعداد خوشه‌های ورودی الگوریتم FCM استفاده شده است. تعداد خوشه‌های نهایی مطابق رابطه ۷ محاسبه می‌شود:

$$K = |unique(labels)\{-1\}| \quad (۷)$$

در این رابطه (۷)، $labels$ مجموعه برچسب‌های تولیدشده توسط الگوریتم DBSCAN است. تابع $unique(labels)$ مقادیر یکتای برچسب‌ها را استخراج می‌کند و مقدار -1 که مربوط به نقاط نویز است، از این مجموعه حذف می‌شود. بنابراین، مقدار K تعداد خوشه‌های واقعی شناسایی شده توسط DBSCAN را نشان می‌دهد. خروجی برچسب گذاری DBSCAN است و مقدار -1 نشان دهنده نقاط نویز است؛ بنابراین در شمارش تعداد خوشه‌ها لحاظ نمی‌شود.

در پیاده سازی این پژوهش، داده‌های ویژگی ابتدا با روش Z-score نرمالیزه شدند. سپس با تنظیم $MinPts=4$ و استفاده از روش زانویی روی منحنی k-dist، مقدار ϵ برابر 0.5 به دست آمد.

۳-۵- الگویابی

یکی از الگوریتم‌های شناخته شده برای استخراج مجموعه الگو، الگوریتم Apriori است. این روش در مواجهه با الگوهای محدود، کارایی مناسبی نشان می‌دهد؛ اما با افزایش تعداد نمونه‌ها و الگوهای زیاد، هزینه‌های محاسباتی به طرز قابل توجهی افزایش می‌یابد. به عنوان مثال، زمانی که هم تعداد نمونه‌ها زیاد است و هم تعداد

ویژگی‌های کاربر و هم از ویژگی‌های فیلم برای پیشنهاد دادن استفاده می‌گردد که در حالت اولیه، فقط براساس ویژگی‌های کاربر، اطلاعات مربوط به فیلم‌ها به کاربر پیشنهاد داده می‌شود.

۴- ارزیابی نتایج

۴-۱- محیط شبیه‌سازی

در جدول (۲) مشخصات محیط شبیه‌سازی از دیدگاه سخت‌افزاری و نرم‌افزاری آورده شده است که این مشخصات در ادامه آورده شده است.

جدول ۲: مشخصات محیط شبیه‌سازی.

ردیف	پارامتر	مقدار/توضیحات
۱	پردازنده (CPU)	intel core i7 8200
۲	حافظه (RAM)	8 gigs
۳	سیستم عامل	Windows 10 - 64 bit
۴	زبان برنامه نویسی	MATLAB 2016a

۴-۲- مجموعه داده

در این پژوهش جهت ارزیابی روش پیشنهادی از مجموعه داده‌ی MovieLens در دو حالت ۱۰۰k و ۱M مورد استفاده قرار گرفته است. مشخصات این مجموعه داده در جدول (۳) آورده شده است.

جدول ۳: مشخصات مجموعه داده‌ی MovieLens.

تعداد فیلم	تعداد کاربر	تعداد نظرات	مجموعه داده
۱۷۰۰	۱۰۰۰	۱۰۰۰۰۰	100 k
۴۰۰۰	۶۰۰۰	۱۰۰۰۰۰۰	1 M

مجموعه داده‌ی MovieLens از سه بخش نظرات، کاربران و فیلم تشکیل شده است که جزییات هر کدام در ادامه آورده شده است.

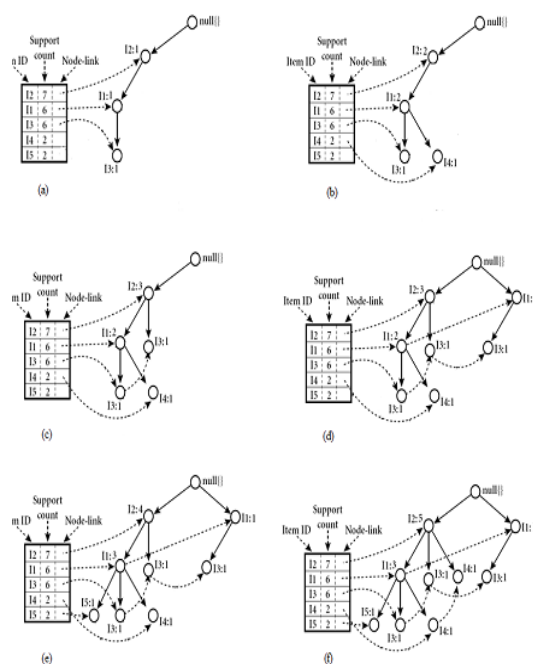
• بخش نظرات: این بخش اطلاعات مربوط به نظرات کاربران بر روی فیلم‌هایی است که تماشا کرده است. مشخصات ذخیره شده برای بخش نظرات در جدول (۴) آورده شده است.

جدول ۴: مشخصات بخش نظرات.

فاکتور اول	فاکتور دوم	فاکتور سوم	فاکتور چهارم
شماره کاربر	شماره فیلم	امتیازدهی به فیلم	زمان تماشای فیلم

همانطور که در جدول (۴) نشان داده شده است، برای هر فیلم، چهار فاکتور شماره کاربر، شماره فیلم، امتیاز دهی به فیلم و زمان تماشای

حال برای ساخت درخت، مانند الگوریتم الگویاب Apriori ابتدا میزان حمایت هر فیلم محاسبه می‌گردد. بعد از محاسبه‌ی معیار حمایت، فیلم‌هایی که شرط حداقل حمایت را داشته باشند انتخاب می‌گردند. در مثال فوق حداقل میزان حمایت ۲ در نظر گرفته شده است. حال براساس میزان حمایت موجود، تمام فیلم‌ها حداقل دوبار مشاهده شده‌اند. از همین رو در گام نخست همه‌ی فیلم‌ها انتخاب گردیده است. بعد از محاسبه‌ی معیار حمایت، فیلم‌ها براساس معیار حمایت از بزرگ به کوچک (نزولی) مرتب می‌شوند. و بعد از آن درخت مورد نظر ایجاد می‌گردد. در شکل (۷) فرآیند ایجاد درخت نمایش داده شده است.



شکل ۷: فرآیند ساخت FP-tree بر اساس معیار حمایت.

