فصلنامه نوآوری در مهندسی و تکنولوژی شماره ۲- دوره ۱- صفحات ۱۰۶-۸۴
Journal of Innovation in Engineering and Technology
ISSN 3060-6098

# بهینه سازی سیستم های توصیه گر فیلم با استفاده از الگوریتم های فراابتکاری فازی

زهرا محمدزاده [۱]*، محمدرضا اسلامی نژاد[۲]
۱ -کارشناس ارشد مهندسی کامپیوتر - نرم افزار، دانشکده مهندسی، موسسه آموزش عالی زند شیراز، استان فارس، ایران .
۲ -عضو هیات علمی گروه کامپیوتر، دانشکده مهندسی، موسسه آموزش عالی زند شیراز، استان فارس، ایران.

**چکیده :**
امروزه سیستم‌های توصیه‌گر (Recommendation Systems) با تحلیل رفتار کاربران، نقش مهمی در پیشنهاد دقیق و سریع اقلام به کاربران ایفا می‌کنند. این سیستم‌ها برای مواجهه با چالش‌هایی نظیر حجم عظیم اطلاعات، مشکل "شروع سرد" ناشی از کمبود داده‌های اولیه کاربران جدید، و پراکندگی اطلاعات طراحی شده‌اند. هدف این پژوهش ارائه روش نوین بهینه‌سازی مبتنی بر الگوریتم‌های فراابتکاری و خوشه‌بندی فازی است که دقت و سرعت سیستم‌های توصیه‌گر را در فروشگاه‌های اینترنتی فیلم افزایش می‌دهد. در روش پیشنهادی، از الگوریتم خوشه‌بندی فازی استفاده شده تا با دسته‌بندی داده‌ها، تعداد مقایسه‌ها کاهش یابد و سرعت توصیه بهبود یابد. برخلاف الگوریتم‌های سنتی که تنها از معیارهای امتیازدهی برای سنجش شباهت کاربران استفاده می‌کنند، این الگوریتم فاکتور زمان را نیز به عنوان پارامتر مهمی برای تحلیل رفتار کاربران در نظر می‌گیرد. این نوآوری به سیستم امکان می‌دهد تغییرات علایق کاربران در طول زمان را به‌خوبی مدیریت کرده و توصیه‌هایی با دقت بالاتر ارائه دهد. نتایج نشان می‌دهند که روش پیشنهادی در مقایسه با روش‌های دیگر، دقت بالاتری (تا ۹۹٪) دارد، میزان پوشش بیشتری ارائه می‌کند و زمان پاسخگویی به درخواست‌ها را به‌طور قابل‌توجهی کاهش می‌دهد. برای بررسی عملکرد، داده‌های واقعی از پایگاه‌های اینترنتی فیلم مانند IMDB استفاده شده و با دو سیستم توصیه‌گر فاقد الگوریتم فراابتکاری و فاقد الگوریتم خوشه‌بندی مقایسه شده است. نتیجه آزمایش‌ها نشان می‌دهد که الگوریتم‌های فراابتکاری و خوشه‌بندی فازی تأثیر مثبتی بر بهینه‌سازی ابعاد داده‌ها، افزایش دقت توصیه و کاهش زمان پردازش دارند. روش‌های فراابتکاری، شامل الگوریتم کلونی زنبور عسل برای کاهش ابعاد، در این پروژه مورد استفاده قرار گرفته‌اند که توانایی یافتن راه‌حل‌های دقیق و بهینه را به‌طور قابل‌توجهی ارتقا داده‌اند. علاوه بر این، خوشه‌بندی سلسله‌مراتبی فازی باعث شده تا داده‌ها تحت وزن‌های مختلف به خوشه‌ها تخصیص یابند و زمینه‌ای برای پیشنهادات پویا و مؤثر فراهم شود. با توجه به نتایج، روش پیشنهادی نه تنها دقت توصیه‌های سیستم را افزایش داده بلکه با کاهش خطای توصیه اقلام کم‌کیفیت، کیفیت کلی سیستم ارتقا یافته است .

