

فیلترکننده مشارکتی فازی ناهموار مبتنی بر کاربر در سیستم‌های پیشنهاددهنده

جواد حمیدزاده^۱، استادیار؛ مریم صادق‌زاده^۲، دانشجوی کارشناسی ارشد

۱- دانشکده مهندسی کامپیوتر و فناوری اطلاعات - دانشگاه صنعتی سجاد - مشهد - ایران - j_hamidzadeh@sadjad.ac.ir

۲- دانشکده مهندسی صنایع - دانشگاه صنعتی سجاد - مشهد - ایران - m.sadeghzadeh123@sadjad.ac.ir

چکیده: امروزه سیستم‌های پیشنهاددهنده، با توجه به کارایی خاص خود، در بسیاری از زمینه‌های علمی، کاربردهای فراوانی پیدا کرده است. نوعی از این سیستم‌ها، سیستم‌های پیشنهاددهنده فازی است، که با ترکیب نظریه فازی، انعطاف آن‌ها بیش‌تر شده است. در این مقاله، یک روش جدید با ترکیب فیلترکننده مشارکتی و مجموعه فازی ناهموار ارائه شده است که در آن، شباهت بر اساس فاصله اقلیدسی محاسبه می‌شود. در واقع با استفاده از مجموعه‌های فازی ناهموار، از اطلاعات کاربرانی که پیش‌بینی بهتری دارند، استفاده می‌شود. همچنین، از الگوریتم رقابت استعماری آشوبی، جهت تنظیم پارامترهای روش پیشنهادی استفاده شده است. این روش ترکیبی با استفاده از مجموعه‌های فازی ناهموار، صحت پیش‌بینی را افزایش می‌دهد. آزمایش‌های انجام‌شده روی مجموعه داده‌های محک نشان می‌دهد که روش پیشنهادی، میزان صحت را نسبت به روش‌های قبلی افزایش داده است.

واژه‌های کلیدی: سیستم‌های پیشنهاددهنده، مجموعه‌های فازی ناهموار، فاصله اقلیدسی، رقابت استعماری آشوبی.

A User Based Fuzzy Rough Collaborative Filtering in Recommender Systems

J. Hamidzadeh¹, Assistant Professor; M. Sadeghzadeh², MSc Student

1- Faculty of Computer Engineering and Information Technology, Sadjad University of Technology, Mashhad, Iran, Email: j_hamidzadeh@sadjad.ac.ir

2- Faculty of Industrial Engineering, Sadjad University of Technology, Mashhad, Iran, Email: m.sadeghzadeh123@sadjad.ac.ir

Abstract: Today, recommender systems, according to their special performance, are widely utilized in several scientific issues. Fuzzy recommender systems are the type of systems in which they are combined with fuzzy theory and give more flexibility to recommender systems. Nonetheless, in this article a new combination method from fuzzy rough set collaborative filtering is presented in which the distance is calculated based on Euclidean distance measure. In fact, by using fuzzy rough sets the applicants' data are used regarding to have better prediction. This combination method via fuzzy rough sets increases the accuracy of prediction. Since, implemented experiments on test data depict recommended method has better validation in comparison with other reputed methods.

Keywords: Recommender system, fuzzy rough set, Euclidean distance, chaos imperialist competitive.

تاریخ ارسال مقاله: ۱۳۹۴/۰۷/۱۲

تاریخ اصلاح مقاله: ۱۳۹۵/۰۲/۱۶ و ۱۳۹۵/۰۳/۲۲

تاریخ پذیرش مقاله: ۱۳۹۵/۰۶/۰۶

نام نویسنده مسئول: جواد حمیدزاده

نشانی نویسنده مسئول: ایران - مشهد - بلوار جلال آل احمد ۶۴ - دانشگاه صنعتی سجاد - دانشکده مهندسی کامپیوتر و فناوری اطلاعات.

۱- مقدمه

کاوش آن می‌تواند موجب شناسایی شباهت بین صفحات مختلف با یکدیگر شود. از این دانش می‌توان برای توصیه صفحات جدیدی که به وبسایت اضافه شده‌اند، استفاده کرد. در مقاله [۴]، یک سیستم توصیه‌گر ترکیبی با استفاده از دو روش پالایش گروهی و مبتنی بر محدودیت ارائه شده است که با توجه به استفاده از روش محدودیت محور، سعی در برطرف ساختن مشکل شروع سرد روش پالایش گروهی را داشته است.

در مقاله [۵] برای سیستم‌های پیشنهاددهنده، روشی بر اساس مجموعه فازی از جمله روش نمایش، معیارهای تشابه و روش‌های تجمعی ارائه شده است. در این روش، مسائل مبهم ناشی از فردیت، ابهام و عدم صحت در داده‌ها به خوبی بررسی می‌شود.

در مقاله‌های [۶، ۷] یک روش ترکیبی برای سیستم‌های پیشنهاددهنده، ارائه شده است. این روش، با ترکیب فیلترکننده رقابتی و الگوریتم‌های مبتنی بر محتوا، از مزایای هر دو روش برخوردار است. سیستم‌های پیشنهاددهنده فازی، برای مدل کردن عدم قطعیت موجود در اطلاعات شخصی کاربر، تنها از مجموعه فازی استفاده می‌کردند. در حالی که استفاده از قوانین ریاضی در کنار آن، می‌تواند روش سودمندی باشد. همچنین، در تحقیقات پیشین، شباهت بین کاربران تنها از روی اطلاعات شخصی آن‌ها محاسبه می‌شد، در حالی که می‌توان از سابقه امتیازدهی‌ها نیز جهت بیان شباهت بین دو کاربر، استفاده کرد.

[۸] یک تعریف ریاضی اصولی از سیستم‌های پیشنهاددهنده فازی به همراه عملیات جبری و ویژگی آن‌ها بیان کرده است. همچنین روشی به نام $HU-FCF$ ارائه کرده است که ایده اصلی آن ادغام کردن درجه‌های شباهت فازی بین کاربران است. این درجات از اطلاعات شخصی با درجه‌های قطعی مبتنی بر کاربر به دست آمده است. بنابراین اطلاعات داخلی (صفات کاربران)، اطلاعات خارجی (فعل و انفعالات بین کاربران) و درجه شباهت‌ها، ارتباط دقیق‌تر بین کاربران را نشان می‌دهد. بعد از محاسبه درجه‌های شباهت نهایی، رتبه‌بندی نهایی بر اساس رتبه‌بندی همسایه‌های کاربر مورد نظر به دست می‌آید. در واقع $HU-FCF$ صحت و انعطاف بهتری نسبت به روش‌های مشابه، با استفاده از درجه‌های شباهت ترکیبی دارد. یکی از معایب این روش زمان محاسبات طولانی است. مشکلات مربوط به تحقیقات سیستم‌های پیشنهاددهنده فازی نشان می‌دهد که سیستم پیشنهاددهنده ترکیبی، می‌تواند کاستی‌های روش‌های فیلترکننده را برطرف سازد.

