بهبود پالایش مشارکتی در سامانه‌های توصیه‌گر با کمک خوشه‌بندی فازی C– میانگین مرتب‌شده و الگوریتم ازدحام ذرات تطبیقی–آشوبی

جواد حمیدزاده*و[1] منا مرادی[2]

[1] دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه سجاد، مشهد، ایران

[2] دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران

چکیده

سامانه‌های توصیه‌گر زیرمجموعه‌ای از سامانه‌های هوشمند پالایش اطلاعات هستند که در فضای اینترنت علایق کاربر را شناسایی کرده و توصیه‌های مرتبط با سلیقهٔ کاربر را ارائه می‌دهند. پالایش مشارکتی مبتنی بر کاربر، از مهم‌ترین انواع سامانه‌های توصیه‌گر است. از مهم‌ترین چالش‌ها در این سامانه‌ها پراکندگی و حجم زیاد داده‌ها است که بر کارایی آن‌ها اثرگذار است. در روش پیشنهادی، برای نخستین‌بار از الگوریتم خوشه‌بندی فازی C- میانگین مرتب‌شده و الگوریتم تکاملی ازدحام ذرات تطبیقی آشوبی برای خوشه‌بندی کاربران استفاده شده است. هدف روش پیشنهادی بهبود میزان خطای پیش‌بینی در مجموعه‌داده‌های حجیم با پراکندگی زیاد و کاهش تأثیر داده‌های پرت و نوفه است. به‌منظور ارزیابی و اثبات کارایی روش پیشنهادی، آزمایش‌هایی روی پایگاه داده‌های واقعی اجرا شده است. نتایج آزمایش‌ها نشان‌دهندهٔ برتری روش پیشنهادی نسبت به روش‌های مرز دانش بر اساس معیارهای میانگین خطای مطلق، جذر میانگین مربعات خطا، نرخ صحت و زمان محاسباتی است.

واژگان کلیدی: سامانه توصیه‌گر، پالایش مشارکتی، خوشه‌بندی فازی، الگوریتم تکاملی، الگوریتم ازدحام ذرات تطبیقی آشوبی

Improving Collaborative Recommender Systems by Integrating Fuzzy C-Ordered Means Clustering and Chaotic Self-Adaptive Particle Swarm Optimization Algorithm

Javad Hamidzadeh*[1]And Mona Moradi[2]

[1] Faculty of Computer Engineering and Information Technology, Sadjad University, Mashhad, Iran

[2] Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran

Abstract

Recommender systems are a subset of intelligent information filtering systems that discover user interests and provide user-friendly recommendations. User-based collaborative filtering recommender systems are one of the most important types of recommender systems. However, they face voluminous data and sparsity problems that negatively affect the performance of the systems. The proposed method aims to improve the rating prediction in large sparse datasets and reduce the negative impact of outliers and noisy data. To cluster users, the proposed method integrates the fuzzy C-ordered means clustering algorithm with a chaotic self-adaptive particle swarm evolutionary algorithm. Huber's M-estimators,

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۹/۱/۱۱ • تاریخ پذیرش: ۱۴۰۲/۱۱/۲۸ • تاریخ انتشار: ۱۴۰۳/۵/۱۰ • نوع مطالعه: پژوهشی

along with incorporating chaotic dynamics into the optimization process, can help mitigate the effects of noisy objective functions. Besides, the proposed method benefits from PSO's exploitation capabilities, allowing it to converge quickly to promising solutions. Additionally, its chaotic nature improves exploration, enabling it to escape local optima and search a broader solution space. The self-adaptive mechanism allows the algorithm to adjust its parameters dynamically during the optimization process. This adaptability enhances its robustness across different problem domains and varying optimization landscapes. The combination of PSO and evolutionary algorithms with chaotic dynamics often leads to faster convergence compared to traditional optimization techniques. This can result in significant time savings, especially for complex optimization problems. Experiments have been conducted on real-world datasets to evaluate and prove the efficiency of the proposed method. Experimental results show the superiority of the proposed method over state-of-the-art methods based on prediction error criteria, accuracy rates, and computational time.

Keywords: Recommender systems; Collaborative filtering; Fuzzy clustering; Evolutionary algorithm; Chaotic self-adaptive particle swarm optimization algorithm

۱-مقدمه

سامانه‌های توصیه‌گر زیرمجموعه‌ای از سامانه‌های پالایش اطلاعات هستند که با شناسایی علایق کاربر در فضای اینترنت، توصیه‌های مناسب و مرتبط را به او ارائه می‌دهند. طراحی و پیاده‌سازی این سامانه‌ها سه مرحله اصلی دارد: جمع‌آوری اطلاعات، یادگیری، پیش‌بینی و توصیه.

جمع‌آوری اطلاعات به‌صورت ضمنی، صریح و یا ترکیبی از این دو صورت می‌پذیرد. پردازش‌های اولیه متناسب با سامانه در مرحله نخست انجام می‌شوند. در مرحله یادگیری، اطلاعات به‌دست‌آمده با روش‌های مناسب پالایش شده و شاخص‌های مناسب جهت توصیه بهینه به کاربر به‌دست می‌آید. در مرحله پیش‌بینی و توصیه بر اساس شباهت کاربر به همسایگانش امتیازهای احتمالی او به قلم‌کالاها[1] پیش‌بینی شده و قلم‌کالا با بالاترین امتیاز به کاربر هدف، توصیه می‌شود. واضح است که استفاده از روش مناسب، منجر به ارائه توصیه‌هایی با دقت و کارایی بالا می‌شود. سامانه‌های توصیه‌گر براساس پالایش منابع اطلاعاتی موجود و تکنیک مورداستفاده به دو دسته اصلی تقسیم می‌شوند: پالایش مبتنی بر محتوا[2] و مبتنی بر پالایش مشارکتی[3].

سامانه‌های مبتنی بر محتوا، بر اساس شباهت محتوای پروفایل کاربری و یا آیتم‌های آن تعریف شده و فرض می‌کنند که اگر کاربری در گذشته به یک آیتم خاص علاقه داشته، به‌احتمال هنوز هم به آن علاقه‌مند است. آیتم‌های مشابه در گروه‌های یکسانی قرار داده می‌شوند؛ همچنین پروفایل‌های کاربران نیز با توجه به تاریخچهٔ خرید آن‌ها یا با پرسیدن علایق کاربر به‌صورت صریح تکمیل می‌شود؛ البته روش‌هایی نیز وجود دارند که به‌طورکامل محتوا محور نبوده و از اطلاعات شخصی کاربر و اطلاعات وی در شبکه‌های اجتماعی استفاده می‌کنند [۱, ۲]. به طور خلاصه، در این گونه از سامانه‌ها آیتم‌ها خوشه‌بندی شده؛ سپس اگر کاربر آیتمی از یک خوشه را بپسندد، سایر آیتم‌های آن خوشه به وی پیشنهاد می‌شود [۳, ۴].

سامانه‌های مبتنی بر پالایش مشارکتی یکی از انواع پرکاربرد سامانه‌های توصیه‌گر بوده و به دو دسته تقسیم می‌شوند[۵]: مبتنی بر مدل[4] که با ایجاد مدل بر روی حجم زیاد داده‌ها موجب کاهش پیچیدگی زمانی و محاسباتی می‌شوند [۶] و مبتنی بر حافظه[5] که با حفظ تمام داده‌های موجود در حافظه، از آن‌ها برای پیش‌بینی سلیقه کاربر هدف استفاده می‌کند [۷]. در سال‌های اخیر با رشد بسیار سریع و وسیع اطلاعات، چالش‌هایی به‌شرح زیر در سامانه‌های توصیه‌گر مطرح شده‌اند [۸]:

پراکندگی داده‌ها: در سامانه‌هایی که نسبت تعداد کاربران و قلم‌داده‌ها به امتیازهای موجود بیشتر است، پراکندگی ماتریس امتیاز برای یافتن شباهت بین کاربران یا قلم‌داده‌ها، پیش‌بینی امتیازها و علایق کاربران و میزان رضایت کاربران مؤثر است. روش‌های مختلفی برای این مشکل معرفی شده است که برخی از آن‌ها از اطلاعات کمکی برای افزایش نرخ صحت پیش‌بینی، استفاده می‌کنند.