یکی از چالش‌های اساسی الگویابی، عدم اطلاعات کافی در مورد کاربر یا فیلم مورد نظر می‌باشد. به عبارتی اگر یک کاربر وارد یک سیستم شود و هیچ مشخصاتی از این کاربر در دسترس نباشد، نمی‌توانیم هیچ کاربری را با این کاربر مطابقت دهیم. وقتی که نتوانیم کاربری را به کاربر مورد نظر تطبیق دهیم، نمی‌توانیم هیچ فیلمی را به کاربر مورد نظر پیشنهاد دهیم. که این مشکل به عنوان مشکل شروع سرد شناخته می‌شود. یکی از راهکارهای حل مشکل شروع سرد، استفاده از دیدگاه‌های مختلف برای پیشنهادی فیلم به کاربر می‌باشد. در این حالت ابتدا به سراغ کاربر می‌رویم. اگر اطلاعات کاربر قابل تطبیق دادن به دیگر کاربران باشد، می‌توانیم متناسب با شخصیت کاربر، فیلم‌هایی را به کاربر نمایش دهیم ولی اگر ویژگی کاربر در دسترس نباشد، براساس فیلم پیشنهادی کاربر اقدام به پیشنهاد فیلم‌های مشابه می‌نماییم. به عبارتی در این حالت هم از

فیلم در نظر گرفته شده است.

- بخش کاربران: در این بخش مشخصات کاربران آورده شده است. که این مشخصات در جدول (۵) آورده شده است:

جدول ۵: مشخصات کاربران.

مشخصات	UserID	Gender	Age	Occupation	Zip-code
مقدار	شماره کاربر	جنسیت کاربر	سن کاربر	شغل کاربر	کدپستی کاربر

همانطور که در جدول (۵) نشان داده شده است، برای هر کاربر مشخصات کاربر ذخیره شده است. در جدول (۶) مقادیر حروفی این مشخصه آورده شده است.

جدول ۶: مشخصات شغلی هر کاربر

شغل	مقادیر عددی	شغل	مقادیر عددی
Administrator	0	Marketing	11
Artist	1	None	12
Doctor	2	Other	13
Educator	3	Programmer	14
Engineer	4	Retired	15
Entertainment	5	Salesman	16
Executive	6	Scientist	17
Healthcare	7	Student	18
Homemaker	8	Technician	19
Lawyer	9	Writer	20
Librarian	10		

- بخش فیلم‌ها: هر فیلم در سه بخش تنظیم می‌گردد که این بخش در جدول (۷) آورده شده است.

بخش ژانر فیلم نیز همانند بخش شغل کاربر نیاز است که به عدد تبدیل گردد. در جدول (۸) ژانرهای موجود در این مجموعه داده به عدد تبدیل گردیده است.

جدول ۷: مشخصات هر فیلم.

MovieID	Title	Genres
شماره فیلم	عنوان فیلم	ژانر فیلم

جدول ۸: مشخصات ژانر فیلم.

ژانر	مقادیر عددی	ژانر	مقادیر عددی
unknown	0	Film-Noir	10
Action	1	Horror	11
Adventure	2	Musical	12
Animation	3	Mystery	13
Children	4	Romance	14
Comedy	5	Sci-Fi	15
Crime	6	Thriller	16
Documentary	7	War	17
Drama	8	Western	18
Fantasy	9		

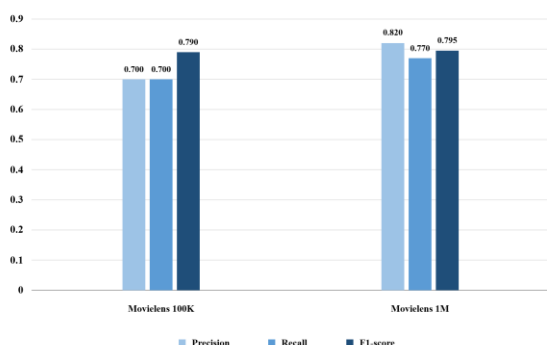
۴-۳- نحوه تقسیم داده‌ها و تنظیمات ارزیابی

در این پژوهش داده‌ها به صورت ۸۰٪ آموزش و ۲۰٪ آزمون تقسیم بندی شدند. تمامی مراحل خوشه‌بندی، استخراج الگو و تولید توصیه‌ها بر روی داده‌های آموزش انجام شد و ارزیابی صرفاً روی مجموعه آزمون گزارش گردید. برای کاهش اثر تصادفی بودن تفکیک، آزمایش‌ها ۵ بار با seed های متفاوت تکرار و میانگین معیارها گزارش شد.

جدول ۹: تنظیم پارامترها.

الگوریتم	پارامتر	مقدار
FP-Growth	(حداقل حمایت) MST	0.6
FP-Growth	(حداقل اعتماد) MCT	0.5
DBSCAN	MinPts, ϵ	4, 0.5
FCM	(ضریب فازی‌سازی) m	2
سیستم توصیه	Top-N	30

جدول (۹) تنظیمات پارامترهای اصلی مورد استفاده در الگوریتم‌های پیشنهادی را نشان می‌دهد. مقدار حداقل حمایت (MST=0.6) به گونه‌ای انتخاب شده است که تنها الگوهای پرتکرار و معنادار استخراج شوند و از تولید قواعد کم‌اهمیت جلوگیری گردد. همچنین حداقل اعتماد (MCT=0.5) تضمین می‌کند که تنها قواعدی با احتمال پیش‌بینی قابل قبول در فرآیند توصیه مورد استفاده قرار گیرند. در الگوریتم DBSCAN مقدار MinPts برابر ۴ در نظر گرفته شد تا خوشه‌های با تراکم واقعی شناسایی شوند و داده‌های پرت حذف گردند. در الگوریتم FCM نیز پارامتر فازی‌سازی m برابر ۲ انتخاب شد که مقدار متداول و پایدار در



شکل ۸. عملکرد روش پیشنهادی بر روی مجموعه داده های MovieLens 100K و MovieLens 1M.

۴-۵- تاثیر خوشه بندی فازی

در جدول (۱۱) عملیات خوشه بندی فازی بهبود یافته مبتنی بر چگالی نسبت به سه روش عدم خوشه بندی و خوشه بندی مبتنی بر Kmeans و خوشه بندی مبتنی بر فازی مورد مقایسه قرار گرفته است که نتایج عملیات خوشه بندی نسبت به عملیات عدم خوشه بندی به مراتب بالاتر می باشد زیرا در روش خوشه بندی، داده های پرت از مجموعه خوشه ها جدا می گردند و باعث می شود الگوریتم بتواند، پیشنهادهای دقیق تری ارائه دهد.