در [۹] بیان می‌شود که سیستم‌های پیشنهاددهنده، نوع خاصی از سیستم‌های اطلاعاتی هستند که به انتخاب مناسب قلم‌داده‌ها، کمک می‌کنند. این سیستم‌ها در حوزه‌های مختلف، به‌منظور فردی‌سازی برنامه‌های کاربردی قلم‌داده‌های مختلفی پیشنهاد می‌دهند. یکی از مشکلات سیستم‌های پیشنهاددهنده، شروع سرد کاربران جدید است. این مشکل زمانی اتفاق می‌افتد که یک کاربر جدید وارد سیستم

در سال‌های اخیر، استفاده از وب از حالت آکادمیک به حالت تجاری تغییر یافته است. این امر باعث می‌شود که حجم عظیمی از اطلاعات، در دسترس کاربر نباشد و یا کاربران از وجود چنین اطلاعاتی مطلع نباشند. سیستم‌های پیشنهاددهنده، که زیرشاخه‌ای از سیستم‌های پشتیبان تصمیم هستند، می‌توانند در راستای یافتن اطلاعات مناسب، به کاربر کمک کنند. این سیستم‌ها، در رشته‌های مختلف، کاربردهای فراوانی دارند و پیشنهادهای خاصی را درباره موضوع مشخصی، که ممکن است برای کاربر جالب باشد، فراهم می‌کنند. سیستم‌های پیشنهاددهنده، کالاهای مورد علاقه کاربران را پیش‌بینی و به آن‌ها کمک می‌کنند تا از میان کالاهای متعدد، کالای مناسب را انتخاب کنند.

سیستم‌های پیشنهاددهنده به‌عنوان یک حوزه تحقیقاتی مستقل در اواسط دهه ۱۹۹۰ ظهور پیدا کردند، زمانی که محققین روی پیشنهادهایی که دقیقاً مبتنی بر ساختار امتیازدهی است، تمرکز داشتند. سیستم‌های پیشنهاددهنده سیستم‌هایی هستند که برای کاربران لیست مرتبی از قلم‌داده‌ها و اطلاعات را فراهم می‌کنند. این سیستم‌ها برای تصمیم‌گیری اینکه کدام قلم‌داده‌ها بر اساس اولویت کاربر در نظر گرفته شوند به کاربران کمک می‌کنند. طبیعی است که این سیستم‌ها بدون در اختیار داشتن اطلاعات کافی و صحیح در مورد کاربران و قلم‌داده‌های مورد نظر آن‌ها قادر به پیشنهاد دادن نمی‌باشند. بنابراین یکی از اساسی‌ترین اهداف آن‌ها، جمع‌آوری اطلاعات گوناگون در رابطه با علاقه کاربران و قلم‌داده‌های موجود در سیستم است. امروزه نیز استفاده از روش‌های شناسایی الگو در این نوع سیستم‌ها مطرح شده است. استفاده از روش‌های خوشه‌بندی داده‌ها و طبقه‌بندی آن‌ها از اهمیت خاصی برخوردار شده است [۱].

اهمیت سیستم‌های پیشنهاددهنده، هنگامی بیش‌تر مشخص می‌شود، که کاربران با انبوهی از اطلاعات روبه‌رو باشند. در چنین شرایطی، سیستم باید پیشنهادهای خود را با در نظر گرفتن چهار مورد به کاربر ارائه دهد که هرکدام از این موارد، زمینه ظهور گروهی از سیستم‌های پیشنهاددهنده را فراهم می‌کند. این موارد شامل شرایط و محیطی که کاربر در آن قرار دارد، نیاز کاربر، دانش سیستم در مورد کاربر و سابقه فعالیت‌های کاربر، است. امروزه این سیستم‌ها، کاربردهای متعددی در پیشنهاد کتاب، سی‌دی، هتل، فیلم، موسیقی، خرید اینترنتی، برنامه‌های تلویزیونی و غیره به کاربران دارند و در دهه اخیر، مقاله‌های متعددی در این راستا منتشر شده است [۲].

در مقاله [۳]، یک روش پیش‌بینی صفحات وب با استفاده از ترکیب خوشه‌بندی و مدل مارکوف ارائه شده است. در این روش، ابتدا نشست‌های کاربران خوشه‌بندی می‌شوند. سپس یک مدل مارکوف برای هر خوشه ایجاد شده و از آن برای پیش‌بینی درخواست‌های آتی کاربران استفاده می‌گردد. در این روش از محتوای صفحات برای خوشه‌بندی استفاده نشده است چرا که استفاده از محتوای صفحات و

در [۱۴]، روش مجموعه ناهموار احتمالی عمومی و خاص به منظور معنی‌شناسی حد آستانه و ویژگی‌های تصمیم، بیان شد. در این روش برخلاف مجموعه ناهموار پاولاک^۴، نواحی مثبت، نواحی مرزی و نواحی منفی، یکنواخت نیستند. این مسئله منجر به کاهش ویژگی‌های تداخلی و ناکارآمد می‌شود. به همین دلیل، معرفی یک حد آستانه احتمالی نیاز است. روش بیان‌شده نه‌تنها یکنواختی نواحی مثبت محتوای تصمیم را تضمین می‌کند، بلکه خطای محلی هر قسمت را نیز کمینه می‌نماید.

در [۱۵]، برای بیان کیفیت نمونه‌ها، مدل انتخاب نمونه‌های اولیه با استفاده از روش مجموعه‌های فازی ناهموار مطرح شده است. این مدل پیش‌پردازش، قصد دارد تا با حذف نمونه‌هایی که باعث پیش‌بینی بد می‌شوند و حفظ نمونه‌هایی با قابلیت پیش‌بینی خوب، منجر به بهبود دسته‌بندی KNN^۵ شود. همچنین در این مقاله، نشان داده شده نمونه‌هایی که منجر به پیش‌بینی بدتر می‌شوند، نمونه‌هایی هستند که روی مرزها، یا در قسمت‌هایی که کلاس‌های مختلف با یکدیگر همپوشانی دارند، واقع شده‌اند و یا نمونه‌های هستند که برچسب‌گذاری نشده‌اند.

مجموعه‌های فازی، پایه مناسبی برای سیستم‌های پیشنهاددهنده هستند. از جمله کاربردهای فازی، می‌توان اشاره کرد به: استفاده از روش‌های مجموعه فازی، برای نمایش محتوای دانش در یک سیستم پیشنهاددهنده، نمایش اولویت‌های کاربر به روش قابل انعطاف برای فرآیندهای تصمیم‌گیری موردی و استفاده از مقادیر فازی برای نمایش ویژگی‌های یک قلم‌داده. منطق فازی، می‌تواند برای خوشه‌بندی قلم‌داده‌ها یا گروه‌بندی کاربران با سلاقی مشابه استفاده شود.

نظریه مجموعه‌های فازی ناهموار در اوایل سال ۱۹۸۰ توسط پاولاک پایه‌گذاری شد. این نظریه، در مواردی که ابهام وجود دارد روش‌هایی را با استفاده از ابزار ریاضی ارائه می‌دهد که نیاز به اطلاعات اضافی در داده‌ها نیست. این روش‌ها، پس از تحلیل اطلاعات پنهان در داده‌ها، اطلاعات غیرمرتبط از مجموعه دادگان را حذف می‌نماید.

در سال‌های اخیر، مجموعه‌های فازی ناهموار موضوع قابل توجهی از دید محققان بوده‌اند و در بسیاری از زمینه‌ها کاربرد دارند. در سال ۱۹۸۱، پاولاک در [۱۶] عملیات تقریب زدن روی مجموعه‌ها، شباهت مجموعه‌ها و همچنین به دست آوردن گنجایش مجموعه‌ها را بررسی نمود. در واقع، هدف وی، توصیف ویژگی مجموعه‌های فازی ناهموار بود. مفهوم مجموعه‌های فازی ناهموار، در درجه اول در شاخه هوش مصنوعی معنا می‌یابد. مفاهیمی از جمله: استدلال استنتاجی، دسته‌بندی خودکار، شناسایی الگو، الگوریتم‌های یادگیری.