مسئله شروع سرد: این مشکل زمانی پیش می‌آید که کاربر برای نخستین‌بار از سامانه استفاده می‌کند یا قلم‌داده جدیدی به محصولات افزوده می‌شود و هیچ اطلاعی از ویژگی‌ها یا سابقه‌ای از امتیازها و تعاملات آن‌ها وجود ندارد. در این دو نوع چالش وجود دارد: در برخی موارد

[1] Item
[2] Content-based
[3] Collabrating Filtering
[4] Model-based
[5] Memory-based

هیچ امتیازی در مورد قلم‌داده یا کاربر فعال وجود ندارد و در برخی دیگر امتیازهای بسیار کمی در مورد آن‌ها موجود است [۹].

مقیاس‌پذیری: با زیادشدن تعداد کاربران و اقلام، منابع محاسباتی برای برطرف‌کردن درخواست‌های جدید با کمبود مواجه می‌شود.

تنوع: یکی از چالش‌ها در ارائهٔ توصیه به کاربران این است که با توجه به تنوع کاربران و قلم‌داده‌ها، علایق مختلف کاربر پوشش داده شود.

امنیت: در روش پالایش مشارکتی با توجه به این‌که اطلاعات کاربران نظیر کالاهای خریداری‌شده، اطلاعات شخصی و امتیازات داده‌شده به اقلام کالا مورد استفاده قرار می‌گیرد، افزایش امنیت باعث می‌شود کاربران با اطمینان بیشتری، اطلاعات دقیق و صحیح ارائه دهند که این مسئله نیز منجر به افزایش کارایی سامانه‌های توصیه‌گر می‌شود.

خوشه‌بندی مبتنی بر منطق فازی یکی از پرکاربردترین روش‌ها در سامانه‌های توصیه‌گر است که در سال‌های اخیر بسیار مورد بررسی و پژوهش پژوهش‌گران واقع شده است [۱۰, ۱۱]. هرچند وجود مشکلاتی نظیر وابستگی نتایج به انتخاب مراکز اولیه و تأثیر داده‌های پرت در واگرایی الگوریتم در این روش‌ها، باعث شده است که الگوریتم‌های تکاملی و محاسبات زیستی مانند کلونی مورچگان، ازدحام جمعیت، الگوریتم‌های ژنتیک نیز مورد توجه پژوهش‌گران قرار گیرند [۱۲, ۱۳].

در روش پیشنهادی، کاربران با سوابق امتیازات مشابه شناسایی و امتیاز احتمالی کاربر هدف به قلم‌داده‌ها به‌گونه‌ای پیش‌بینی می‌شود که نرخ خطای پیش‌بینی کاهش و نرخ صحت پیش‌بینی‌ها بهبود یابد؛ ازاین‌رو، با توجه به پراکندگی و مقیاس‌پذیری داده‌های موجود، کاربران مشابه به‌کمک الگوریتم خوشه‌بندی فازی که در آن از معیار فاصله مقاوم به نوفه بهره گرفته شده است، در یک گروه قرار می‌گیرند. فرایند تشخیص خوشه‌های مطلوب به‌وسیلهٔ الگوریتم بهینه‌سازی ذرات انجام و با استفاده از رویکرد تطبیقی جستجوهای محلی و سراسری بهینه می‌شود. استفاده از تابع آشوب نیز به افزایش سرعت هم‌گرایی کمک می‌کند. ساختار مقاله به صورت زیر است: در بخش دوم، مروری بر کارهای مرتبط پیشین مورد بررسی، در بخش سوم، ابزارهای اصلی مورد استفاده ارائه شده و در بخش چهارم، جزئیات روش پیشنهادی نشان داده شده است. نتایج آزمایش‌ها، ارزیابی روش پیشنهادی و مقایسهٔ آن با سایر روش‌ها در بخش پنجم مطرح شده است. نتیجه‌گیری و بیان کارهای آینده در بخش آخر ذکر شده‌است.

۲- مروری بر کارهای مرتبط پیشین

تاکنون روش‌های مختلفی جهت حل سامانه‌های توصیه‌گر ارائه شده‌است. در [۱۴]، سامانه توصیه‌گری با استفاده از الگوریتم C-میانگین فازی (FCM) کاربران را خوشه‌بندی نموده و به حل مشکل شروع سرد و پراکندگی پرداخته است. همچنین، قلم داده‌ای که توسط همسایگان کاربر هدف در خوشه امتیاز بالاتری دریافت کرده به او توصیه می‌شود. در روش یادشده، اندازه‌گیری شباهت با استفاده از ضریب همبستگی پیرسون و محاسبه پیش‌بینی با میانگین وزنی انجام می‌شود. ضعف این روش، حساسیت به انتخاب مراکز اولیه و داده‌های پرت است. در [۱۵] از روش خوشه‌بندی C-میانگین فازی (FCM) برای سامانه توصیه‌گر صفحات وب استفاده شده است. با درنظرگرفتن اطلاعات موجود در الگوهای استخراج شدهٔ کاربران در صفحه‌های وب، توصیه‌هایی به کاربران ارائه می‌شود. در این روش، N خوشهٔ برتر برای پردازش برگزیده می‌شوند. این انتخاب براساس میانگین تعلق کاربر به مراکز خوشه‌ها صورت می‌گیرد. به ازای هر نوع از صفحات وب، وزن‌ها محاسبه و بهترین توصیه برای کاربر هدف، پیش‌بینی می‌شود. مسئله پراکندگی در [۱۶] توسط الگوریتم *k*means بهبودیافته مورد بررسی قرار گرفته است. با بهره‌گیری از فاصله اقلیدسی بهبود یافته، میزان شباهت میان کاربر هدف و همسایگانش در یک خوشه افزایش و عملکرد سامانه بهبود یافته است. در این روش، کاربران فقط می‌توانند در یک خوشه عضویت داشته باشند که باعث می‌شود نرخ صحت امتیاز پیش‌بینی‌شده به‌وسیلهٔ سامانه توصیه‌گر کاهش یابد.

الگوریتم‌های تکاملی و محاسبات زیستی نقش مهمی در بهینه‌سازی الگوریتم‌های توصیه‌گر داشته و درصورت تجمیع با الگوریتم‌های خوشه‌بندی مانند *k*means، FCM و نقشه خودسازماندهی (SOM) روش‌های مناسبی را ارائه می دهند [۱۷]. برای مثال، در [۱۸] از یک رویکرد مبتنی بر تکاملی ژنتیک برای بهینه‌سازی توصیه‌ها بر اساس ترجیحات شخصی استفاده و همچنین، در [۱۹] با استفاده از الگوریتم ژنتیک،

پیشنهادهای مرتبط با سفر ارائه شد. در [۲۰] الگوریتم ترکیبی *k*Means-PSO-FCM برای سامانه‌های توصیه‌گر فیلم ارائه شده‌است. به‌منظور کاهش پیچیدگی محاسباتی، ابتدا مجموعه‌داده براساس انواع فیلم‌های موجود تقسیم‌بندی می‌شود. مجموعه‌داده در این مقاله پس از تقسیم‌بندی براساس ژانر، در ۱۹ بخش جدا قرار گرفته و هر بخش، شامل کاربرانی است که به آن نوع فیلم علاقه‌مندند و امتیاز داده‌اند. هر بخش و ماتریس امتیاز جدید، برای ایجاد مراکز خوشه با استفاده از *k*means و PSO به‌کار برده می‌شوند. با استفاده از الگوریتم *k*means، کاربران موجود در هر فایل بر مبنای امتیازهایشان خوشه‌بندی و k مرکز تصادفی انتخاب می‌شود. همچنین، با استفاده از PSO و تعریف تابع برازش مبتنی بر فاصله بین کاربران، مراکز خوشه‌ها بهینه شده و بهترین موقعیت محلی و عمومی برای کاربران در خوشه‌ها به دست می‌آید. این مراکز بهینه در الگوریتم FCM برای خوشه‌بندی نهایی کاربران مورد استفاده قرار می‌گیرد و درنتیجه کاربران با بیشترین شباهت به یک خوشه اختصاص می‌یابند. در مرحله بعد با توجه به کل امتیازها و براساس فاصله اقلیدسی، خوشهٔ کاربر هدف به‌دست می‌آید و امتیازهای مرکز خوشه یا میانگین کل نمونه‌ها برای آن قلم‌داده به‌عنوان امتیاز کاربر هدف در نظر گرفته می‌شود. سامانه یادشده از حساسیت به انتخاب مراکز اولیه و به دام‌افتادن در کمینه‌های محلی رنج می‌برد. در مراجع [۲۱] ترکیب الگوریتم *k*means و الگوریتم مکاشفه‌ای کلونی زنبور مصنوعی[1] ABC-KM، [۲۲] ترکیب الگوریتم FCM و الگوریتم بهینه‌سازی گرگ‌های خاکستری[2]، [۲۳]ترکیب الگوریتم خوشه‌بندی *k*means با الگوریتم بهینه‌سازی فاخته[3]، [۲۴, ۲۵] الگوریتم بهینه‌سازی خفاش[4] و ترکیب آن با FCM، برای بهبود خوشه‌بندی کاربران، کاهش میزان خطای پیش‌بینی امتیازها و درنتیجه بهینه‌سازی سامانه توصیه‌گر به‌کار برده شده و چالش‌های پراکندگی و مقیاس‌پذیری مورد بررسی قرارگرفته است.