**کلمات کلیدی:** سیستم توصیه گر، الگوی رفتاری کاربران، الگوریتم های فرا ابتکاری، الگوریتم های فازی

فصلنامه نوآوری در مهندسی و تکنولوژی شماره ۲- دوره ۱- صفحات ۱۰۶-۸۴
Journal of Innovation in Engineering and Technology
ISSN 3060-6098

**مقدمه**

با توجه به رشد روز افزون داده‌ها و اطلاعات در دنیای امروز، نیاز به فیلتر کردن اطلاعات برای استفاده مطلوب از آن‌ها بیش از پیش احساس می‌شود و بدون راهنمایی و هدایت درست، ممکن است انتخاب‌هایی غلط و یا غیر بهینه از میان آن‌ها داشته باشیم. در این میان سیستم‌های توصیه‌گر تأثیر بسزایی در راهنمایی و هدایت کاربران در میان حجم عظیمی از انتخاب‌های ممکن، برای رسیدن به گزینه مفید و موردعلاقه آن‌ها را دارند. سیستم‌های توصیه‌گر را زیرمجموعه‌ای از سیستم‌های پشتیبان تصمیم می‌دانند و از آن‌ها به عنوان سیستم‌های اطلاعاتی یاد می‌کنند که توانایی تحلیل رفتارهای گذشته و ارائه توصیه‌هایی برای مسائل جاری را دارا هستند. به زبان ساده‌تر در سیستم‌های توصیه‌گر تلاش بر این است تا با حدس زدن شیوه تفکر کاربر به کمک اطلاعاتی که از نحوه رفتار وی یا کاربران مشابه وی و نظرات آن‌ها موجود است، مناسب‌ترین و نزدیک‌ترین کالا به سلیقه او شناسایی و پیشنهاد شود. سیستم‌های توصیه‌گر، سیستم‌هایی هستند که برای پیشنهاد کردن آیتم‌هایی بکار برده می‌شوند که انتظار می‌رود این آیتم‌ها موردعلاقه کاربر قرار گیرند.

سیستم توصیه گر یا سامانه‌ی پیشنهادگر با تحلیل رفتار کاربر خود، اقدام به پیشنهاد مناسب‌ترین اقلام (داده، اطلاعات، کالا و ... ) به وی می‌نماید. این‌گونه سیستم‌ها درواقع جهت حل مشکلات ناشی از حجم فراوان و رو به رشد اطلاعات ارائه‌شده است و به کاربر خود کمک می‌کند تا در میان حجم عظیم اطلاعات، سریع‌تر به هدف خود دست یابد. سیستم‌های توصیه گر به کاربرانی که از بین حجم بالای اطلاعات به دنبال نوعی خاص از اطلاعات مرتبط با اولویت‌هایشان هستند، پیشنهادات شخصی شده‌ای را ارائه می‌دهد. این نوع سیستم‌ها با توانایی که در جمع‌آوری اطلاعات از رفتار و حرکت کاربران، دسته‌بندی و تفسیر آن‌ها دارند، امکانی فراهم آورده‌اند که کاربران با صرف زمان و انرژی کمتر به اطلاعات مناسب‌تری دسترسی پیدا کنند.