هدف ما در این مقاله، ارائه روشی هوشمند برای افزایش صحت در سیستم‌های پیشنهاددهنده است. در این روش، پیش‌بینی امتیازها بر اساس شباهت بین کاربران است که این شباهت، با فاصله اقلیدسی به دست می‌آید. کاربران مشابه، امتیازات نزدیک به هم کسب خواهند کرد. با این وجود، مشکل اصلی، مطلع نبودن از میزان شباهت کاربران

می‌شود. محققان در این حالت با محدودیت مجموعه دادگان، تخمین میزان بهینگی خوشه‌ها، معیار شباهت، کاربران بی‌ربط و انتخاب درجه تعلق، روبرو هستند. روش ++FCF-HU روشی است که با مشکلات یکپارچه‌سازی چندین گروه، مقابله نموده است و با استفاده از پردازش‌های خاص، باعث ترکیب مزیت‌های مختلف هر گروه و از بین بردن معایب آن‌ها، می‌شود.

در مقاله [۱۰]، از منطق فازی توسعه‌یافته، برای شناسایی در سیستم‌های غیرخطی استفاده شده است. [۱۱] نیز از مجموعه‌های فازی ناهموار به‌عنوان ابزاری سودمند، جهت پیش‌بینی در سیستم‌های پیشنهاددهنده، نام برده است. این مقاله به بررسی دو ویژگی پیشنهادها با استفاده از مجموعه‌های فازی ناهموار پرداخته است. اولی صحت پیشنهادها برای اکثر کاربران و دومی برجسته‌سازی پیشنهادهای عمومی است. افزایش صحت پیشنهادها، کیفیت سیستم‌های پیشنهاددهنده را برای بیش‌تر کاربران بالا می‌برد. در مجموعه‌های فازی ناهموار، این دو ویژگی توسط دو حد آستانه α و β کنترل می‌شود. یکی از مسائل مهم، تخمین مقادیر مؤثر دو حد آستانه مزبور بر اساس دو ویژگی بیان شده است. روش GTPS^۷، برای تخمین راه حل متعادل بین صحت و عمومیت، با استفاده از پیاده‌سازی یک بازی، این دو حد آستانه را به دست می‌آورد.

تشخیص دارویی، به‌عنوان یک پردازش مهم در داروهای بالینی بیان می‌شود که با استفاده از آن می‌توان بیماری‌ها را بر اساس علائم آن‌ها تخمین زد. افزایش صحت در تشخیص بیماری امری بسیار مهم است. محققان در این زمینه از روش‌های کامپیوتری از قبیل مجموعه‌های فازی شهودی و سیستم‌های پیشنهاددهنده استفاده کرده‌اند. از آنجا که داده‌های دارویی اغلب مبهم بوده و کامل نیستند، تنها استفاده از دو روش بیان‌شده، پیشرفتی در صحت تشخیص ندارد. بدین‌منظور این دو روش با هم ترکیب شده و روشی دیگر به نام IFRS^۲ به وجود آمده است [۱۲]. IFRS شامل تعاریفی از IFRS ترکیبی فازی شهودی، پیشنهاد ماتریس شباهت فازی شهودی، درجه شباهت فازی شهودی و به دست آوردن فرمول جهت پیش‌بینی مقادیر بر مبنای درجه شباهت فازی شهودی و بیان یک روش فیلترکننده مشارکتی فازی شهودی برای پیش‌بینی بیماری‌های ممکن، است.

در بیش‌تر سیستم‌های موجود، به‌طور ضمنی پایه دسته‌بندی‌ها را، دودویی در نظر می‌گیرند. قالب دیگری در [۱۳] برای ایجاد سیستم‌های پیشنهاددهنده بیان شده است که تصمیم سه راه را، با مجموعه درخت‌های تصادفی ادغام می‌نماید. در این روش تنظیماتی منطقی برای موقعیت و مقدار منفی حد آستانه به نام α^* و β^* انجام می‌شود. سپس یک مجموعه درخت‌های تصادفی برای محاسبه احتمال P ایجاد می‌گردد و در نهایت با استفاده از α^* ، β^* و P می‌توان رفتار پیشنهاددهنده را تخمین زد.

تعریف ۱: سیستم‌های پیشنهاددهنده (RS^{δ}) [۸]

اگر U را مجموعه تمام کاربران و I را مجموعه‌ای از کالاهای سیستم در نظر بگیریم، تابع سودمندی R روی $U_1 \subset U$ و $I_1 \subset I$ به صورت زیر نگاشت شده است:

$$R: U_1 \times I_1 \rightarrow p \quad (1)$$

$$(u_1, i_1) \mapsto R(u_1, i_1)$$

که $R(u_1, i_1)$ یک عدد صحیح غیرمنفی یا یک عدد حقیقی داخل محدوده‌ای خاص است. همچنین P مجموعه رتبه‌بندی‌های موجود در سیستم است. بنابراین RS سیستمی است که دو تابع زیر را آماده می‌کند:

(الف) پیش‌بینی: قابلیت تخمین $R(u^*, i^*)$ برای هر $(u^*, i^*) \in (U, I) \setminus (U_1, I_1)$

(ب) پیشنهاد: قابلیت انتخاب $i^* \in I$ ، $i^* = \arg \max_{i \in I} R(u, i)$ برای هر $u \in U$.

هدف ما در این مقاله، معرفی روشی هوشمند جهت افزایش صحت سیستم‌های پیشنهاددهنده است.

تعریف ۲: سیستم‌های پیشنهاددهنده چندمعیاره $(MCRS^{\gamma})$ [۸]

سیستم‌هایی که توابع پایه شبیه به RS فراهم می‌کنند؛ اما از ضوابط چندگانه پیروی می‌کنند. به عبارت دیگر تابع سودمندی، به صورت زیر تعریف می‌شود:

$$R: U_1 \times I_1 \rightarrow P_1 \times P_2 \times \dots \times P_k \quad (2)$$

$$(u_1, i_1) \mapsto (R_1, R_2, \dots, R_k)$$

که R_i ($i = 1, \dots, k$) رتبه‌بندی کاربران $u_1 \in U_1$ و $i_1 \in I_1$ است که از ضوابط چندگانه پیروی می‌کند. در این مورد، پیشنهاد مطابق با ضوابط داده شده انجام می‌شود.

سیستم‌های پیشنهاددهنده چندمعیاره می‌توانند برای پیش‌بینی رتبه‌بندی یک کاربر به فیلمی که از قبل رأی داده نشده، کمک کنند. همچنین این نوع سیستم‌ها، فیلم مورد علاقه کاربر را از طریق رتبه‌بندی موجود پیشنهاد می‌دهند.

تعریف ۳: سیستم‌های پیشنهاددهنده فازی (FRS^{λ}) [۸]

مجموعه‌های فازی، یک پایه نظری مناسب برای بسیاری از سیستم‌های پیشنهاددهنده ایجاد کرده است. روش‌های خوشه‌بندی فازی، که قلم‌داده‌ها را به بیش از یک خوشه دسته‌بندی می‌کند، به منظور ارائه واقع‌بینانه‌تر محتوای پروفایل‌ها استفاده می‌شوند. در سیستم پیشنهاددهنده مشارکتی، خوشه‌بندی فازی نیز، برای گروه‌بندی کاربران با علاقه‌های مشابه استفاده می‌شود.

اگر U را مجموعه تمام کاربران، I را مجموعه‌ای از کالاها و C را مجموعه‌ای از تعاریف فازی در نظر بگیریم، تابع سودمندی R ، روی $U_1 \subset U$ و $I_1 \subset I$ و $C_1 \subset C$ به صورت زیر نگاشت می‌شود:

$$R: U_1 \times I_1 \times C_1 \rightarrow P_1 \times P_2 \times \dots \times P_k \quad (3)$$

$$(u_1, i_1, \{c_1^*, \mu_c^*\}) \in (U, I, C) \setminus (U_1, I_1, C_1)$$

به یکدیگر دارند. زیرا در بسیاری از کاربردها، این مقدار را به طور دقیق نمی‌توان محاسبه کرد و باید به جای آن‌ها، اطلاعات دیگری مانند اطلاعات شخصی کاربران، استفاده شود. همچنین، بعضی ویژگی کاربران، مقدار پیوسته‌ای ندارند. بنابراین، برای محاسبه شباهت بین کاربران، لازم است منطق فازی را با سیستم پیشنهاددهنده ادغام کنیم. از سویی دیگر، انتخاب‌های بعضی از کاربران برای پیش‌بینی کالاها در آینده، درست نیست و منجر به پیش‌بینی‌های اشتباه می‌شوند. در چنین شرایطی، مجموعه‌های فازی ناهموار ابزاری مؤثر در حذف نظر چنین کاربرانی است تا پیش‌بینی‌های بهتری را ارائه دهد.