در شکل ۹، عملکرد روش های مختلف خوشه بندی بر روی مجموعه داده MovieLens 100K مقایسه شده است.

جدول ۱۱: ارزیابی روش پیشنهادی بر روی مجموعه داده ی MovieLens

	Method	Precision	Recall	F Measure	O (N)
100K	No Cluster	0.36	0.38	0.369	O (N)
	K-Means	0.64	0.59	0.613	O(MKIN)
	FCM	0.75	0.72	0.734	O(MKIN)
	FCM-DBSCAN	0.78	0.76	0.77	O(M ² +MKIN)
1M	No Cluster	0.41	0.34	0.371	O (N)
	K-Means	0.63	0.60	0.614	O(MKIN)
	FCM	0.74	0.76	0.749	O(MKIN)
	FCM-DBSCAN	0.82	0.77	0.795	O(M ² +MKIN)

خوشه بندی فازی محسوب می شود. مقدار ϵ برابر ۰.۵ تعیین شده است. این مقدار پس از نرمالیزاسیون Z-score ویژگی های سن، جنسیت و شغل و با استفاده از روش زانویی (knee point) روی منحنی k-dist به دست آمده است. در نهایت، تعداد اقلام پیشنهادی به هر کاربر (Top-N) برابر ۳۰ در نظر گرفته شد.

۴-۴- معیارهای ارزیابی

در این بخش معیارهای مورد استفاده جهت ارزیابی روش پیشنهادی آورده شده است که این معیارها شامل معیارهای صحت، فراخوان و معیار F می باشد که در ادامه روابط این معیارها آورده شده است.

$$Precision = \frac{TP}{TP + FP} \quad (۸)$$

$$Recall = \frac{TP}{TP + FN} \quad (۹)$$

$$F Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (۱۰)$$

در این روابط (۸) (۹) (۱۰)، (TP) تعداد فیلم های پیشنهاد شده ای است که واقعاً مورد علاقه کاربر بوده اند، (FP) تعداد فیلم های پیشنهاد شده ای است که مورد علاقه کاربر نبوده اند، و (FN) تعداد فیلم های مرتبطی است که سیستم پیشنهاد نداده است. در این پژوهش، فیلم هایی با امتیاز ۴ و ۵ به عنوان آیتم های مرتبط در نظر گرفته شدند.

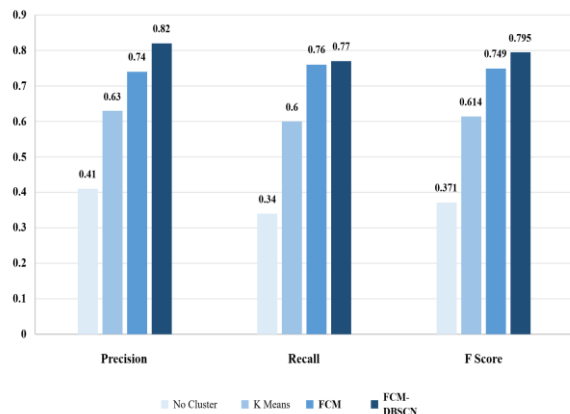
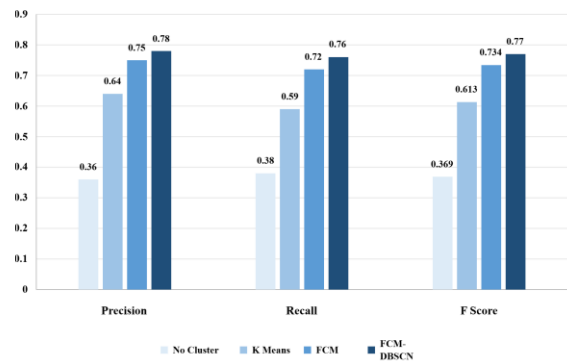
در این بخش نتایج روش پیشنهادی بر روی مجموعه داده ی MovieLens در حالت ۱۰۰k و ۱M مورد مقایسه قرار گرفته است که نتایج این آزمایش ها در جدول (۱۰) آورده شده است.

جدول ۱۰: ارزیابی روش پیشنهادی بر روی مجموعه داده ی MovieLens

	Precision	Recall	F Measure
100 k	0.78	0.76	0.77
1 M	0.82	0.77	0.795

در شکل (۸) عملکرد روش پیشنهادی بر روی مجموعه داده های MovieLens 100K و MovieLens 1M بر اساس معیارهای Precision، Recall و F1-score نشان داده شده است. در ادامه نتایج روش پیشنهادی در حالت های مختلف مورد مقایسه قرار می گیرد.

مجموعه داده‌ی MovieLens در دو حالت ۱۰۰ K و ۱M آورده شده است.



شکل ۹. مقایسه عملکرد روش‌های خوشه‌بندی بر روی مجموعه داده MovieLens 100K.

شکل ۱۰. مقایسه عملکرد روش‌های خوشه‌بندی بر روی مجموعه داده MovieLens 1M.

نتایج ارائه شده در جدول ۱۳ نشان می‌دهد که روش پیشنهادی در تمامی معیارهای ارزیابی شامل دقت (Precision)، بازخوانی (Recall) و معیار F1 عملکرد مطلوبی از خود نشان داده است. این نتایج بیانگر آن است که ترکیب الگوریتم‌های DBSCAN، خوشه‌بندی فازی FCM و استخراج الگوهای پرتکرار با استفاده از FP-Growth توانسته است کاربران دارای علائق مشابه را با دقت بیشتری شناسایی کرده و کیفیت توصیه‌ها را بهبود بخشد. همچنین، استفاده از خوشه‌بندی فازی موجب شده است که هر کاربر با درجات عضویت متفاوت در چندین خوشه قرار گیرد و در نتیجه ترجیحات واقعی کاربران با دقت بیشتری مدل‌سازی شود.

همان‌گونه که در جدول ۱۳ مشاهده می‌شود، عملکرد الگوریتم FP-Growth نسبت به الگوریتم Apriori برتر است. دلیل این برتری آن است که FP-Growth بدون تولید مجموعه‌های کاندیدا عمل کرده و با استفاده از ساختار درخت FP، الگوهای پرتکرار را به صورت مستقیم استخراج می‌کند. این ویژگی موجب کاهش زمان اجرا، کاهش تعداد دفعات پیمایش پایگاه داده و بهینه‌تر شدن مصرف حافظه، به‌ویژه در مجموعه داده‌های بزرگ، می‌شود.