۳- مروری بر ابزارهای مورد استفاده

در این بخش ابزارهای مورد استفاده در روش پیشنهادی معرفی می‌شوند.

۳-۱- الگوریتم خوشه‌بندی C-میانگین مرتب‌شده فازی (FCOM)

یکی از الگوریتم‌های پرکاربرد خوشه‌بندی با کمک تئوری فازی الگوریتم FCM است، هرچند این روش نقاط ضعفی نظیر لزوم مشخص‌بودن تعداد خوشه‌ها در ابتدا، تأثیرگذاری مراکز اولیه تصادفی در گیرافتادن در کمینه‌های محلی و همچنین حساسیت به داده‌های پرت و خطا را دارد. با هدف حل این مشکلات، روش خوشه‌بندی C-میانگین مرتب‌شده فازی (FCOM)[5] [۲۶] معرفی شد. این الگوریتم از معیار فاصله مقاوم به نوفه لگاریتمی زیر که دارای پیچیدگی محاسباتی کم است، استفاده می‌کند:

$$\mathscr{L}_{\mathrm{LOG}}(e) = \begin{cases} 0 & e = 0 \\ log(1+e^2) & e \neq 0 \end{cases} \quad (۱)$$

با فرض دراختیارداشتن n داده D بعدی و c مرکز خوشه، تابع هدف FCOM طبق رابطه (۲) بیان می‌شود. وزن هر داده در FCOM که با $\beta_{ik} \in \{0-1\}$ نشان داده شده است، برحسب فاصله تا مرکز خوشه محاسبه می‌شود. هرچه داده به مرکز خوشه نزدیک‌تر باشد، وزن بیشتر و هرچه دورتر باشد، وزن کمتری می‌گیرد:

$$J_{\mathrm{fcom}}(V.U) = min \sum_{i=1}^{c} \sum_{k=1}^{n} \beta_{\mathrm{ik}} (u_{\mathrm{ik}})^m \mathcal{D}(x_k.v_i) \quad (۲)$$

$$\mathcal{D}(x_k.v_i) = \mathscr{L}(x_k - v_i) = \sum_{d=1}^{D} \{\mathcal{D}(x_{\mathrm{kd}}.v_{\mathrm{id}})\}_{k=1,i=1}^{k=n,i=c} \quad (۳)$$

$$\mathcal{J}_{\mathrm{fcom}} = \left\{ \begin{matrix} U \in R_{\mathrm{cn}} | \forall_{\substack{1\le i\le c \\ 1\le k\le n}} u_{\mathrm{ik}} \in \{0,1\}; \\ \sum_{i=1}^{c} \beta_{\mathrm{ik}} u_{\mathrm{ik}} = f_k; \ 0 < \sum_{k=1}^{n} u_{\mathrm{ik}} < n \end{matrix} \right\} \quad (۴)$$

که در آن m درجه فازی‌سازی، v_{il} مکان تقریبی بعد d ام مرکز خوشه i ام و β_{ik} وزن تخصیص داده‌شده به‌وسیلهٔ خوشه i ام به داده k ام است. با جای‌گذاری رابطه (۳) در (۲)، تابع هدف FCOM به‌صورت زیر بیان می‌شود:

$$\forall_{1\le k\le N} f_k = \beta_{1k} \vee \beta_{2k} \vee \ldots \vee \beta_{\mathrm{ck}} \quad (۵)$$

$$\forall_{\substack{1\le k\le n \\ 1\le i\le c}} u_{\mathrm{ik}} = \frac{f_k \mathcal{D}(x_k.v_i)^{\frac{1}{1-m}}}{\left\{\sum_{j=1}^{c} \beta_{\mathrm{jk}} \mathcal{D}(x_k.v_j)^{\frac{1}{1-m}}\right\}} \quad (۶)$$

[1] Artificial Bee Colony (ABC)
[2] Grey Wolf Optimizer (GWO)
[3] Cuckoo Optimization Algorithm (COA)
[4] BAT Algorithm (BA)
[5] Fuzzy C_Ordered Means (FCOM)

با استفاده از عملگر S-norm، برایند تمام وزن‌هایی که به‌وسلهٔ تمام خوشه‌ها به داده k ام داده می‌شود، محاسبه می‌شود:

میزان تعلق داده k ام به خوشه i ام با (۷) تعیین می‌شود:

در تکرار t، بعد d ام از مرکز خوشه i ام توسط (۷) محاسبه می‌شود:

با درنظرگرفتن تابع جای‌گشت $\pi:\{1.2 \dots . n\} \rightarrow \{1.2 \dots . n\}$، شرط زیر برقرار است:

که در آن، مقدار $e_{i\pi(j)l}^{\{t-1\}}$ فاصله داده $\pi(j)$ ام از مرکز خوشه i ام با توجه به بعد d ام است که در تکرار $\{t-1\}$ ام محاسبه شده است.

با جایگزینی β_{ik} با مقدار α که از رابطه (۱۱) به دست می‌آید و در نظر گرفتن رابطه (۱۰)، چنانچه شرط $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ برقرار باشد، تأثیر داده‌های پرت با وزن کم نسبت داده‌شده کاهش می‌یابد:

$$\alpha_k = \left\{\left(\frac{p_c n - k}{2p_l n} + 0.5\right) \wedge 1\right\} \vee 0 \qquad (۱۱)$$

که در آن، $p_l = 0.2$ و $p_c = 0.5$ است.

بنابراین، روابط (۷) و (۸) به‌صورت (۱۲) و (۱۴) بازنویسی می‌شوند:

$$\underset{\substack{1\leq d\leq D\\ 1\leq i\leq c}}{\forall} v_{\mathrm{il}}^{(t)} = \frac{\sum_{k=1}^{n} \hat{\alpha}_k (u_{\mathrm{ik}})^m h_{\mathrm{ikd}}^{(t)} x_{\mathrm{kd}}}{\sum_{k=1}^{n} \hat{\alpha}_k (u_{\mathrm{ik}})^m h_{\mathrm{ikd}}^{(t)}} \qquad (۱۲)$$

$$\hat{\alpha}_k = \alpha_{\pi(\pi^{-1}(k))} = \alpha_k \qquad (۱۳)$$

$$h_{\mathrm{ikd}}^{(t)} = \begin{cases} 0 & x_{\mathrm{kd}} - v_{id}^{(t-1)} = 0 \\ \dfrac{\mathscr{L}\left(x_{\mathrm{kd}} - v_{id}^{(t-1)}\right)}{\left(x_{\mathrm{kd}} - v_{id}^{(t-1)}\right)^2} & x_{\mathrm{kd}} - v_{id}^{(t-1)} \neq 0 \end{cases} \qquad (۱۴)$$

جزئیات این روش به تفضیل در مرجع [۲۶] آمده است. به‌طور خلاصه، گام‌های FCOM به‌صورت ذیل است:

گام ۱– دادن مقادیر اولیه به مراکز خوشه‌ها (ابعاد مختلف) به‌صورت $v_{id}^{(0)} = 0$ و تنظیم‌کردن اندیس تکرار $t = 1$.