برای تنظیم بهینه پارامترهای روش پیشنهادی در این مقاله، از الگوریتم رقابت استعماری آشوبی استفاده شده است. الگوریتم رقابت استعماری، مانند سایر الگوریتم‌های تکاملی، یک جمعیت اولیه دارد. جمعیت اولیه کشور نامیده می‌شود که بر دو نوع است: مستعمره و استعمارگر. این دو نوع جمعیت با هم تعدادی امپراتوری تشکیل می‌دهند. اساس این الگوریتم، رقابت استعمارگراییه میان این امپراتوری‌ها است. در طول این رقابت، امپراتوری‌های ضعیف از هم فرو می‌پاشند و قوی‌ترها مستعمرات بقیه را صاحب می‌شوند.

رقابت استعمارگراییه، به طور مطلوب به ایالتی همگرا می‌شود که در آن تنها یک امپراتوری وجود دارد، مستعمرات آن در یک موقعیت قرار دارند و هزینه یکسان به عنوان استعمارگر دارد. مانند همه الگوریتم‌های قرار گرفته در این دسته، الگوریتم رقابت استعماری نیز، مجموعه اولیه‌ای از جواب‌های احتمالی را دارا است. این جواب‌های اولیه در الگوریتم ژنتیک به نام «کروموزوم»، در الگوریتم ازدحام ذرات به نام «ذره» و در الگوریتم رقابت استعماری با عنوان «کشور» شناخته می‌شوند. الگوریتم رقابت استعماری، با روند خاصی، این جواب‌های اولیه (کشورها) را به تدریج بهبود داده و در نهایت جواب مناسب را برای مسئله بهینه‌سازی (کشور مطلوب) در اختیار می‌گذارد [۱۷].

نظریه آشوب [۱۸]، یک زمینه مطالعاتی در ریاضیات است، که در چندین رشته از جمله هواشناسی، جامعه‌شناسی، فیزیک، مهندسی، اقتصاد، زیست‌شناسی و فلسفه کاربرد دارد. آشوب، نمونه‌ای از رفتار پویای سیستم‌های غیرخطی است که در زمینه‌های مختلف علوم، مانند کنترل آشوب، همگام‌سازی، شناسایی الگو، نظریه بهینه‌سازی نیز استفاده شده است [۱۹].

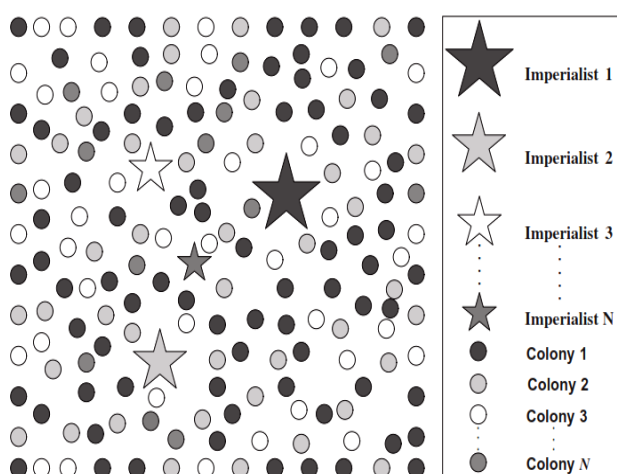
ساختار ادامه مقاله به صورت زیر است:

در بخش دوم، تعاریف و مسئله بیان شده است. مروری بر الگوریتم رقابت استعماری آشوبی در بخش سوم ارائه شده است. در بخش چهارم، جزئیات روش پیشنهادی نشان شده است. نتایج آزمایش‌ها، ارزیابی و مقایسه آن‌ها در بخش پنجم نشان داده شده است. نتیجه‌گیری و بیان کارهای آینده در بخش آخر ذکر شده است.

۲- تعاریف و بیان مسئله

در این بخش، تعاریف لازم برای بیان مسئله ارائه شده است.

که باید بهینه شوند. هرکدام از این متغیرها یک «کشور» نامیده می‌شوند. می‌خواهیم این کشورها را به دو دسته استعمارگر و مستعمره تقسیم کنیم. تعدادی از بهترین عناصر جمعیت به‌عنوان امپریالیست انتخاب می‌شوند، باقی‌مانده جمعیت نیز به‌عنوان مستعمره، در نظر گرفته می‌شوند. از آن پس امپراتوری‌ها به‌منظور افزایش قدرت خود، به جذب مستعمرات سایر امپراتوری‌ها می‌پردازند و با یکدیگر رقابت می‌کنند. هر امپراتوری که قوی‌تر باشد قدرت بیشتری برای جذب کشورهای مستعمره دارد، در نتیجه امپراتوری‌های قوی‌تر همان‌طور که در شکل ۱ مشاهده می‌کنید تعداد مستعمرات بیشتری دارند [۱۷]. روند حذف امپراتوری‌های ضعیف و افزایش تعداد مستعمره‌های کشورهای قوی ادامه می‌یابد و حد نهایی رقابت زمانی است که تنها یک امپراتوری واحد وجود داشته باشد.



شکل ۱: امپراتوری اولیه [۱۷]

در ابتدای کار تعداد $N_{country}$ کشور اولیه به‌صورت تصادفی انتخاب می‌کنیم. برای هر کشور تابع هزینه را محاسبه کنیم و با استفاده از رابطه (۷) هزینه نرمالیزه آن‌ها را به دست می‌آوریم. در حل این مسئله بهینه‌سازی برای یافتن بهترین کشورها باید به دنبال کشورهایی با کم‌ترین هزینه باشیم [۱۷].

$$C_n = \max_i \{c_i\} - c_n \quad (7)$$

که در آن c_n ، هزینه امپریالیست $\max_i \{c_i\}$ بیش‌ترین هزینه میان امپریالیست‌ها و c_n هزینه نرمالیزه‌شده این امپریالیست می‌باشد. طبق این رابطه هرچه هزینه یک امپریالیست بیش‌تر باشد، هزینه نرمال شده آن کم‌تر است. با داشتن هزینه نرمالیزه، قدرت نسبی نرمالیزه هر امپریالیست، به کمک رابطه (۸) محاسبه شده و بر مبنای آن، کشورهای مستعمره بین امپریالیست‌ها تقسیم می‌شوند، به این صورت که N_{imp} تا را از بهترین (قوی‌ترین) اعضای این جمعیت را به‌عنوان امپریالیست و باقی‌مانده N_{col} تا از کشورها، مستعمراتی را تشکیل می‌دهند که هرکدام به یک امپراتوری تعلق دارند.

که $\mu_c \in [0, 1]$ مقدار عضویت تعریف c_i است. $R_i (i = 1, \dots, k)$ امتیاز کاربر u_i به کالای i_1 در تعریف c_i با ضوابط P_i با درجه تعلق فازی $\mu_i \in [0, 1]$ است. FRS سیستمی است که دو تابع اساسی زیر را در اختیار می‌گذارد:

- (الف) پیش‌بینی: قابلیت تخمین زدن $R(u^*, i^*, \{c_i^*, \mu_c^*\})$ برای هر $(u^*, i^*, \{c_i^*, \mu_c^*\}) \in (U, I, C) \setminus (U_1, I_1, C_1)$
 (ب) پیشنهاد: قابلیت انتخاب $i^* \in I$ $i^* = \arg \max_{i \in I} R(u, i, \{c, \mu_c\})$ برای هر $u \in U$ و $\{c, \mu_c\} \in C$ و یک میزان معین.