جدول ۱۳: ارزیابی روش الگویاب بر روی مجموعه داده‌ی MovieLens.

	Method	Precision	Recall	F Measure	O (N)
100K	Apriori	0.64	0.63	0.634	$O(2^{N_{ML}} + M^2 + M K I N)$
	FP Growth	0.78	0.76	0.77	$O(M^2 + M K I N)$
1M	Apriori	0.69	0.64	0.664	$O(2^{N_{ML}} + M^2 + M K I N)$
	FP Growth	0.82	0.77	0.795	$O(M^2 + M K I N)$

در شکل (۱۰)، عملکرد همان روش‌ها بر روی مجموعه داده MovieLens 1M گزارش شده است.

در جدول (۱۲) معناداری روش پیشنهادی نسبت به روش‌های خوشه‌بندی سنجیده شده است که این معناداری در ادامه آورده شده است.

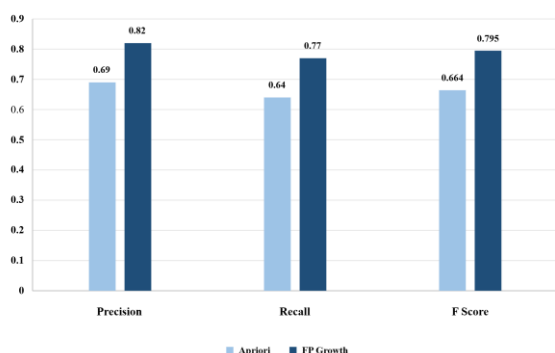
جدول ۱۲: معناداری روش پیشنهادی بر روی مجموعه داده‌ی MovieLens بر روی مجموعه 100k.

Comparison	t-value	df	Value-P
Proposed vs No Cluster	59.58	9	< 0.0001
Proposed vs K-Means	21.59	9	< 0.0001
Proposed vs FCM	7.4	9	< 0.0004
FP Growth	0.86	0.83	0.845

به منظور بررسی عملکرد روش پیشنهادی FCM-DBSCAN با سایر روش‌ها (No Cluster، K-Means و FCM)، از آزمون آماری t-test با درجه آزادی ۹ استفاده شد. نتایج نشان داد که اختلاف عملکرد روش پیشنهادی در مقایسه با هر سه روش مذکور از نظر آماری بسیار معنادار است، به طوری که مقادیر t به ترتیب برابر با 59.58، 21.59 و 7.40 به دست آمد و تمامی مقادیر P کمتر از 0.001 بودند. این نتایج بیانگر آن است که بهبود مشاهده شده در معیار ارزیابی، ناشی از تصادف نبوده و روش پیشنهادی با سطح اطمینان بالایی عملکرد بهتری نسبت به روش‌های مقایسه‌ای ارائه می‌دهد.

۴-۶- تاثیر الگوریتم الگویاب

در این بخش مقایسه‌ای بین دو الگوریتم الگویاب FP-Growth و Apriori آورده شده است. نتایج این آزمایش‌ها بر روی



شکل ۱۲. مقایسه عملکرد الگوریتم‌های Apriori و FP-Growth بر روی MovieLens 1M.

در جدول (۱۴) معناداری روش پیشنهادی نسبت به روش‌های الگویی سنجیده شده است که این معناداری در ادامه آورده شده است.

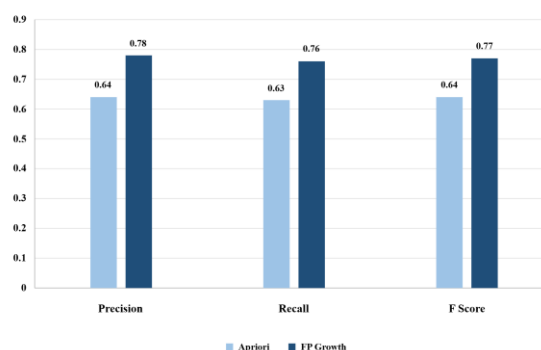
جدول ۱۴: معناداری روش پیشنهادی بر روی مجموعه داده‌ی MovieLens بر روی مجموعه 100k.

	t-value	df	P Value
FP-Growth vs Apriori	20.46	9	< 0.0001

مطابق نتایج ارائه شده در جدول ۱۴، روش پیشنهادی در مقایسه با روش‌های مرجع و مطالعات پیشین عملکرد بهتری را نشان می‌دهد. مقدار بالاتر معیار F1 در روش پیشنهادی بیانگر تعادل مناسب میان دقت (Precision) و بازخوانی (Recall) و در نتیجه کیفیت بالاتر توصیه‌ها است. این نتایج نشان می‌دهد که ترکیب خوشه‌بندی فازی FCM، تعیین تعداد بهینه خوشه‌ها با استفاده از DBSCAN و استخراج الگوهای پرتکرار توسط الگوریتم FP-Growth توانسته است اثرات ناشی از پراکندگی داده‌ها و مسئله شروع سرد را به میزان قابل توجهی کاهش دهد. همچنین، نتایج آزمون آماری حاکی از معناداری برتری روش پیشنهادی نسبت به الگوریتم‌های پایه است. مقدار آماره t برابر با ۲۰،۴۶ و درجه آزادی ۹ نشان‌دهنده اختلاف قابل توجه بین عملکرد الگوریتم FP-Growth و الگوریتم Apriori است. از سوی دیگر، مقدار سطح معناداری ($P\text{-value} < 0.0001$) بیان می‌کند که این تفاوت با اطمینان بسیار بالا از نظر آماری معنادار بوده و ناشی از تصادف یا خطای نمونه‌گیری نیست. بنابراین، نتایج حاصل نه تنها برتری روش پیشنهادی را از نظر شاخص‌های ارزیابی تأیید می‌کند، بلکه اعتبار آماری این برتری را نیز نشان می‌دهد و بیانگر توانایی بیشتر روش پیشنهادی در شناسایی ترجیحات کاربران و تولید توصیه‌های دقیق‌تر و قابل اعتمادتر است.