گام ۲– محاسبه فواصل داده‌ها از مراکز– $e_{ikd}^{(t-1)} = x_{kd}$ $v_{id}^{(t-1)}$ و ضرایب $h_{ikd}^{(t)}$ با استفاده از رابطه (۸) برای همه داده‌ها.

گام ۳– مرتب‌کردن تابع $\pi(k)$ به‌صورت صعودی طبق رابطه (۱۰).

گام ۴– محاسبه ضرایب α_k با استفاده از عبارت (۱۱).

گام ۵– به‌روزرسانی مراکز با استفاده از عبارت (۱۲).

گام ۶– اگر $\left\| v_{id}^{(t)} - v_{id}^{(t-1)} \right\|_2^2 > \varepsilon$ است، پس $t = t+1$ و رفتن به مرحلهٔ ۲، درغیراین‌صورت $\beta_{ik} = \widehat{\alpha_k}$ و خاتمه الگوریتم.

۲-۳- الگوریتم بهینه‌سازی ذرات تطبیقی آشوبی

یکی از الگوریتم‌های شناخته‌شده محاسبات زیستی، الگوریتم ازدحام ذرات[1] است که ابزاری مناسب برای

$$\underset{\substack{1\leq d\leq D\\ 1\leq i\leq c}}{\forall} v_{\mathrm{il}}^{(t)} = \frac{\sum_{k=1}^{n} \beta_{\mathrm{ik}} (u_{\mathrm{ik}})^m h_{\mathrm{ikd}}^{(t)} x_{\mathrm{kd}}}{\sum_{k=1}^{n} \beta_{\mathrm{ik}} (u_{\mathrm{ik}})^m h_{\mathrm{ikd}}^{(t)}} \qquad (۷)$$

$$h_{\mathrm{ikd}}^{(t)} = \begin{cases} 0 & e_{\mathrm{ikd}}{}^{(t-1)} = 0 \\ \dfrac{\mathscr{L}\left(e_{\mathrm{ikd}}{}^{(t-1)}\right)}{\left(e_{\mathrm{ikd}}{}^{(t-1)}\right)^2} & e_{\mathrm{ikd}}{}^{(t-1)} \neq 0 \end{cases} \qquad (۸)$$

$$e_{\mathrm{ikl}}{}^{(t-1)} = x_{\mathrm{kl}} - v_{il} \qquad (۹)$$

حل مسائل بهینه‌سازی است [۲۷]. مراحل الگوریتم عبارتند از:

گام ۱– مقداردهی اولیه یک جمعیت از ذرات با موقعیت‌ها و سرعت‌های تصادفی در ابعاد D در

$$\left|e_{i\pi(1)d}^{\{t-1\}}\right| \leq \left|e_{i\pi(2)d}^{\{t-1\}}\right| \leq \left|e_{i\pi(3)d}^{\{t-1\}}\right| \dots . \leq \left|e_{i\pi(n)d}^{\{t-1\}}\right| \qquad (۱۰)$$

فضای جستجو؛

$$v_{id}^{t+1} = w_t v_{id}^t + m_1 r_{1d}(p_{id} - x_{id}^t) + m_2 r_{2d}(p_{gd} - x_{id}^t) \qquad (۱۵)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \qquad (۱۶)$$

گام ۲– ارزیابی برازش هر ذره از جمعیت توسط تابع هدف تعریف‌شده؛

گام ۳– ارزیابی بهترین تجربه شخصی هر ذره از طریق مقایسه برازش آن در تکرار فعلی با بهترین برازش قبلی. چنانچه مقدار موقعیت فعلی مناسب‌تر است، بهترین تجربه ذره را به‌روز کن؛

گام ۴– ارزیابی بهترین تجربه گروهی از طریق شناسایی ذره‌ای که بهترین مقدار برازش را داشته است. موقعیت این ذره به‌عنوان بهترین تجربه گروهی در نظر گرفته می‌شود؛

گام ۵– به‌روزرسانی سرعت هر ذره با عبارت (۱۵) و موقعیت با استفاده از عبارت (۱۶).

که در آن v_t سرعت در تکرار t، w وزن اینرسی جهت کنترل جستجو (وزن‌های کمتر برای جستجوی محلی و

[1] Particle swarm optimization (PSO)

وزن‌های بیشتر برای جستجوی سراسری تنظیم می‌شوند)، r_1 و r_2 اعداد تصادفی، p_i بهترین مکان شخصی ذره i ام، p_g بهترین مکان گروهی و x_t موقعیت ذره در تکرار فعلی هستند. همچنین، m_1 و m_2 اعداد تصادفی بوده و میزان اهمیت و وزن تجربه شخصی و گروهی را مشخص می‌کنند.

گام ۶- توقف الگوریتم درصورت رسیدن به شرط توقف ازپیش تعیین‌شده؛ درغیراین صورت، رفتن به گام ۲.

در مقاله‌های [۲۸, ۲۹] الگوریتم بهینه‌سازی ازدحام ذرات تطبیقی[1] (APSO) با هدف بهبود جستجوهای سراسری[2] و محلی[3] معرفی شده است. در نسل نخست این روش، ذرات به‌صورت تصادفی انتخاب شده و جواب‌های دقیقی برای مقادیر بهینهٔ محلی و سراسری وجود نداشته؛ بنابراین جواب‌ها در همان مسیر تصادفی اولیه به‌روزرسانی شده و کم‌تر از تجربیات به‌دست‌آمده استفاده می‌کنند. با تکامل نسل‌ها می‌توان با استفاده از رابطه (۱۷) از جواب‌های تصادفی اولیه فاصله گرفت و جواب‌ها را بر اساس بهینه‌های محلی و سراسری به‌دست آورد:

$$w_t = w_{max} - t\left(\frac{w_{max} - w_{min}}{K_{max}}\right) \qquad (۱۷)$$

که در آن، K_{max} بیشینه تعداد نسل‌ها، t اندیس نسل کنونی و w_{max}=۱ و w_{min}=۰.۱ است.

۴- روش پیشنهادی

هدف روش پیشنهادی بهبود کارایی سامانه توصیه‌گر در مواجهه با مجموعه‌داده‌های با مقیاس و پراکندگی زیاد است؛ ازاین‌رو، الگوریتم بهینه‌سازی ازدحام ذرات تطبیقی آشوبی[4] (CAPSO) معرفی و در سامانه توصیه‌گر پیشنهادی به‌کار گرفته شده است. در ادامه، جزییات روش پیشنهادی ارائه می‌شود.

۴-۱- پیش‌پردازش داده‌های اولیه

ماتریس امتیازات دارای پراکندگی زیادی است؛ زیرا به‌طورمعمول کاربران تنها به بخش اندکی از قلم‌کالاهای خریداری‌شده امتیازدهی کرده، بنابراین اطلاعات کافی و لازم برای ارائه پیشنهاد در سامانه موجود نیست. جهت حل این مشکل، روش پیشنهادی به‌صورت زیر قلم‌کالای j را که توسط کاربر هدف امتیازدهی نشده است، امتیازدهی می‌کند:

ابتدا تعداد M همسایه کاربر هدف شناسایی می‌شوند. بدین منظور، میزان شباهت دو کاربر x و y با استفاده از ضریب همبستگی پیرسون محاسبه می‌شود:

$$\text{sim } x,y = \frac{\sum_j r_{x,j} - \bar{r}_x \times r_{y,j} - \bar{r}_y}{\sqrt{\sum_j (r_{x,j} - \bar{r}_x)^2 \times \sum_j (r_{y,j} - \bar{r}_y)^2}} \qquad (۱۸)$$

که در آن $r_{x,j}$ امتیازی است که کاربر x به قلم‌کالای j داده است. $\bar{r}_x$ میانگین کل امتیازهایی است که کاربر x به قلم‌کالاها داده است.

پس از آن، کاربران همسایه به‌ترتیب مقدار شباهتشان با کاربر هدف به‌صورت نزولی مرتب شده و تعداد M کاربر با بیشترین مقدار sim به عنوان نزدیک‌ترین همسایگان کاربر هدف در نظر گرفته می‌شوند؛ پس از آن، میانگین امتیازات کاربران همسایه طبق رابطه (۱۹) محاسبه و به‌عنوان امتیاز قلم‌کالای j در نظر گرفته می‌شود:

$$Rate_{feature\ User_i, feature_j} = \frac{\sum_{Item_k \in feature_j}^{M} rating\ User_i, Item_k}{t_{Ni}} \qquad (۱۹)$$

که t_{Ni} تعداد قلم‌کالاهایی که در نوع j، کاربر i امتیاز داده است.