تعریف ۴: فضای تقریب [۱۶]

زوج (U, R) را که در آن U مجموعه معلوم به نام جهان و R یک رابطه غیر قابل تشخیص است در نظر بگیرید. در واقع یک رابطه هم‌ارزی روی U است. $A = (U, R)$ را فضای تقریب می‌نامند. اگر $(x, y) \in R$ ، $x, y \in U$ ، در این صورت می‌گوییم x, y در A غیر قابل تشخیص است.

تعریف ۵: مجموعه‌های ناهموار [۱۶]

اگر $X, X \subseteq U$ می‌تواند با استفاده از اطلاعات موجود در مجموعه خصوصیت P تخمین زده شود و حد پایین P و حد بالای P را به دست آورد:

$$\underline{P}X = \{x \mid [x]_p \subseteq X\} \quad (4)$$

$$\overline{P}X = \{x \mid [x]_p \cap X \neq \emptyset\}$$

دوتایی $(\overline{P}X, \underline{P}X)$ به‌عنوان مجموعه ناهموار شناخته می‌شود.

تعریف ۶: مجموعه‌های فازی ناهموار [۲۰]

با ترکیب مجموعه فازی و مجموعه ناهموار، می‌توان روی فازی‌سازی تقریب‌های بالا و پایین تمرکز کرد.

تقریب یکسانی بین اشیا با مدل کردن رابطه فازی R در X ، درجه شباهت هر زوج اشیا را مشخص می‌کند. اگر R را رابطه فازی و A را مجموعه فازی در نظر بگیریم، تعریف کلی زیر را برای تقریب‌های پایین و بالای A توسط R می‌توان بیان کرد:

$$(R \downarrow A)(x) = \inf_{y \in X} t(R(x, y), A(y)) \quad (5)$$

$$(R \uparrow A)(x) = \sup_{y \in X} \tau(R(x, y), A(y)) \quad (6)$$

که t ، یک عملگر مشمول فازی^۹ و τ ، یک عملگر اشتراک فازی^{۱۰} است. در این مقاله، با استفاده از مجموعه‌های فازی ناهموار، از اطلاعات کاربرانی که پیش‌بینی بهتری دارند، استفاده می‌شود.

در این مقاله، مقدار حد آستانه θ با استفاده از روش رقابت استعماری آشوبی به دست آمده است.

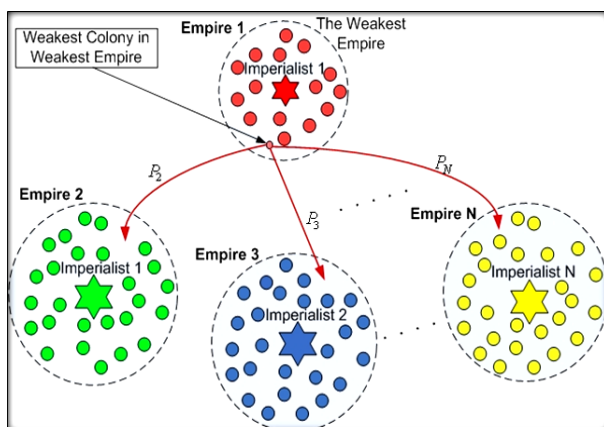
۳- الگوریتم رقابت استعماری آشوبی

الگوریتم رقابت استعماری (ICA) در سال ۲۰۰۷ توسط آقای اسماعیل آتش‌پز و مرحوم پرفسور کارولوکس ارائه شده است. این الگوریتم در ابتدا تعدادی جمعیت اولیه تصادفی به‌صورت آرایه‌ای از متغیرها دارد

در حین حرکت مستعمرات به سمت کشور استعمارگر، ممکن است بعضی از این مستعمرات به مقادیری در تابع هزینه برسند که هزینه کمتری را نسبت به مقدار تابع هزینه در موقعیت امپریالیست، تولید می‌کنند. در این حالت، کشور استعمارگر و کشور مستعمره، جای خود را با همدیگر عوض کرده و الگوریتم با کشور استعمارگر در موقعیت جدید ادامه یافته و این بار این کشور امپریالیست جدید است که شروع به اعمال سیاست همگون‌سازی بر مستعمرات خود می‌کند. قدرت یک امپراتوری برابر است با قدرت کشور استعمارگر، به‌اضافه درصدی از قدرت کل مستعمرات آن که مقدار آن را به کمک رابطه (۱۰) محاسبه می‌کنیم.

$$T.C_n = \text{Cost}(\text{imperialist}_n) + \xi \text{mean}\{\text{Cost}(\text{colonies of empire}_n)\} \quad (10)$$

در این رابطه $T.C_n$ هزینه کل امپراتوری n ام و ξ عددی مثبتی است که معمولاً بین صفر و یک و نزدیک به صفر در نظر گرفته می‌شود. در جریان رقابت‌های امپریالیستی، به مرور زمان امپراتوری‌های ضعیف مستعمرات خود را از دست داده و امپراتوری‌های قوی‌تر، این مستعمرات را تصاحب کرده و بر قدرت خویش می‌افزایند. در نهایت هر امپراتوری که نتواند بر قدرت خود بیفزاید و قدرت رقابت خود را از دست بدهد، در جریان رقابت‌های امپریالیستی، حذف خواهد شد. برای مدل کردن این واقعیت، فرض می‌کنیم که امپراتوری در حال حذف، ضعیف‌ترین امپراتوری موجود است. بدین ترتیب، در تکرار الگوریتم، یکی یا چند تا از ضعیف‌ترین مستعمرات ضعیف‌ترین امپراتوری را برداشته و برای تصاحب این مستعمرات، رقابتی را میان کلیه امپراتوری‌ها ایجاد می‌کنیم. مستعمرات مذکور، لزوماً توسط قوی‌ترین امپراتوری، تصاحب نخواهند شد، بلکه امپراتوری‌های قوی‌تر، احتمال تصاحب بیشتری دارند. شکل ۳ رقابت امپراتوری‌ها برای جذب کشور مستعمره ضعیف را نشان می‌دهد.



شکل ۳: شمای کلی رقابت استعماری [۱۷]

امپراتوری‌های قرار گرفته در قسمت پایین شکل برای تصاحب یکی از مستعمرات ضعیف امپراتوری یک که ضعیف‌ترین امپراتوری

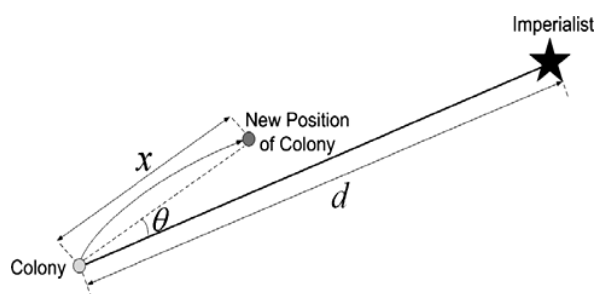
$$P_n = \left| \frac{C_n}{\sum_{i=1}^{N_{col}} C_i} \right| \quad (8)$$

از دید دیگر، تعداد مستعمرات برای هر امپراتوری نشان‌دهنده قدرت آن می‌باشد. از این جهت قدرت نرمالیزه شده یک امپریالیست، نسبت مستعمراتی است که توسط آن امپریالیست اداره می‌شود. بنابراین تعداد اولیه مستعمرات یک امپریالیست را به کمک رابطه (۹) به دست می‌آوریم.

$$N.C_n = \text{round}\{p_n \cdot (N_{col})\} \quad (9)$$

در این رابطه $N.C_n$ تعداد اولیه مستعمرات یک امپراتوری و C_n تعداد کل کشورهای مستعمره موجود در جمعیت کشورهای اولیه است. از این جهت که $N.C_n$ بیانگر تعداد کشورهاست که باید مقدار آن عددی صحیح باشد، لذا از تابع round استفاده می‌شود. این تابع نزدیک‌ترین عدد صحیح به یک عدد اعشاری را می‌دهد. تعداد $N.C$ مستعمره اولیه را به صورت تصادفی انتخاب کرده و به امپریالیست n ام می‌دهیم. با داشتن حالت اولیه تمام امپراتوری‌ها، الگوریتم رقابت استعماری شروع می‌شود. هر امپراتوری تلاش می‌کند کشورهای مستعمره بیشتری را به سمت خود جذب کند؛ زیرا بقای یک امپراتوری، وابسته به قدرت آن در جذب مستعمرات امپراتوری‌های رقیب خواهد بود.