از منظر پیچیدگی محاسباتی نیز تفاوت قابل توجهی میان دو رویکرد وجود دارد. در روش مبتنی بر Apriori همراه با خوشه‌بندی FCM و DBSCAN، مرتبه زمانی کلی سیستم برابر با $O(2^{NML} + M^2)$ است که در آن M تعداد کاربران، N تعداد فیلم‌ها، L متوسط طول تراکنش، K تعداد خوشه‌ها و I تعداد تکرارهای الگوریتم FCM را نشان می‌دهد. در این رابطه، جمله نمایی 2^{NML} مربوط به تولید مجموعه‌های کاندیدا در Apriori بوده و عامل اصلی افزایش هزینه محاسباتی سیستم محسوب می‌شود. در مقابل، در روش پیشنهادی مبتنی بر FP-Growth، مرتبه زمانی سیستم به $O(M^2 + MKIN)$ کاهش می‌یابد که در آن بخش M^2 مربوط به DBSCAN و بخش MKIN مربوط به FCM است. بنابراین حذف فرآیند تولید کاندیداها در FP-Growth باعث کاهش چشمگیر پیچیدگی زمانی شده و امکان اجرای کارآمد سیستم را حتی در مجموعه داده‌های بزرگ و با تعداد کاربران بالا فراهم می‌کند. این نتایج نشان می‌دهد که استفاده از FP-Growth در کنار FCM و DBSCAN نه تنها دقت توصیه‌ها را افزایش داده، بلکه کارایی محاسباتی سیستم را نیز بهبود بخشیده است.

در شکل (۱۱)، عملکرد الگوریتم‌های Apriori و FP-Growth بر روی مجموعه داده 100K MovieLens مقایسه شده است.



شکل ۱۱. مقایسه عملکرد الگوریتم‌های Apriori و FP-Growth بر روی MovieLens 100K.

در شکل (۱۲)، عملکرد الگوریتم‌های Apriori و FP-Growth بر روی مجموعه داده 1M MovieLens مقایسه شده است.

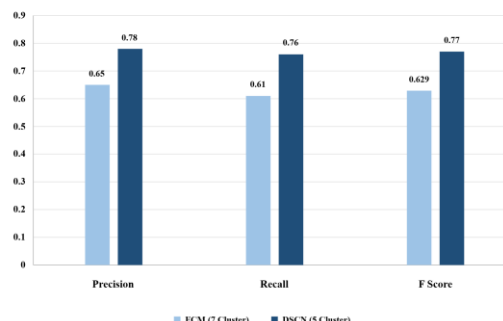
۷-۴- بررسی تعداد خوشه

در این بخش تعداد خوشه‌های بدست آمده از الگوریتم DBSCAN و بدون استفاده از الگوریتم خوشه‌بندی DBSCAN بدست آمده است که این تعداد در جدول (۱۵) آورده شده است.

جدول ۱۵: تاثیر تعداد خوشه بر روی مجموعه داده‌ی MovieLens

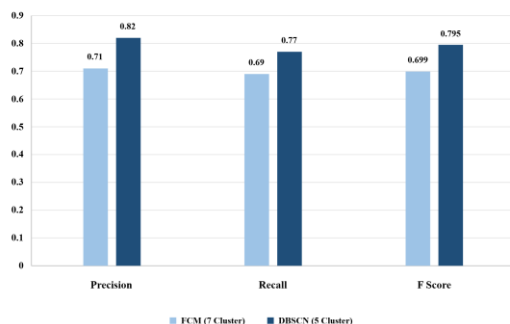
	Method	Precision	Recall	F Measure
100K	FCM (7 Cluster)	0.65	0.61	0.629
	FCM -DBSCAN (5 Cluster)	0.78	0.76	0.77
1M	FCM (7 Cluster)	0.71	0.69	0.699
	FCM -DBSCAN (5 Cluster)	0.82	0.77	0.795

همانطور که در جدول (۱۵) نشان داده شده است، استفاده از خوشه‌بندی باعث بهبود معنادار معیارهای Precision، Recall و F-Measure نسبت به حالت بدون خوشه‌بندی شده است. در روش No Cluster با مرتبه زمانی $O(N)$ ، به دلیل بررسی کلی آیتم‌ها بدون تفکیک ساختاری، دقت پیشنهادها پایین‌تر است. با به کارگیری K-Means و FCM، اگرچه پیچیدگی زمانی به صورت $O(MKIN)$ افزایش می‌یابد، اما به دلیل گروه‌بندی کاربران مشابه، کیفیت توصیه‌ها به طور قابل توجهی بهبود می‌یابد. در نهایت، روش پیشنهادی FCM-DBSCAN با مرتبه زمانی $O(M^2+MKIN)$ ، بیشترین دقت را ارائه داده است. هرچند این روش از نظر محاسباتی سنگین‌تر است، اما جداسازی داده‌های پرت و تعیین بهینه تعداد خوشه‌ها باعث افزایش عملکرد سیستم توصیه گر شده است. نتایج در هر دو مجموعه داده 100K و 1M نشان می‌دهد که بهبود کیفیت توصیه‌ها با افزایش پیچیدگی زمانی همراه بوده است. در شکل (۱۳)، اثر تعیین تعداد خوشه با DBSCAN بر عملکرد FCM در مجموعه داده MovieLens 100K نشان داده شده است.



شکل ۱۳. اثر تعیین تعداد خوشه با DBSCAN بر عملکرد FCM در MovieLens 100K

در شکل (۱۴)، اثر تعیین تعداد خوشه با DBSCAN بر عملکرد FCM در مجموعه داده MovieLens 1M نشان داده شده است.



شکل ۱۴. اثر تعیین تعداد خوشه با DBSCAN بر عملکرد FCM در MovieLens 1M

۸-۴- مقایسه با پژوهش‌های پیشین

در این بخش نتایج روش پیشنهادی بر روی مجموعه داده‌ی MovieLens با نتایج بدست آمده از مقاله مرجع [۱۵] مورد ارزیابی و مقایسه قرار خواهد گرفت. نتایج این ارزیابی در جدول (۱۶) آورده شده است.