۴-۲- خوشه‌بندی کاربران با ماتریس نرمال امتیازها

خوشه‌بندی کاربران با FCOM انجام می‌شود. از آنجا که بررسی هم‌زمان چند جواب، به کاهش تعداد تکرارها در رسیدن به جواب بهینه کمک می‌کند، روش پیشنهادی از ایده جستجوی مکاشفه‌ای الگوریتم APSO جهت بررسی مجموعه‌جواب‌ها بهره می‌برد؛ بنابراین، خوشه‌بندی کاربران تکرارشونده بوده و در هر تکرار آن مراکز خوشه‌ها و ماتریس میزان تعلق کاربران به خوشه‌ها به‌روزرسانی می‌شود.

برای مقداردهی اولیه ذرات از عبارت (۲۰) استفاده می‌شود:

$$x_{i,j}(0) = x_{min,j} + \zeta_j\left(x_{max,j} - x_{min,j}\right) \quad \forall j = 1, \dots, n_x, \forall i = 1, \dots, n_s \qquad (۲۰)$$

x_{max} و x_{min} بیشترین و کمترین مقدار هر بعد و ζ یک عدد تصادفی در بازه [۰,۱] است. بردارهای سرعت اولیه را می‌توان با عدد صفر مقداردهی کرد؛ سپس با استفاده از الگوریتم FCOM و رابطه (۷)، مقادیر اولیه میزان تعلق

[1] Adaptive Particle Swarm Optimization (APSO)
[2] Exploration
[3] Exploitation
[4] Chaotic Adaptive particle swarm optimization

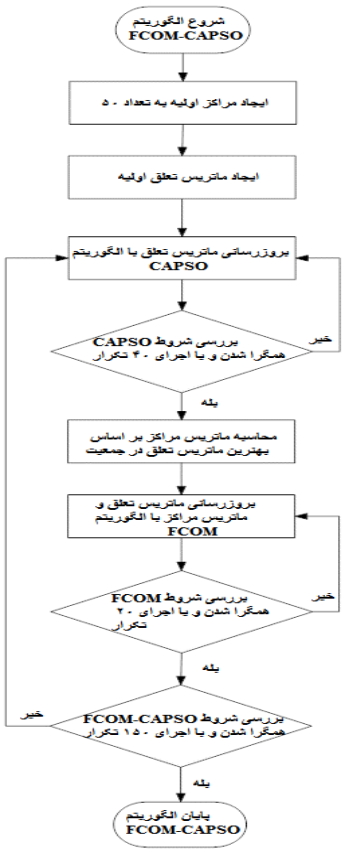
نمونه‌ها که عددی در بازهٔ [۰,۱] است محاسبه می‌شود. به‌روزرسانی جواب‌ها با تکرارهای پی‌درپی در کل الگوریتم به تعداد تکرار ثابت و مشخص صورت می‌پذیرد و در هر مرحله ماتریس مراکز و ماتریس تعلق به‌روزرسانی می‌شود. در هر تکرار، ابتدا میزان تعلق کاربران به‌وسیله الگوریتم CAPSO به‌روزرسانی شده تا شرط پایانی آن (بیشینه تکرار الگوریتم یا عدم‌تغییر جواب‌ها) برقرار شود. با استفاده از (۷)، ماتریس تعلق به‌ازای مراکز خوشه مرتباً به‌روزرسانی می‌شود. مراکز خوشه در طی اجرای این الگوریتم به ازای هر بار اجرا ثابت هستند.

نکته مهم آن است که در رابطه (۱۵) از اعداد تصادفی استفاده شده است و این امکان وجود دارد که مراکز اولیه بسیار پراکنده شده و در محدودهٔ داده‌ها واقع نشوند. بر اساس نظریه آشوب می‌توان رفتارهای آشوبی را جایگزین رفتارهای تصادفی کرد [۳۰, ۳۱]. آشوب یک رفتار قطعی شبه‌تصادفی است که تکرار در آن وجود ندارد. می‌توان اعداد تصادفی مورد استفاده در الگوریتم تکاملی را با توابع آشوب که با شرایط اولیه تصادفی شروع به کار می‌کنند، جایگزین کرد. از دلایل برتری این رویکرد، تولید داده‌هایی با پراکندگی بیشتر در فضای جستجو و جستجو در مرزها است (برخلاف سامانه تصادفی یکنواخت که کل فضای جستجو به‌جز نواحی مرزی را پوشش می‌دهند). همچنین، موجب تسریع در هم‌گرایی الگوریتم به جواب بهینه می‌شود. با درنظرگرفتن این موضوع، رابطه (۱۵) به‌صورت رابطه (۲۱) بازنویسی شده، همچنین طبق رابطه (۲۲) [۳۲] ضرایب آشوب c_1 و c_2 تولید می‌شوند:

$$v_{i+1}^d = w_i v_i^d + c_{1i} r_1^d \left(x_{p_1}^d - x_i^d\right) + c_{2i} r_2^d \left(x_g^d - x_i^d\right) \quad (۲۱)$$

$$c_{1i}, c_{2i} = \frac{sin(2n\pi x_i - 1)}{2} \quad (۲۲)$$

مقدار w در طی اجرا از ۱ تا ۰ کاهش می‌یابد و از طریق عبارت (۱۸) مقداردهی می‌شود. در مقادیر بیشتر، الگوریتم به جمعیت‌ها اجازه اکتشاف می‌دهد و در مقادیر کمتر، شانس برای رسیدن به مقادیر بهینهٔ محلی افزایش می‌یابد [۲۸]. به‌دلیل لزوم برقراری شروط منطق فازی در ماتریس مذکور، چنان‌چه در حین به‌روزرسانی مقادیر میزان تعلق خارج از بازه [۰,۱] گردند، نرمال‌سازی انجام می‌شود. پس از پایان حلقه مربوط به الگوریتم CAPSO، ماتریس مراکز با استفاده از عبارت (۸) به‌دست می‌آید. به‌روزرسانی در الگوریتم FCOM با عبارت‌های محاسبه مراکز و میزان تعلق ذرات انجام می‌شود. در این مرحله نیز با استفاده از تابع برازش، مراکز خوشه‌ها و میزان تعلقشان مورد بررسی قرار می‌گیرند. چنان‌چه تغییری در مراکز خوشه ایجاد نشود، الگوریتم FCOM به پایان می‌رسد. در انتها، درصورت برقرارنبودن شرط (بیشینه تعداد تکرار)، ماتریس تعلق کاربران دوباره به الگوریتم CAPSO ارسال و دوباره دو الگوریتم به‌صورت تکرارشونده اجرا می‌شوند. در پایان خروجی الگوریتم خوشه‌بندی ماتریس مراکز و ماتریس میزان تعلق کاربران به خوشه‌های بهینه به‌گونه‌ای به‌دست می‌آید که تابع هزینه FCOM کمترین مقدار را داشته باشد. درنتیجه کاربران با بیشترین شباهت در خوشه‌ها قرارمی‌گیرند. ماتریس تعلق یک ماتریس $C \times U$ است، که در آن C مراکز خوشه‌ها و U کاربران است. هر درایهٔ این ماتریس بیان‌گر میزان تعلق کاربر به خوشه است. در روش پیشنهادی با استفاده از بیشینه میزان تعلق، خوشه‌ای که کاربر عضو آن است، شناسایی می‌شود. همسایگان او کاربرانی هستند که آن‌ها نیز در همان خوشه قرار گیرند. در گام بعدی روش پیشنهادی، از روش پالایش مشارکتی برای ادامه مراحل سامانه توصیه‌گر استفاده شده است. شکل (۱) روندنمای خوشه‌بندی الگوریتم پیشنهادی را نمایش می‌دهد:



(شکل-۱) روندنمای خوشه‌بندی کاربران در روش پیشنهادی

(Figure-1) The flowchart of the clustering stage in the proposed method

۴-۴- پیش‌بینی امتیازهای کاربر هدف

هدف از این مرحله، پیش‌بینی امتیازدهی کاربر هدف به قلم‌داده هدف است. روش پیشنهادی برای پیش‌بینی امتیاز مقداردهی نشده از (۲۳) استفاده می‌کند:

nn تعداد کاربران هم‌خوشه در همسایگی کاربر هدف i، $\text{Ratings}_{y,i}$ امتیاز کاربر y به قلم‌داده i و $\overline{Ratings}$ میانگین امتیازهای کاربر است. شناسایی کاربران با بیشترین شباهت در هر خوشه، توسط ضریب همبستگی پیرسون انجام می‌شود [۳۳].