شکل ۲ شمای کلی حرکت مستعمرات به سمت کشور امپریالیست را نشان می‌دهد. در این شکل، همان‌طور که مشاهده می‌کنید فاصله میان استعمارگر و مستعمره با d نشان داده شده است. کشور مستعمره^۱، به اندازه x واحد در جهت خط واصل مستعمره به استعمارگر^۲، حرکت می‌کند، x می‌تواند عددی تصادفی با توزیع یکنواخت (و یا هر توزیع مناسب دیگر) باشد. علاوه بر این می‌توانیم زاویه تصادفی θ را بر راستای حرکت کشور مستعمره به سمت کشور استعمارگر متأثر کنیم، به این صورت که کشور مستعمره با انحراف θ در مسیر، به حرکت خود ادامه می‌دهد. مقدار θ معمولاً به صورت تصادفی ارائه می‌شود. این شکل چگونگی اعمال زاویه θ را بر راستای حرکت کشور مستعمره نشان می‌دهد.



شکل ۲: شمای کلی حرکت مستعمرات به سمت امپریالیست با اعمال زاویه θ [۱۷]

۴- روش پیشنهادی

در این بخش، الگوریتم FRCF (شکل ۵) را ارائه می‌دهیم. هدف این روش، افزایش نرخ صحت پیش‌بینی نسبت به روش‌های قبلی، از جمله روش HU-FCF است. ایده اصلی روش FRCF، یافتن داده‌هایی است که بر اساس آن‌ها می‌توان پیش‌بینی‌های بهتری را انجام داد. این کار با استفاده از مجموعه‌های فازی ناهموار و به دست آوردن مقدار α که در [۱۴] آمده است انجام می‌شود. پس از یافتن و جدا کردن این داده‌ها در مجموعه‌ای مجزا (s)، درجه شباهت بین کاربران مجموعه آموزش و مجموعه آزمون، توسط معکوس معیار فاصله اقلیدسی محاسبه می‌شود. در کاربرد سیستم‌های پیشنهاددهنده، می‌توان نمودار p را ویژگی یک فرد موجود در مجموعه آموزش و q را ویژگی یک فرد در مجموعه آزمون در نظر گرفت. بسته به دامنه مسئله خاص، رتبه‌بندی نهایی به نزدیک‌ترین مقدار آن در دامنه، همراه خطای حد آستانه تخمین زده خواهد شد.

ورودی:

- اطلاعات شخصی کاربر: $U = \{U_1, \dots, U_N\}$
- که هر کاربر به صورت $U_i = \{U_i^1, \dots, U_i^I\} (i = 1, \dots, N)$ است، N تعداد کاربران، I تعداد ویژگی‌های شخصی
- مجموعه اقلام: $I = \{I_1, \dots, I_M\}$ که M تعداد اقلام را نشان می‌دهد.
- تاریخچه رتبه‌بندی: $R = \{R(U_i, I_j) | U_i \in U; I_j \in I\}$

در الگوریتم FRCF، ابتدا با فراخوانی تابع رقابت استعماری آشوبی مقدار مناسب پارامتر حد آستانه مشخص می‌شود. سپس انتخاب داده‌هایی که به وسیله آن‌ها بهتر می‌توان پیش‌بینی کرد (مجموعه s). محاسبه معیار α برای تک تک داده‌ها با استفاده از رابطه (۱۳):

$$\alpha_k = \max_{i=1}^n \frac{1}{\max_{j=1}^m \delta_{a_j}(x, i)}, \text{ where } \delta_{a_j}(x, y) = (a_j(x) - a_j(y))^2 \quad (13)$$

سپس حفظ داده‌هایی که مقدار α آن‌ها بیش‌تر از حد آستانه θ است (تشکیل مجموعه s). تخمین زدن درجه‌های شباهت بین کاربران مجموعه آموزش و کاربران مجموعه آزمایش با استفاده از معیار فاصله اقلیدسی:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (14)$$

درجه شباهت را معکوس فاصله اقلیدسی در نظر می‌گیریم:

$$SIM(p, q) = \frac{1}{d(p, q)} \quad (15)$$

محاسبه امتیاز نهایی توسط رابطه (۱۶):

$$R(a, i^*) = \bar{r}_a + \frac{\sum_{b \in U_{[a]}} SIM(a, b) * (r_{b, i^*} - \bar{r}_b)}{\sum_{b \in U_{[a]}} SIM(a, b)} \quad (16)$$

می‌باشد؛ با هم به رقابت می‌پردازند. برای مدل‌سازی رقابت میان امپراتوری‌ها برای تصاحب این مستعمرات، باید احتمال تصاحب هر امپراتوری را با در نظر گرفتن هزینه کل امپراتوری محاسبه می‌کنیم. ابتدا از روی هزینه کل امپراتوری، هزینه کل نرمالیزه شده آن را به کمک رابطه (۱۱) تعیین می‌کنیم.

$$NTC_n = \max_i \{T.C_i\} - T.C_n \quad (11)$$

در این رابطه $T.C_n$ ، هزینه کل امپراتوری n ام و NTC_n نیز، هزینه کل نرمالیزه شده آن امپراتوری می‌باشد. هر امپراتوری که $T.C_n$ کم‌تری داشته باشد NTC_n بیش‌تری خواهد داشت. در حقیقت $T.C_n$ معادل هزینه کل یک امپراتوری و NTC_n معادل قدرت کل آن می‌باشد. امپراتوری با کم‌ترین هزینه، دارای بیش‌ترین قدرت است. با داشتن هزینه کل نرمالیزه شده، احتمال (قدرت) تصاحب مستعمره رقابت، توسط هر امپراتوری، با استفاده از رابطه (۱۲) محاسبه می‌شود.

$$p_n = \frac{NTC_n}{\sum_{i=1}^n NTC_i} \quad (12)$$

پس از مدتی، همه امپراتوری‌ها، سقوط کرده و تنها یک امپراتوری خواهیم داشت و بقیه کشورها تحت کنترل این امپراتوری واحد، قرار می‌گیرند. در چنین موقعیتی رقابت امپریالیستی به پایان رسیده و به عنوان یکی از شروط توقف الگوریتم متوقف می‌شود. در شکل ۴، شبه‌کد الگوریتم رقابت استعماری آشوبی نشان داده شده است.

```

Function CICA (problem)
Input : Population_size, Problem_size, EP_size, beta, P_R
Output : S_best

Population ← phi;
EmpiresPopulation ← phi;

for i = 1 to population_size do
    C_iPosition ← RandomPosition (Problem_size);

    if i ≤ EP_size then
        EmpiresPopulation ← C_iPosition;
    else
        C_w ← GetWorstSolution (EmpiresPopulation)
        if Fitness (C_iPosition) > Fitness (C_wPosition) then
            ReplaceSolution (EmpiresPopulation, C_j, C_w)
        else
            C_iEmpire ← AssignAnEmpire (EmpiresPopulation);
        end
    end
end
Population ← C_j;
end
    
```

شکل ۴: الگوریتم رقابت استعماری

محاسبه خطای حد آستانه در ادامه:

$$\Delta = 100 \times \frac{|R(a, i^*) - d|}{\max\{R(a, i^*), d\}} \quad (17)$$

که $d \in D$ نزدیک‌ترین مقدار $R(a, i^*)$ است.