جدول ۱۶: مقایسه روش پیشنهادی با روش مرجع [۱۵] بر روی مجموعه داده‌ی MovieLens

	Method	Precision	Recall	F Measure
100K	Strategy 1	0.58	0.56	0.57
	Strategy 2	0.59	0.58	0.58
	Strategy 3	0.57	0.69	0.61
	Strategy 4	0.54	0.70	0.59
	Strategy 5	0.69	0.64	0.66
	Strategy 6	0.79	0.62	0.69
	Strategy 7	0.67	0.63	0.63
	Our Method	0.78	0.76	0.77
1M	Strategy 1	0.63	0.63	0.62
	Strategy 2	0.64	0.64	0.63
	Strategy 3	0.59	0.73	0.67
	Strategy 4	0.59	0.55	0.56
	Strategy 5	0.65	0.71	0.64
	Strategy 6	0.64	0.73	0.66
	Strategy 7	0.6	0.79	0.66
	Our Method	0.82	0.77	0.795

همانطور که در جدول (۱۶) نشان داده شده است، Strategy 1 تا Strategy 7 مطابق نام‌گذاری و نتایج گزارش شده در پژوهش مرجع [۱۵] آورده شده‌اند. برای حفظ انطباق با مقاله مرجع، نام‌گذاری همان پژوهش در جدول حفظ شده است. نتایج نشان می‌دهد روش پیشنهادی در هر دو مجموعه داده MovieLens 100K و MovieLens 1M عملکرد بهتری نسبت به روش مرجع داشته است.

جغرافیایی، شغل یا حوزه تحصیلی می‌باشد. سپس فیلم‌هایی را پیشنهاد می‌دهیم که کاربران مشابه از نظر جمعیتی پسندیده‌اند. در این حالت مشخصات کاربران توسط الگوریتم خوشه‌بندی پیشنهادی، خوشه‌بندی می‌گردند. سپس خوشه‌ی کاربر جدید را مشخص می‌کنیم. بعد از مشخص شدن خوشه‌ی کاربر جدید، K کاربر که از نظر اطلاعات جمعیتی به کاربر مورد نظر نزدیک می‌باشند را انتخاب می‌کنیم. بعد از انتخاب کاربرهای مشابه، از میان فیلم‌های پر تکراری که در این مجموعه کاربر موجود می‌باشد، به کاربر جدید پیشنهاد می‌دهیم. برای ارزیابی این بخش ۱۰۰ کاربر به صورت تصادفی از مجموعه داده انتخاب گردیده است. نتایج آزمایش‌ها بر روی این ۱۰۰ کاربر در جدول (۱۸) آورده شده است.

جدول ۱۸: پیشنهاد فیلم به کاربر براساس اطلاعات جمعیتی کاربر.

روش پیشنهاد فیلم	Precision	Recall	F Measure
اطلاعات جمعیتی کاربر	0.435	0.38	0.405

بر اساس نتایج ارائه شده در جدول (۱۸)، روش پیشنهاد فیلم مبتنی بر اطلاعات جمعیتی کاربران عملکرد متوسطی را در مواجهه با چالش شروع سرد نشان داده است. مقدار صحت برابر با ۰٫۴۳۵، بیانگر آن است که از میان تمام فیلم‌های پیشنهاد داده شده به کاربران ناشناس، حدود ۴۳٫۵ درصد آنها مورد پسند واقع شده‌اند که برای یک سناریوی شروع سرد بدون داشتن هیچگونه تاریخچه‌ای از کاربر، نتیجه قابل قبولی محسوب می‌شود. با این حال، مقدار فراخوان با ۰٫۳۸، نشان می‌دهد که این روش تنها ۳۸ درصد از فیلم‌های مرتبط و مورد علاقه کاربر را شناسایی کرده است. معیار F Measure نیز با مقدار ۰٫۴۰۵، میانگینی از این دو معیار را ارائه می‌دهد. این ارقام حاکی از آن است که اگرچه رویکرد جمعیتی می‌تواند به عنوان نقطه شروعی برای توصیه فیلم به کاربران جدید به کار گرفته شود، اما برای افزایش پوشش و شناسایی بهتر فیلم‌های مرتبط، نیاز به ترکیب با روش‌های تکمیلی دیگر احساس می‌شود.

• روش ترکیبی با پرسشنامه اولیه: در این حالت فرض می‌گردد که در ثبت‌نام کاربر، از کاربر درخواست می‌گردد که چند فیلم مورد علاقه خود را انتخاب کند، به چند فیلم پیشنهادی امتیاز دهد، ژانرهای مورد علاقه خود را مشخص کند، در این حالت نیز براساس داده‌های موجود، k کاربر را که از نظر سلیقه‌ای به کاربر مورد نظر همخوانی دارد، به عنوان مرجع اصلی در نظر گرفته می‌شود. در نهایت از میان کاربران مورد نظر، فیلم‌هایی که بیشترین تکرار را در میان کاربران دارد به عنوان فیلم‌های پیشنهادی به کاربر ارائه می‌گردد. در جدول (۱۹) این نتایج آورده شده است.

این بهبود را می‌توان به ترکیب خوشه‌بندی فازی با استخراج الگوهای پرتکرار توسط الگوریتم FP-Growth نسبت داد.

۴-۹- رویکردهای مقابله با شروع سرد برای کاربران جدید

در این بخش راهکاری‌های عملی برای مقابله با کاربر جدید در نظر گرفته شده است همچنین در سناریوی واقعی شروع سرد کاربر، مجموعه‌ای از کاربران به صورت کامل از داده‌های آموزش حذف شدند و هیچ‌گونه تاریخچه امتیازدهی آن‌ها در مرحله آموزش استفاده نشد. سپس برای این کاربران جدید، سه رویکرد (مبتنی بر محتوا، مبتنی بر اطلاعات جمعیتی، و مبتنی بر پرسشنامه اولیه) اجرا و معیارهای ارزیابی گزارش گردید. به طور مشابه، برای سناریوی شروع سرد آیتیم، مجموعه‌ای از فیلم‌ها تنها در داده‌های آزمون در نظر گرفته شدند و توصیه‌ها بدون استفاده از تعاملات آموزشی مربوط به آن آیتیم‌ها ارزیابی گردید، که در ادامه انواع آن شرح داده می‌شود.

• رویکرد مبتنی بر محتوا: بدون داشتن اطلاعات کاربر و سلیقه کاربر، می‌توانیم فیلم‌های پربیننده، فیلم‌های با بالاترین امتیاز در ژانرهای مختلف و فیلم‌های جدید و محبوب را به کاربر پیشنهاد بدهیم. نتایج این آزمایش در جدول (۱۷) آورده شده است.

جدول ۱۷: پیشنهاد فیلم به کاربر براساس محتوا.