در سیستم‌های توصیه گر با مقیاس دنیای واقعی به دلیل خرید مقطعی کاربران(مانند خرید وسایل الکترونیکی) و یا به‌طورکلی کثرت تعداد آیتم‌های موجود جهت پیشنهاد، مشتری‌ها نوعاً درصد کمی از آیتم‌ها را نرخ‌گذاری می‌کنند. این مشکل را پراکندگی می‌نامیم. این مشکل زمانی رخ خواهد داد که تعداد نرخ‌ها در مقایسه با تعداد آیتم‌ها کوچک باشد. مسئله "شروع آهسته" نیز مشکل دیگری است که می‌توان آن را نوع خاصی از مشکل پراکندگی به‌حساب آورد. در ابتدای شروع به کار، سیستم تا مدتی نمی‌تواند پاسخ مناسبی ارائه کند؛ زیرا پایگاه دانش آن از تعداد محدودی نمونه ساخته‌شده است. این مشکل همچنین در هنگام کار با آیتم جدید و یا کاربر جدید نیز پیش می‌آید. برای بسیاری از مسائل ارائه‌شده، با گذر زمان و استفاده‌ی مکرر کاربران، پایگاه دانش با استفاده از اطلاعات کاربران تقویت می‌شود ولی در این فاصله‌ی زمانی، کارایی سیستم ممکن است از سطح قابل قبولی برخوردار نباشد.

در این پژوهش یک سیستم جدید بر مبنای الگوریتم های فراابتکاری فازی در فروشگاه های اینترنتی فیلم معرفی می‌گردد که دقت و سرعت توصیه را به خوبی افزایش می‌دهد. این اتفاق باید در زمان فیلترسازی انجام پذیرد. در فیلترسازی برای یافتن نزدیک‌ترین همسایگان باید شباهت بین تمام آیتم‌های موجود در سیستم محاسبه شود، که همین امر تأثیر بسزایی در کاهش سرعت این مدل دارد. که برای حل این مشکل از الگوریتم خوشه بندی استفاده می شود. در

خوشه بندی با استفاده از معیارهای مختلف موجود در کالاها می توان حجم عظیمی از داده ها را به صورت تفکیک شده در آورد و با این کار تعداد مقایسه های لازم جهت پیدا کردن کالای مورد نظر کاهش یافته و در نتیجه سرعت سیستم بالا می رود.

با استفاده از الگوریتم های خوشه بندی داده های موجود در سیستم به دسته های مختلف تقسیم بندی می شوند که این امر باعث می شود تا سیستم جهت پیدا کردن کالای مورد نظر کاربر تنها نیاز به وارد شدن به دسته مورد علاقه کاربر را داشته و کالا را از آنجا انتخاب و نمایش دهد. این کار باعث می شود تا در مقایسه با روش عادی تعداد مقایسه های کمتری انجام شده و در نتیجه سرعت و دقت سیستم بالا رود. در اینجا از الگوریتم خوشه بندی فازی استفاده می شود. در این حالت، با وزن‌های مختلف، داده‌ها به خوشه‌های مختلف تخصیص می‌یابد، لذا در هنگامی که می‌خواهیم یک آیتم را به کاربر پیشنهاد دهیم، ممکن است آیتم‌های با امتیاز بالا درون چند خوشه‌ی مختلف به کاربر تخصیص یابد.

علاوه بر این، ما بجز معیار شباهت و امتیازدهی، از فاکتور زمان نیز به عنوان پارامتر مهمی برای اندازه گیری تشابه کاربران برای خوشه بندی استفاده می کنیم. در روش‌های سنتی، شباهت بر اساس امتیازدهی تعیین می شود، بدین گونه که اگر دو کاربر به یک آیتم رای بدهند سیستم نتیجه می‌گیرد که شباهتی بین این دو کاربر وجود دارد. نوآوری ما از این جنبه است که ما از یک روش جدید در خوشه‌بندی استفاده کرده‌ایم که علاوه بر معیار شباهت و امتیازدهی، فاکتور زمان را نیز به عنوان پارامتر مهمی برای اندازه‌گیری تشابه کاربران در نظر می گیرد. این رویکرد، به خصوص مواقعی که علایق کاربران در طول زمان تغییر می کند بسیار کارآمد است. رتبه‌بندی آیتم‌ها بر اساس الگوی رفتاری کاربران در زمان‌های مختلف، امکان ارائه توصیه‌های بهتر و با دقت بیشتری را فراهم می‌سازد.