۵-۴- ارائه فهرستی از توصیه‌ها برای کاربر هدف

جهت ارائه فهرستی از اقلام پیشنهادی، N قلم‌کالا از قلم‌کالاهایی که برای کاربر هدف بیشترین امتیاز پیش‌بینی‌شده را دارند توصیه می‌شود. چنانچه ماتریس امتیاز قلم‌کالا حجیم و پراکنده بوده و بر اساس یک ویژگی از قلم‌کالاها متراکم شده تا کاهش حجم صورت گرفته باشد، می‌توان امتیاز پیش‌بینی‌شده برای هر نوع از آن ویژگی، به‌عنوان‌مثال هر نوع فیلم را برای فیلم‌هایی که در آن نوع وجود دارند، در نظر گرفت و برای قلم‌کالاهایی که در بیش از یک نوع قرار دارند، میانگین را محاسبه کرد. در پایان، N قلم‌کالا با بالاترین امتیاز به کاربر هدف، توصیه می‌شود.

۵- ارزیابی روش پیشنهادی

آزمایش‌ها در و محیط Matlab و بر روی سامانه‌ای با شانزده گیگابایت رم و پردازنده Core i7 1255U انجام شده است. روش ارزیابی در همه آزمایش‌ها مبتنی بر روش اعتبارسنجی متقاطع ده‌بخشی[1] بوده است. در آزمایش‌ها مقدار m در الگوریتم FCOM برابر با دو و شرط خاتمه رسیدن به هم‌گرایی و یا اجرای بیشینه بیست تکرار است. در الگوریتم CAPSO تعداد ذرات پنجاه و شرط خاتمه رسیدن به شرط هم‌گرایی و یا و یا اجرای بیشینه چهل قرار داده شده است. شرط خاتمه کل الگوریتم ترکیبی، رسیدن به هم‌گرایی یا اجرای بیشینه ۱۵۰ است.

۵-۱- مشخصات مجموعه‌داده‌ها

برای انجام آزمایش‌ها و مقایسه روش پیشنهادی با سایر روش‌ها، از مجموعه‌داده‌های Movielenz[2]، Book-Crossing[3] و Jester[4] استفاده شده است. مشخصات مجموعه‌داده‌های یادشده در جدول (۱) بیان شده است.

۵-۲- معیارهای ارزیابی

به‌منظور ارزیابی عملکرد روش پیشنهادی میانگین مطلق خطا (۲۵) و خطای جذر میانگین مربعات (۲۶) و نرخ صحت (۲۷) محاسبه شده‌اند.

$$\text{MAE} = \frac{\sum_{x,i} \left|\text{Pred}_{x,i} - \text{r}_{x,i}\right|}{\text{N}} \qquad (۲۵)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{x,i} \text{Pred}_{x,i} - \text{r}_{x,i}^{\,2}}{\text{N}}} \qquad (۲۶)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (۲۷)$$

که در آن TP نشان‌دهنده آن است که کالا درخواست شده و سامانه آن را به‌درستی پیشنهاد داده، TN نشان‌دهنده آن است که کالا درخواست نشده و سامانه به‌درستی آن را پیشنهاد نداده است، FP نشان‌دهنده آن است که کالا درخواست نشده، اما سامانه آن را به‌اشتباه پیشنهاد داده است، FN نشان‌دهنده آن است که کالا درخواست شده اما سامانه آن را به‌اشتباه پیشنهاد نداده است.

(جدول- ۱) مشخصات مجموعه‌داده‌ها

(Table -1) The characteristics of datasets

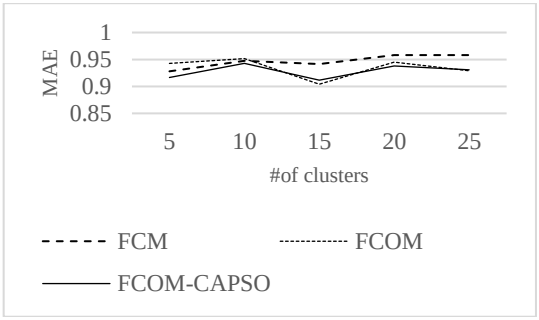
نام مجموعه داده	کاربران	اقلام	امتیازها	پراکندگی	امتیاز
MovieLens(100K)	۹۴۳	۱۶۸۲	۱۰۰۰۰۰	۹۴٪	۵-۱
MovieLens1M	۶۰۴۰	۳۹۰۰	۱۰۰۰۲۰۹	۹۳٪	۵-۱
Book-Crossing	۲۷۸۸۵۸	۲۷۱۳۷۹	۱۱۴۹۷۸۰	۹۸٪	۱۰-۱
Jester	۲۴۹۳۸	۱۰۰	۶۲۳۴۵۰	۷۵٪	۱۰ تا -۱۰

۵-۳- نتایج و تفسیر آن‌ها

به‌منظور بررسی تأثیر تعداد خوشه‌ها بر معیارهای ارزیابی میانگین مطلق خطا، جذر میانگین مربعات و نرخ صحت الگوریتم پیشنهادی FCOM-CAPSO با دو الگوریتم FCM و FCOM با تعداد خوشه‌های متفاوت (۵،۱۰،۱۵،۲۰،۲۵) مقایسه گردید. نتایج در شکل‌های ۲-۴ مشاهده می‌شود. در تکرارهای متوالی از اجرای الگوریتم‌ها، روش پیشنهادی به‌طور میانگین در تعداد خوشه‌های متفاوت بهتر عمل کرده و موجب بهبود کارایی سامانه می‌شود. نرخ صحت روش پیشنهادی با تعداد تکرار بیشتر و افزایش تعداد خوشه‌ها افزایش می‌یابد.

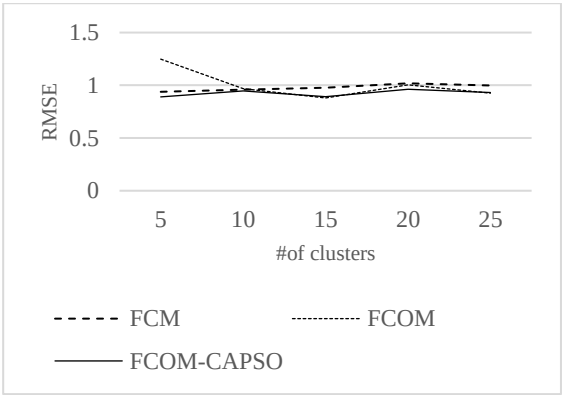
[1] K-Fold Cross Validation
[2] https://grouplens.org
[3] http://www2.informatik.uni-freiburg.de/~cziegler/BX
[4] http://www.ieor.berkeley.edu

با توجه به استفاده از مقداردهی اولیه تصادفی در این سه الگوریتم مشاهده می‌شود که الگوریتم FCM به‌دلیل حساسیت آن به انتخاب مراکز اولیه، عملکرد ضعیف‌تری دارد. همچنین از آن‌جا که FCOM امکان به دام افتادن در بهینه‌های محلی را دارد، نسبت به انتخاب مراکز اولیه حساس است. علاوه‌بر آن، از مزایای الگوریتم‌های تکاملی و همچنین تئوری آشوب بهره‌مند نیست؛ بنابراین نسبت به الگوریتم FCOM-CAPSO نتایج ضعیف‌تری دارد.



شکل۲-میانگین مطلق خطا با تعداد خوشه‌های متفاوت

Figure 2- MAE results for varying cluster numbers



(شکل-۳) خطای جذر میانگین مربعات با تعداد خوشه‌های متفاوت

(Figure -3) RMSE results for varying cluster numbers.