روش پیشنهاد فیلم	Precision	Recall	F Measure
محتوا	0.352	0.308	0.328

نتایج ارائه شده در جدول (۱۷) جهت پیشنهاد فیلم مبتنی بر محتوا، عملکرد نسبتاً ضعیفی را در توصیه فیلم به کاربران نشان می‌دهد. مقدار صحت برابر با ۰٫۳۵۲، بیانگر آن است که تنها حدود ۳۵ درصد از فیلم‌های پیشنهاد شده با سلیقه کاربر مطابقت داشته‌اند که رقم چندان مطلوبی نیست. مقدار فراخوان با ۰٫۳۰۸، نیز حاکی از آن است که این روش تنها ۳۰ درصد از فیلم‌های مرتبط با علایق کاربر را شناسایی کرده است. معیار F Measure نیز با مقدار ۰٫۳۲۸، میانگینی از این دو معیار را نشان می‌دهد. این نتایج می‌تواند نشان‌دهنده ناکارآمدی ویژگی‌های محتوایی انتخاب شده یا عدم تطابق آن‌ها با معیارهای سلیقه کاربران در مجموعه داده مورد آزمایش باشد. به نظر می‌رسد رویکرد مبتنی بر محتوا به تنهایی نمی‌تواند راهکار مناسبی برای مسئله شروع سرد باشد و نیاز به ترکیب با روش‌های دیگر یا بهبود استخراج ویژگی‌های محتوایی دارد.

• استفاده از اطلاعات جمعیتی: اگر کاربر ناشناخته است، حداقل می‌توان از اطلاعات پایه‌ای آن در مرحله ثبت‌نام دریافت استفاده کرد که این اطلاعات شامل سن و گروه سنی، جنسیت، موقعیت

جدول ۱۹: پیشنهاد فیلم به کاربر براساس پرسشنامه.

روش پیشنهاد فیلم	Precision	Recall	F Measure
براساس پرسشنامه	0.654	0.627	0.640

نتایج ارائه شده در جدول (۱۹) نشان دهنده عملکرد بسیار خوب روش پیشنهاد فیلم مبتنی بر پرسشنامه است. این روش توانسته حدود ۶۵ درصد از فیلم های پیشنهادی را به درستی مطابق با سلیقه کاربر انتخاب کند که بیانگر قدرت بالای آن در ارائه پیشنهاد های مرتبط است. مقدار فراخوان برابر با ۰.۶۲۷، نیز حاکی از آن است که این روش موفق به شناسایی حدود ۶۳ درصد از فیلم های مورد علاقه کاربر شده است. معیار F Measure با مقدار ۰.۶۴۰، نشان دهنده تعادل مطلوب بین صحت و فراخوان و عملکرد هماهنگ مدل است.

۵- نتیجه گیری

در این پژوهش، روشی نوین برای مواجهه با مشکل شروع سرد در سیستم های توصیه گر ارائه شد. روش پیشنهادی با بهره گیری از ترکیب خوشه بندی فازی که در آن برای تعیین تعداد بهینه خوشه ها از الگوریتم DBSCAN استفاده شده است و الگوریتم FP-Growth، تلاش می کند با استخراج الگوهای رفتاری کاربران در خوشه های مشابه، توصیه هایی دقیق و هدفمند ارائه دهد. این رویکرد امکان تولید پیشنهاد های مناسب را حتی در شرایط کمبود اطلاعات مربوط به کاربران جدید فراهم می سازد. ارزیابی روش پیشنهادی بر روی مجموعه داده MovieLens در دو مقیاس ۱۰۰K و ۱M و با استفاده از معیارهای صحت (Precision)، فراخوان (Recall) و F-measure انجام شد. نتایج نشان داد که در مقیاس ۱۰۰K، مقادیر صحت، فراخوان و F-measure به ترتیب برابر با ۰.۷۸، ۰.۷۶ و ۰.۷۷ بوده است. همچنین در مقیاس ۱M این مقادیر به ترتیب ۰.۸۲، ۰.۷۷ و ۰.۷۹۵ به دست آمده است که بیانگر بهبود عملکرد با افزایش حجم داده ها است. مقایسه نتایج با روش های مرتبط نیز حاکی از کارایی و برتری نسبی روش پیشنهادی می باشد. علاوه بر این، برای مقابله با مشکل شروع سرد کاربران جدید، سه راهکار شامل رویکرد مبتنی بر محتوا، استفاده از اطلاعات جمعیت شناختی و روش ترکیبی مبتنی بر پرسشنامه اولیه مورد بررسی قرار گرفت. نتایج نشان داد روش مبتنی بر پرسشنامه اولیه با دقت ۰.۶۵۴ نسبت به سایر راهکارها عملکرد مطلوب تری داشته است. با وجود نتایج امیدوارکننده، چالش هایی نظیر مقیاس پذیری در داده های بزرگ مقیاس و مدیریت آیت های جدید همچنان نیازمند پژوهش های تکمیلی و بهبود های بیشتر است. در مجموع، یافته های این پژوهش نشان می دهد ترکیب تکنیک های خوشه بندی فازی و استخراج الگو می تواند راهکاری مؤثر برای کاهش اثرات مشکل شروع سرد در

سیستم های توصیه گر باشد.

این روش، با هدف بهبود دقت و کارایی سیستم های توصیه گر در شرایط شروع سرد انجام شده است و تلاش می کند با بهره گیری از ترکیب خوشه بندی فازی مبتنی بر DBSCAN و استخراج الگوهای پرتکرار، چارچوبی مؤثر و عملی برای کاهش اثرات این چالش ارائه دهد. در حالی که بسیاری از روش های نوین مبتنی بر یادگیری عمیق نیازمند حجم بالایی از داده، تنظیم پارامترهای پیچیده و هزینه محاسباتی سنگین هستند، رویکرد پیشنهادی با تمرکز بر ساختاردهی هوشمند داده ها و کشف الگوهای رفتاری معنادار، راهکاری با پیچیدگی محاسباتی کنترل شده ارائه می کند. این ویژگی سبب می شود روش پیشنهادی در عین سادگی پیاده سازی، عملکردی رقابتی و قابل اتکا در سناریوهای شروع سرد کاربران و آیت ها داشته باشد.

۶- کارهای آینده

در پژوهش های آتی می توان روش پیشنهادی را بر روی مجموعه داده های بزرگ تر و در حوزه های کاربردی مختلف مورد ارزیابی قرار داد تا میزان تعمیم پذیری آن بررسی شود. همچنین ترکیب چارچوب ارائه شده با روش های یادگیری عمیق و مدل های مبتنی بر گراف می تواند به استخراج بهتر ترجیحات کاربران و بهبود کیفیت توصیه ها منجر شود. بهره گیری از اطلاعات زمینه ای کاربران نظیر ویژگی های جمعیت شناختی، داده های شبکه های اجتماعی و بازخوردهای ضمنی نیز از دیگر مسیرهای توسعه این پژوهش خواهد بود. علاوه بر این، بررسی عملکرد روش پیشنهادی در محیط های بلادرنگ و سیستم های توصیه گر مقیاس پذیر می تواند به عنوان یکی از موضوعات مهم تحقیقات آینده مورد توجه قرار گیرد.