در این رابطه x_k ، k امین عدد آشوبی است که اگر مقدار اولیه x مخالف مقادیر 0 ، 0.25 ، 0.5 ، 0.75 و 1 باشد تولید می‌شود. همچنین مقدار a به صورت تجربی 4 در نظر گرفته می‌شود.

در الگوریتم رقابت استعماری آشوبی، از رابطه (۱۷)، جهت تولید دنباله آشوبی استفاده شده است.

$$\gamma_{n+1} = a\gamma_n(1 - \gamma_n) \quad (19)$$

که در آن a پارامتر کنترل و γ پارامتر آشوبی در تکرار n ام است. در نهایت با اضافه کردن متغیر آشوبی، موقعیت مستعمره به صورت زیر به دست می‌آید:

$$\{\theta\}_{new} = \{\theta\}_{old} + \beta \times d \times \{cm\} \times \{V\} \quad (20)$$

که در آن cm متغیر آشوبی است.

۵- آزمایش‌ها

برای انجام آزمایش‌ها و مقایسه روش پیشنهادی با سایر روش‌ها، از مجموعه داده‌های معرفی شده در جدول ۱ استفاده شده است. مجموعه داده MovieLens [۲۱] شامل دو مدل 100K و 1M است، که حاوی امتیازات کاربران به یک مجموعه از فیلم‌ها است. این امتیازات در پنج سطح مختلف از ۱ (بد) تا ۵ (عالی) است. در مجموعه داده MovieLens100k، 100,000 امتیاز از ۹۴۳ کاربر در مورد ۱۶۸۲ فیلم ثبت شده است. هر کاربری، حداقل به ۲۰ فیلم امتیاز داده است. در مجموعه داده MovieLensIM، 1,000,209 امتیاز از ۶۰۴۰ کاربر در مورد ۳۹۰۰ فیلم ثبت شده است. در دو مجموعه داده فوق، تعداد ویژگی‌ها سه بوده که مشتمل بر شناسه کاربر، شناسه فیلم و امتیاز است. مجموعه داده Book-Crossing [۲۲] نیز دربرگیرنده امتیازات کاربران در مورد یک مجموعه از کتاب‌ها است. این امتیازات در ده سطح مختلف از ۱ تا ۱۰ است. در این مجموعه داده نیز، تعداد ویژگی‌ها سه بوده که مشتمل بر شناسه کاربر، شناسه کتاب و امتیاز است. مجموعه داده Jester، در برگیرنده امتیاز کاربران به یک مجموعه صدتایی جُک است [۲۳]. مشخصات مجموعه داده‌های فوق در جدول ۱ نمایش داده شده است. پیاده‌سازی‌ها در محیط Matlab انجام شده و روش ارزیابی در همه آزمایش‌ها، مبتنی بر روش اعتبارسنجی متقاطع ۳تایی^{۱۳} بوده است. ارزیابی این روش، با محاسبه میانگین صحت انجام می‌شود. سطح آستانه مجموعه فازی ناهموار، با استفاده از روش رقابت استعماری آشوبی به دست آمده است.

جدول ۱: مشخصات مجموعه داده‌ها

نام مجموعه داده	تعداد کاربران	تعداد ویژگی‌ها	تعداد ارقام (کالاها)	تعداد امتیازها
MovieLens 100K [۲۱]	۹۴۳	۳	۱۶۸۲	۱۰۰۰۰۰
MovieLens IM [۲۱]	۶۰۴۰	۳	۳۹۰۰	۱۰۰۰۲۰۹
Book-Crossing [۲۲]	۲۷۸۸۵۸	۳	۲۷۱۳۷۹	۱۱۴۹۷۸۰
Jester [۲۳]	۲۴۹۳۸	۳	۱۰۰	۶۲۳۴۵۰

```

FRCF:
// Set theta parameter using CICA function
Theta= CICA( );
// selectdata( );
// calculate Alpha for each user k and compose AlphaList
Pos_A^alpha(y) = min_{x in X} t(R_A^alpha(x, y), R_d(x, y))
alpha(x) = sup{alpha in [0, +inf]} | POS_A^alpha(x) < 1 | =
max_{i=1}^n \frac{1}{\max_{j=1}^m \delta_{a_j}(x, i)}, where
\delta_{a_j}(x, y) = (a_j(x) - a_j(y))^2

//Keep data which has alpha more than //threshold (Theta)

// Calculate user similarity
d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_m - q_m)^2}
SIM(p, q) = \frac{1}{d(p, q)}
//Calculate user rating
R(a, i^*) = \bar{r}_a + \frac{\sum_{b \in U(a)} SIM(a, b) * (r_{b, i^*} - \bar{r}_b)}{\sum_{b \in U(a)} |SIM(a, b)|}

// Calculate error rate
// R is user rating estimated
// d is the real user rating
\Delta = 100 \times \frac{|R(a, i^*) - d|}{\max\{R(a, i^*), d\}}
    
```

شکل ۵: الگوریتم روش پیشنهادی FRCF

در حل این مسئله از الگوریتم رقابت استعماری آشوبی جهت تعیین مقدار بهینه حد آستانه استفاده شده است. به منظور افزایش قابلیت‌های الگوریتم رقابت استعماری، برای محاسبه جهت حرکت کشور مستعمره به سمت امپراتوری (θ)، روش آشوبی به کار برده شده است. در الگوریتم رقابت استعماری برای متغیر θ از اعداد تصادفی استفاده می‌شود. استفاده از نگاشت‌های آشوبی، به جای الگوریتم‌های تولیدکننده اعداد تصادفی برای محاسبه θ ، باعث همگرایی بهتر نتایج می‌شود. به کمک روش آشوبی می‌توان مسئله مصرف بالای حافظه را در روش‌های تولیدکننده اعداد تصادفی بهبود بخشید. نگاشت استفاده شده در روش آشوبی نگاشت لجستیک می‌باشد. رابطه این الگو به صورت یک دینامیک غیرخطی از جمعیت زیستی با رفتار آشوبی به دست می‌آید [۱۷]:

$$x_{k+1} = ax_k(1 - x_k) \quad (18)$$

در ادامه این تحقیق، می‌توان خوشه‌بندی کاربران و تأثیر آن بر روی افزایش نرخ صحت پیش‌بینی امتیازها را در کاربردهای مختلف بررسی نمود. با توجه به زمان محاسبه بالای روش پیشنهادی، در ادامه تحقیق فوق، می‌توان در جهت کاهش حجم محاسبات روش پیشنهادی گام برداشت.