در ادامه نتایج آزمایش‌ها در پیش‌بینی امتیاز کاربر هدف به‌طور میانگین در پنج‌بار تکرار و انتخاب تعداد خوشه ثابت پانزده درمقایسه با روش‌های دیگر ارائه شده است.

در جدول (۲) مقایسه میانگین مطلق خطا، در جدول (۳) مقایسه جذر میانگین مربعات خطا و در جدول (۴) مقایسه نرخ صحت بین روش‌های رقیب [۱۴, ۱۶, ۲۰, ۲۲] و روش پیشنهادی بر روی مجموعه‌داده‌های MovieLens100k، MovieLens1M، Book-Crossing و jester نمایش داده شده است. مشاهده می‌شود که روش پیشنهادی به‌دلیل بهره‌گیری از الگوریتم FCOM که مقاوم به نوفه و داده های پرت است، کمترین میزان خطا در تمامی مجموعه‌داده‌ها را نشان می دهد.

(جدول-۲) میانگین مطلق خطا

(Table -2) MAE

مجموعه داده / روش	MovieLens 100k	MovieLens 1M	Book-Crossing	Jester
UBCF [14]	۰.۹۶	۰.۹۷۰۱	۰.۹۴۷۱	۰.۹۵۲
KMCF [16]	۰.۹۵۱۰	۰.۹۶۳۲	۰.۹۴۲۹	۰.۹۴۴۱
FCM [20]	۰.۹۴۶۶	۰.۹۳۱۲	۰.۹۴۰۲	۰.۹۲۹
FCM-PSO [22]	۰.۹۳۶۸	۰.۹۴۷۲	۰.۹۳۹۱	۰.۹۱۸۷
FCOM_CAPSO	**۰.۸۸۱۱**	**۰.۸۴۲۸**	**۰.۸۶۱۶**	**۰.۸۰**

(جدول- ۳) جذر میانگین مربعات خطا

(Table- 3) RMSE

مجموعه داده / روش	MovieLens 100k	MovieLens 1M	Book-Crossing	Jester
UBCF [14]	۰.۴۴	۰.۴۲۵۴	۰.۴۰۲۴	۰.۵
KMCF [16]	۰.۴۵۲۱	۰.۴۳۲۰	۰.۴۴۳۰	۰.۵۶۱
FCM [20]	۰.۴۸	۰.۴۷۱۲	۰.۴۷۶۹	۰.۵۹۸
FCM-PSO [22]	۰.۴۹۷۸	۰.۴۸۱۱	۰.۴۸۲۱	۰.۶۱
FCOM_CAPSO	**۰.۵۲۴۰**	**۰.۶۱۴۵**	**۰.۵۱۲۶**	**۰.۶۸۶**

(جدول-۴) نرخ صحت

(Table- 4)Accuracy

مجموعه داده / روش	MovieLens 100k	MovieLens 1M	Book-Crossing	Jester
UBCF [14]	۱.۰۰۸۰	۱.۱۴۲۹	۱.۱۳۰۴	۱.۰۰۱
KMCF [16]	۰.۹۷۸۱	۰.۹۸۷۱	۰.۹۹۲۳	۰.۹۷۴۱
FCM [20]	۰.۹۶۹۱	۰.۹۸۶۰	۰.۹۸۷۹	۰.۹۶
FCM-PSO [22]	۰.۹۵	۰.۹۷۵۵	۰.۹۸۰۲	۰.۹۴۶
FCOM_CAPSO	**۰.۸۹**	**۰.۹۰**	**۰.۸۸۱**	**۰.۸۴**

(جدول- ۵) زمان محاسباتی (ثانیه)

(Table- 5) Computational time (s)

مجموعه داده / روش	MovieLens 100k	MovieLens 1M	Book-Crossing	Jester
UBCF [14]	۱۰۷.۳۳	۱۰۹.۹۲	۱۱۰.۴۷	۱۰۷.۴۴
KMCF [16]	۸۰.۸۶	۸۱.۸۵	۷۴.۶۹	۸۸.۶۹
FCM [20]	۶۷.۱۹	**۶۸.۶۲**	۶۹.۸۳	۷۵.۳۶
FCM-PSO [22]	۸۳.۰۴	۸۴.۶۳	۸۱.۵	۸۴.۴۵
FCOM_CAPSO	**۶۳.۰۳**	۷۴.۲۵	**۶۵.۰۱**	**۶۳.۲۷**

همان‌طور که جداول (۳) و (۴) نشان می‌دهند، میزان خطا در روش پیشنهادی کاهش یافته است. استفاده از روش‌های فازی و بهره‌گیری از ماهیت آن که کاربران می‌توانند به خوشه‌های مختلفی شباهت داشته باشند، باعث بهبود در انتخاب همسایگان شده است. بهره‌مندی از روش تکاملی و خوشه‌بندی فازی به‌طور هم‌زمان اثر داده‌های پراکنده، نوفه و دادهٔ پرت را برای حجم زیاد داده‌ها در ماتریس امتیاز کاهش می‌دهد و از به‌دام‌افتادن در بهینه‌های محلی جلوگیری می‌کند؛ بنابراین میزان خطای سامانه کاهش و نرخ صحت و کارایی آن بهبود یافته است. با توجه به استفاده از مقداردهی اولیه تصادفی و حساسیت الگوریتم FCM به انتخاب مراکز اولیه، عملکرد ضعیف‌تری از آن مشاهده می‌شود. در پایگاه داده jester با پراکندگی کمتر، میزان خطا کاهش و نرخ صحت پیش‌بینی افزایش می‌یابد. روش پیشنهادی با بهره‌مندی از معیار فاصله لگاریتمی مقاوم به نوفه، توانسته است مرحله یافتن کاربران همسایه را برخلاف پراکندگی و مقیاس بالای داده‌ها در ماتریس امتیاز بهبود دهد. در روش پیشنهادی، امتیازهای پیش‌بینی‌شده برای کاربران با امتیازهای موجود تفاوت کمتری دارد و نرخ صحت نسبت به روش‌های موجود بهبود یافته است؛ به‌علاوه، نتایج حاصل از زمان محاسباتی بر حسب ثانیه در جدول (۵) آورده شده است. همان‌طور که مشاهده می‌شود، روش پیشنهادی در همه مجموعه‌داده‌ها به جز MovieLens1M به‌دلیل استفاده از دو رویکرد مهم آشوبی (به جای استفاده از اعداد تصادفی) و همچنین جستجوی مکاشفه‌ای کمترین زمان پردازش را صرف نموده است.

۶- نتیجه‌گیری و کارهای آینده

در روش پیشنهادی با هدف بهبود کارایی در سامانه‌های توصیه‌گر، روش پالایش مشارکتی مبتنی بر تئوری فازی معرفی گردیده است که با بهره‌گیری از الگوریتم تکاملی ازدحام ذرات تطبیقی-آشوبی، فضای جستجو را بهبود داده و همزمان چند جواب احتمالی را مورد بررسی قرار می‌دهد. این رویکرد به افزایش سرعت منجر شده و می‌تواند به حل چالش‌های مقیاس‌پذیری و پراکندگی در آن‌ها بپردازد. در روش پیشنهادی که ترکیبی از خوشه‌بندی C-میانگین فازی مرتب‌شده و الگوریتم تکاملی ازدحام ذرات تطبیقی-آشوبی است، پس از انجام پردازش اولیه داده‌ها، کاربران خوشه‌بندی شده و کاربران مشابه با کاربر هدف شناسایی شده و امتیاز به قلم کالا پیش‌بینی می‌شود. علاوبراین، امکان ارائه فهرستی از اقلام پیشنهادی به کاربر هدف وجود دارد. با به‌کارگیری معیارهای ارزیابی متداول، روش پیشنهادی با چهار روش رقیب مقایسه گردید. نتایج بیان‌گر بهبود نرخ صحت توسط روش پیشنهادی بوده است. همچنین، میزان خطای پیش‌بینی امتیازهای نامشخص برای کاربر هدف، کاهش یافته است. بهره‌مندی از مزایای روش‌های خوشه‌بندی فازی و ترکیب آن با الگوریتم تکاملی، به نتایج بهتر و بهبود کارایی سامانه توصیه‌گر منجر شده است.