مراجع

- [1] R. D. Gottapu and L. V. Sriram Monangi, "Point-Of-Interest Recommender System for Social Groups," *Procedia Comput. Sci.*, vol. 114, pp. 159–164, 2017.
- [2] J. Gao, T. Zhang, and C. Xu, "A unified personalized video recommendation via dynamic recurrent neural networks," in *Proc. 2017 ACM Multimed. Conf.*, 2017, pp. 127–135.
- [3] S. Lakshmi and Dr. T. Adi Lakshmi, "Recommendation Systems: Issues and challenges," *Int. J. Comput. Sci. Inform. Technol. (IJCSIT)*, vol. 5, no. 4, pp. 5771–5772, 2014.
- [4] F. Fadhel Aljunid and D. H. Manjaiah, "A Survey on Recommendation Systems for Social Media Using big data Analytics," *Int. J. Latest Trends Eng. Technol., Special Issue (SACAIM 2017)*, pp. 48–58.
- [5] J. Ma, G. Li, M. Zhong, X. Zhao, L. Zhu, and X. Li, "LGA: latent genre aware micro-video recommendation on social media," *Multimed. Tools Appl.*, vol. 77, no. 3, pp. 2991–3008, 2018.
- [6] P. Priyanga and A. R. N. Banu Kamal, "Methods of Mining the Data from Big Data and Social Networks Based on Recommender

- [23] X. Guo, T. Zhang, F. Wang, X. Wang, X. Zhang, X. Liu, and Z. Cui, "MMHCL: Multi-Modal Hypergraph Contrastive Learning for Recommendation," arXiv preprint arXiv:2504.16576, 2025.
- [24] C. He, Y. Liu, Q. Li, C. Hong, W. Zhong, and X. Yao, "M²VAE: Multi-Modal Multi-View Variational Autoencoder for Cold-start Item Recommendation," arXiv preprint arXiv:2508.00452, 2025.
- [7] M. Wasid and R. Ali, "An improved recommender system based on multi-criteria clustering approach," *Procedia Comput. Sci.*, vol. 131, pp. 93–101, 2018.
- [8] M. H. Mohamed, M. H. Khafagy, and M. H. Ibrahim, "Recommender Systems Challenges and Solutions Survey," in *Proc. 2019 Int. Conf. Innov. Trends Comput. Eng. ITCE*, Feb. 2019, pp. 149–155.
- [9] M. Tsagkias, T. H. King, S. Kallumadi, V. Murdock, and M. de Rijke, "Challenges and research opportunities in eCommerce search and recommendations," *ACM SIGIR Forum*, vol. 54, no. 1, pp. 1–23, Feb. 2021.
- [10] F. Merabet and D. Benmerzoug, "QoS prediction for service selection and recommendation with a deep latent features autoencoder," *Comput. Sci. Inf. Syst.*, vol. 19, no. 2, pp. 709–733, 2022.
- [11] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei, and A. Darwesh, "A systematic study on the recommender systems in the e-commerce," *IEEE Access*, vol. 8, pp. 115694–115716, 2020.
- [12] A. Pawlicka, M. Pawlicki, R. Kozik, and R. S. Choraś, "A systematic review of recommender systems and their applications in cybersecurity," *Sensors*, vol. 21, no. 15, p. 5248, Aug. 2021.
- [13] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, "A review on deep learning for recommender systems: Challenges and remedies," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 1–37, Jun. 2019.
- [14] E. Kannout, M. Grodzki, and M. Grzegorowski, "Utilizing frequent pattern mining for solving cold-start problem in recommender systems," in *Proc. 17th Conf. Comput. Sci. Intell. Syst. (FedCSIS)*, Sofia, Bulgaria, Sep. 2022, pp. 217–226.
- [15] E. Kannout, M. Grodzki, M. Grodzki, and H. Nguyn, "Clustering-Based Frequent Pattern Mining Framework for Solving Cold-Start Problem in Recommender Systems," *IEEE Access*, 2024, pp. 13678–13698.
- [16] M. Grodzki and M. Grzegorowski, "Utilizing Frequent Pattern Mining for Solving Cold-Start Problem in Recommender Systems," in *Proc. 17th Conf. Comput. Sci. Intell. Syst.*, 2022, pp. 217–226.
- [17] M. L. Lalu Zazuli Azhar Mardedi, Kartarina, & Moch. Syahrir, "Analisis Perbandingan Algoritma Fp-Growth dan Tpq-Apriori dalam Menentukan Rule Based Terbaik untuk Sistem Rekomendasi Produk," *Jurnal Sistem Informasi dan Telematika*, Vol. 14, No. 2, pp. —, Jul. 2024, doi:10.35200/ex.v14i2.112.
- [18] E. Kannout, M. Grodzki, and M. Grzegorowski, "Towards addressing item cold-start problem in collaborative filtering by embedding agglomerative clustering and FP-Growth into the recommendation system," *Comput. Sci. Inf. Syst.*, vol. 20, no. 4, pp. 1343–1366, 2022.
- [19] N. Musa Al-Rezaei Golian, "Designing a hybrid recommender system to solve the cold start problem using clustering and chaotic particle swarm algorithm," *Sci. J. Soft Comput.*, pp. 50–61, 2024.
- [20] M. Kawai, H. Sato, and T. Shiohama, "Topic model-based recommender systems and their applications to cold-start problems," *Expert Syst. Appl.*, vol. 202, p. 117129, Sep. 2022.
- [21] M. Pulis and J. Bajada, "Siamese neural networks for content-based cold-start music recommendation," in *Proc. 15th ACM Conf. Recommender Syst.*, New York, NY, USA: ACM, 2021, pp. 719–723.
- [22] J. Hu, T. An, Z. Yu, J. Du, and Y. Luo, "Contrastive Learning for Cold Start Recommendation with Adaptive Feature Fusion," in *Proc. 2025 5th Int. Conf. Consumer Electronics and Computer Engineering (ICCECE)*, Dongguan, China, Feb. 28–Mar. 2, 2025, doi:10.1109/ICCECE65250.2025.10985689.