مراجع

- [1] J. Hamidzadeh, R. Monsefi and H. Sadoghi Yazdi, "IRAHC: instance reduction algorithm using hyperrectangle clustering," *Pattern Recognition*, vol. 48, p. 1878-1889, 2015.
- [2] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," *Expert Systems with Applications*, vol. 39, pp. 10059-10072, 2012.
- [3] سیامک عبدالله‌زاده، محمدعلی بالافر و لیلی محمدخانی، «استفاده از خوشه‌بندی و مدل مارکوف جهت پیش‌بینی درخواست آتی کاربر در وب»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۵، شماره ۳، صفحات ۸۹-۹۶، ۱۳۹۴.
- [4] مصطفی رجب‌زاده و رضا رافع، «ارائه یک سیستم توصیه‌گر ترکیبی برای تجارت الکترونیک»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۵، شماره ۴، صفحات ۸۵-۹۱، ۱۳۹۴.
- [5] A. Zenebe and A. F. Norcio, "Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems," *Fuzzy Sets and Systems*, vol. 160, pp. 76-94, 2009.
- [6] J. P. Lucas, A. Laurent, M. N. Moreno, and M. Teisseire, "A fuzzy associative classification approach for recommender systems," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, pp. 579-617, 2012.
- [7] J. P. Lucas, A. Laurent, M. N. Moreno and M. Teisseire, "A fuzzy associative classification approach for recommender systems," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, pp. 579-617, 2012.
- [8] Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang and J. Lu, "A hybrid fuzzy-based personalized recommender system for telecom products/services," *Information Sciences*, vol. 235, pp. 117-129, 2013.
- [9] L. H. Son, "HU-FCF: A hybrid user-based fuzzy collaborative filtering method in Recommender Systems," *Expert Systems with Applications*, vol. 41, pp. 6861-6870, 2014.
- [10] L. H. Son, "HU-FCF++: A novel hybrid method for the new user cold-start problem in recommender systems," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 207-222, 2015.
- [11] فرناز صباحی و محمدرضا اکبرزاده توتونچی، «شناسایی سیستم‌های غیرخطی بر اساس منطق فازی توسعه‌یافته»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۴، شماره ۱، صفحات ۲۳-۳۲، ۱۳۹۳.
- [12] N. Azam and J. Yao, "Game-theoretic rough sets for recommender systems," *Knowledge-Based Systems*, vol. 72, pp. 96-107, 2014.

نتایج به دست آمده از اعمال الگوریتم ارائه‌شده بر روی مجموعه داده‌های جدول ۱، در جدول ۲ نمایش داده شده است.

جدول ۲: میانگین نرخ صحت

نام مجموعه داده	ZN [۵]	Lucas [۶]	HU-FCF [۸]	GTRS [۱۱]	FRCF
MovieLens 100K	٪۵۵/۰	٪۸۱/۴	٪۸۲/۱	٪۸۴/۱۱	٪۹۴/۵۸
MovieLens 1M	٪۶۲/۱۶	٪۷۸/۴۰	٪۷۸/۴۳	٪۸۵/۴۶	٪۹۵/۷۳
Book-Crossing	٪۸۳/۴۰	٪۷۹/۴۱	٪۷۷/۴۶	٪۷۳/۲۹	٪۸۱/۵۰
Jester	٪۵۱/۲۰	٪۶۶/۱۳	٪۵۴/۸۰	٪۶۵/۱۰	٪۶۵/۱۰

جدول ۳: زمان اجرای روش‌ها (ثانیه)

نام مجموعه داده	ZN [۵]	Lucas [۶]	HU-FCF [۸]	GTRS [۱۱]	FRCF
MovieLens 100K	۱۸/۳	۱۵/۳	۲۲/۷	۴۱	۳۰
MovieLens 1M	۹۳	۷۸	۱۱۳	۱۸۳	۲۹۵
Book-Crossing	۱۱۵	۹۶	۱۳۲	۲۷۴	۳۱۰
Jester	۵۰/۱	۸۷	۱۱۲	۲۵۴	۲۸۵

با توجه به اینکه روش‌های مطرح‌شده قبلی برای رسیدن به جواب، فاصله‌ها را با استفاده از روش‌های متفاوت به دست آورده‌اند؛ در جدول ۲، میانگین نتایج نشان داده شده است. در بعضی از الگوریتم‌ها مانند HU-FCF، در برخی حالت‌ها، نتایج به‌دست‌آمده از روش‌های مطرح‌شده قبلی بدتر بوده است، اما در الگوریتم FRCF با استفاده از مجموعه‌های فازی ناهموار و همچنین فاصله اقلیدسی، نرخ صحت افزایش یافته است. ضمناً، زمان محاسبه طولانی، از معایب روش ارائه‌شده می‌باشد که در جدول ۳ نشان داده شده است، البته در صورتی که پارامتر صحت، اهمیت بیش‌تری داشته باشد؛ روش ارائه شده، روش مناسب‌تری است.

۶- نتیجه‌گیری و کارهای آینده

در این مقاله، افزایش نرخ صحت پیش‌بینی روش فیلترکننده موجود در سیستم‌های پیشنهاددهنده، نشان داده شده است. با بررسی تحقیقات مشابه، مشاهده شد که از مجموعه‌های فازی ناهموار، به‌منظور حذف کاربرانی که پیش‌بینی‌های نادرست را به دنبال دارند، استفاده نکرده‌اند. روش ارائه‌شده در این مقاله، قابلیت اعمال به هر مسئله کاربردی را دارد. الگوریتم FRCF ترکیبی از الگوریتم HU-FCF و الگوریتم FRFS است که با استفاده از فاصله اقلیدسی به‌منظور تخمین شباهت کاربران، نتایج بهتری نسبت به الگوریتم‌های پیشین، را دارا است.

- [۱۹] حسین شایقی و علی قاسمی، «پیش‌بینی قیمت روزانه برق با شبکه عصبی بهبودیافته مبتنی بر تبدیل موجک و روش آشوبناک جستجوی گرانشی»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۵، شماره ۴، صفحات ۱۰۵-۱۱۵، ۱۳۹۴.
- [20] E. Ott, *Chaos in Dynamic Systems*, Cambridge UK: Cambridge University Press, 2002.
- [21] A. M. Radzikowska and E. Kerre, "A Comparative study of fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 126, pp. 137-155, 2002.
- [22] *GroupLens Research*, MovieLens, 2014, Available at: <http://grouplens.org/>.
- [23] C. N. Ziegler, S. M. Mcnee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," *Proceedings of the 14th ACM International Conference on World Wide Web*, pp. 22-32, 2005.
- [24] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: a constant time collaborative filtering algorithm," *Proceedings of the 14th ACM International Conference on World Wide Web*, pp. 22-32, 2005.
- [13] L. H. Son and N. T. Thong, "Intuitionistic fuzzy recommender systems: an effective tool for medical diagnosis," *Knowledge-Based Systems*, 2014.
- [14] H. Zhang and F. Min, "Three-way recommender systems based on random forests," *Knowledge-Based Systems*, 2015.
- [15] Y. Sang, J. Liang and Y. Qian, "Decision-theoretic rough sets under dynamic granulation," *Knowledge-Based Systems*, 2015.
- [16] N. Verbiest, C. Cornelis, and F. Herrera, "FRPS: a fuzzy rough prototype selection method," *Pattern Recognition*, vol. 46, pp. 2770-2782, 2013.
- [17] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341-356, 1982.
- [18] S. Talatahari, B. Farahmand Azar, R. Sheikholeslami, and A. H. Gandomi, "Imperialist competitive algorithm combined with chaos for global optimization," *Communications in Nonlinear Science Numerical Simulation*, vol. 17, pp. 1312-1319, 2012.

زیرنویس‌ها

¹ Hybrid User-based Fuzzy Collaborative Filtering

² Game-Theoretic Rough Sets for Recommender Systems

³ Intuitionistic Fuzzy Recommender Systems

⁴ Pawlak

⁵ K Nearest Neighbor

⁶ Recommender Systems

⁷ Multi-Criteria Recommender Systems

⁸ Fuzzy Recommender Systems

⁹ Fuzzy Implicator

¹⁰ T-norm

¹¹ Colony

¹² Imperialist

¹³ 3-Fold Cross-Validation