به‌عنوان اقدامات آینده، می‌توان به استفاده از اطلاعات اضافی کاربران نظیر ارتباطات در شبکه‌های اجتماعی جهت ارائه پیشنهادهای نزدیکتر به سلایق کاربر هدف، ارائه راه‌کارهای بهینه‌تر جهت انتخاب همسایگان و همچنین مسئله شروع سرد ارائه نمود.

۷- مراجع 7-Reference

[1] P. Lops, M. Polignano, C. Musto, A. Silletti, and G. Semeraro, "ClayRS: An end-to-end framework for reproducible knowledge-aware recommender systems," Information Systems, vol. 119, p. 102273, 2023, doi: https://doi.org/10.1016/ j.is.2023.102273.

[2] J.-C. Zhang, A. M. Zain, K.-Q. Zhou, X. Chen, and R.-M. Zhang, "A review of recommender systems based on knowledge graph embedding," Expert Systems with Applications, vol. 250, p. 123876, 2024, doi: https://doi.org/10.1016 /j.eswa.2024.123876.

[3] A. Ghannadrad, M. Arezoumandan, L. Candela, and D. Castelli, "Recommender systems for science: A basic taxonomy," 2022, pp. 1-8.

[4] K. Saini and A. Singh, "A content-based recommender system using stacked LSTM and an attention-based autoencoder," Measurement: Sensors, p. 100975, 2023, doi: https://doi.org/ 10.1016/ j.measen.2023.100975.

[5] P. B. Thorat, R. Goudar, and S. Barve, "Survey on collaborative filtering, content-based filtering and hybrid recommendation system," International Journal of Computer Applications, vol. 110, no. 4, 2015.

[6] M. Moradi and J. Hamidzadeh, "Ensemble-based Top-k Recommender System Considering Incomplete Data," Journal of AI and Data Mining, 2019.

[7] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," Advances in artificial intelligence, vol. 2009, 2009.

[8] C. C. Aggarwal, Recommender systems. Springer, 2016.

[9] p. bahrani, B. Minaei Bidgoli, H. Parvin, M. Mirzarezaee, and A. Keshavarz, "An Ontological Hybrid Recommender System for Dealing with Cold Start Problem," jsdp, vol. 19, no. 1, pp. 1-18, 2022, doi: 10.52547/jsdp.19.1.1.

[10]R. Yera and L. Martinez, "Fuzzy tools in recommender systems: A survey," International Journal of Computational Intelligence Systems, vol. 10, no. 1, pp. 776-803, 2017.
[11]E. G. Muñoz, J. Parraga-Alava, J. Meza, J. J. Proaño Morales, and S. Ventura, "Housing fuzzy recommender system: A systematic literature review," Heliyon, vol. 10, no. 5, p. e26444, 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e26444.
[12] F. Houshmand-Nanehkaran, S. M. Lajevardi, and M. Mahlouji-Bidgholi, "Optimization of fuzzy similarity by genetic algorithm in user-based collaborative filtering recommender systems," Expert Systems, vol. 39, no. 4, p. e12893, 2022, doi: https://doi.org/10.1111/exsy.12893.
[13]m. robati anaraki and n. riahi, "An evolutionary approach for automating the selection of optimum Algorithm in Collaborative Filtering Recommender Systems," jsdp, vol. 20, no. 1, pp. 59-78, 2023. [Online]. Available: http://jsdp.rcisp.ac.ir/article-1-1206-en.html.
[14]H. Koohi and K. Kiani, "User based Collaborative Filtering using fuzzy C-means," Measurement, vol. 91, pp. 134-139, 2016.
[15] R. Katarya and O. P. Verma, "An effective web page recommender system with fuzzy c-mean clustering," Multimedia Tools and Applications, vol. 76, no. 20, pp. 21481-21496, 2017.
[16] X. Li and D. Li, "Improved Hybrid Collaborative Filtering Algorithm Based on K-Means," in International Conference on Applications and Techniques in Cyber Security and Intelligence, 2018: Springer, pp. 928-934.
[17] N. K. Manikandan and M. Kavitha, "A content recommendation system for e-learning using enhanced Harris Hawks Optimization, Cuckoo search and DSSM," Journal of Intelligent & Fuzzy Systems, no. Preprint, pp. 1-14, 2023.
[18] B. Alhijawi and Y. Kilani, "A collaborative filtering recommender system using genetic algorithm," Information Processing & Management, vol. 57, no. 6, p. 102310, 2020, doi: https://doi.org/10.1016/j.ipm.2020.102310.
[19] R. Paulavičius, L. Stripinis, S. Sutavičiūtė, D. Kočegarov, and E. Filatovas, "A novel greedy genetic algorithm-based personalized travel recommendation system," Expert Systems with Applications, vol. 230, p. 120580, 2023.
[20] R. Katarya and O. P. Verma, "A collaborative recommender system enhanced with particle swarm optimization technique," Multimedia Tools and Applications, vol. 75, no. 15, pp. 9225-9239, 2016.
[21] R. Katarya, "Movie recommender system with metaheuristic artificial bee," Neural Computing and Applications, vol. 30, no. 6, pp. 1983-1990, 2018.
[22] R. Katarya and O. P. Verma, "Recommender system with grey wolf optimizer and FCM," Neural Computing and Applications, vol. 30, no. 5, pp. 1679-1687, 2018.
[23] R. Katarya and O. P. Verma, "An effective collaborative movie recommender system with cuckoo search," Egyptian Informatics Journal, vol. 18, no. 2, pp. 105-112, 2017.
[24]S. Yadav and S. Nagpal, "An Improved Collaborative Filtering Based Recommender System using Bat Algorithm," Procedia computer science, vol. 132, pp. 1795-1803, 2018.
[25]V. Vellaichamy and V. Kalimuthu, "Hybrid Collaborative Movie Recommender System Using Clustering and Bat Optimization," International Journal of Intelligent Engineering and Systems, vol. 10, no. 5, pp. 38-47, 2017.
[26]J. M. Leski, "Fuzzy c-ordered-means clustering," Fuzzy Sets and Systems, vol. 286, pp. 114-133, 2016.
[27]J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proceedings of ICNN'95 - International Conference on Neural Networks, 27 Nov.-1 Dec. 1995 1995, vol. 4, pp. 1942-1948 vol.4, doi: 10.1109/ICNN.1995.488968.
[28] Z.-H. Zhan, J. Zhang, Y. Li, and H. S.-H. Chung, "Adaptive particle swarm optimization," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 6, pp. 1362-1381, 2009.
[29] Y. Shi and R. C. Eberhart, "Fuzzy adaptive particle swarm optimization," in Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546), 2001, vol. 1: IEEE, pp. 101-106.
[30] R. Sadeghi and J. Hamidzadeh, "Automatic support vector data description," Soft Computing, vol. 22, no. 1, pp. 147-158, 2018.
[31] J. Hamidzadeh, R. Sadeghi, and N. Namaei, "Weighted support vector data description based on chaotic bat algorithm," Applied Soft Computing, vol. 60, pp. 540-551, 2017.
[32] B. Alatas, E. Akin, and A. B. Ozer, "Chaos embedded particle swarm optimization algorithms," Chaos, Solitons & Fractals, vol. 40, no. 4, pp. 1715-1734, 2009.
[33] R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, X.-Z. Gao, and V. Indragandhi, "A hybrid quantum-induced swarm intelligence clustering for the urban trip recommendation in smart city," Future Generation Computer Systems, vol. 83, pp. 653-673, 2018.

جواد حمیدزاده در حال حاضر دانشیار دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه سجاد است. از علاقه‌مندی‌های ایشان می‌توان به یادگیری ماشین، رایانش نرم، بازشناسی الگو و شبکه‌های رایانه‌ای اشاره کرد.

نشانی رایانامه ایشان عبارت است از:

J_hamidzadeh@sadjad.ac.ir

منا مرادی دارای مـدرک دکتـرای کامپیوتر در گرایش هوش مصنوعی و رباتیک اسـت. از علاقه‌منـدی‌های ایشان می‌توان به یادگیری ماشـین، محاسبات نـرم و یـادگیری انتقـالی اشاره کرد.

نشانی رایانامه ایشان عبارت است از:

mmoradi@semnan.ac.ir