از این رو به نظر می‌رسد نوآوری‌های بکار رفته در این پژوهش ، جهت استفاده از خوشه‌بندی فازی قادر به افزایش دقت توصیه‌های ارائه شده، می‌باشد. نتایج پیاده سازی روش پیشنهادی و اعمال آن بر روی مجموعه داده های واقعی بیانگر کیفیت مناسب روش پیشنهادی در مقایسه با روش پایه و سایر مطالعات است.

## بیان مسئله

با توسعه سیستم‌های اطلاعاتی، داده به یکی از منابع پراهمیت سازمان‌ها مبدل گشته است. بنابراین روش‌ها و تکنیک‌هایی برای دستیابی کاربر به داده، اشتراک داده، استخراج اطلاعات از داده و استفاده از این اطلاعات، موردنیاز می‌باشد. با ایجاد و گسترش وب و افزایش چشمگیر حجم اطلاعات، نیاز به این روش‌ها و تکنیک‌ها بیش‌ازپیش احساس می‌شود. سیستم توصیه‌گر یک سیستم ابتکاری است که اطلاعات مفید را پیشنهاد می‌دهد و می‌تواند در دامنه‌های گوناگون بکار رود. سیستم توصیه‌گر یا سامانه‌ی پیشنهادگر با تحلیل رفتار کاربر خود، اقدام به پیشنهاد مناسب‌ترین اقلام (داده، اطلاعات، کالا و ... ) به وی می‌نماید. این‌گونه سیستم‌ها در واقع جهت حل مشکلات ناشی از حجم فراوان و رو به رشد اطلاعات ارائه شده است و به کاربر خود کمک می‌کند تا در میان حجم عظیم اطلاعات، سریع‌تر به هدف خود دست یابد. این نوع سیستم‌ها با توانایی‌ای که در جمع‌آوری اطلاعات از رفتار و حرکت کاربران، دسته‌بندی و تفسیر آن‌ها دارند، امکانی فراهم آورده‌اند که کاربران با صرف زمان و انرژی کمتر به اطلاعات مناسب‌تری دسترسی پیدا کنند.

روش فیلترسازی جمعی[۱](CF) با بهره‌برداری از اطلاعات مربوط به رفتار گذشته یا عقاید موجود جامعه کاربران، محصولاتی را پیش‌بینی می کند که احتمالاً موردپسند یا علاقه کاربر جاری سیستم خواهد بود. فیلترسازی جمعی از پرکاربردترین تکنیک‌ها در سیستم‌های توصیه‌گر می‌باشد که توصیه‌هایی را با کمترین تأخیر به کاربران ارائه می‌دهد. با وجود این هنوز هم این مدل از سرعت کم رنج می‌برد. سیستم‌های توصیه‌گر فیلترسازی مبتنی بر آیتم پیش‌بینی را از طریق سابقه همسایگان آیتم‌ها محاسبه می‌کنند که انتخاب درست نزدیک‌ترین همسایگان که بیشترین شباهت را با آیتم مورد نظر داشته باشند تأثیر زیادی بر افزایش دقت پیش‌بینی‌ها خواهد داشت. در فیلترسازی جمعی برای یافتن نزدیک‌ترین همسایگان باید شباهت بین تمام آیتم‌های موجود در سیستم محاسبه شود، که همین امر تأثیر بسزایی در کاهش سرعت این مدل دارد. با توجه به این شرایط، خوشه‌بندی داده‌ها می‌تواند تعداد جفت مقایسه‌های لازم را کاهش دهد و تنها با محاسبه شباهت بین اعضایی که در یک خوشه قرار دارند پیش‌بینی‌ها را تولید نمایند. در این شرایط و با توجه به عدم قطعیت داده‌ها، ممکن است آیتم‌ها به چندین خوشه تعلق داشته باشند. در این شرایط الگوریتم‌های خوشه‌بندی فازی بهتر می‌توانند خوشه‌بندی را انجام دهند.

در این پژوهش ما یک سیستم جدید بر مبنای خوشه‌بندی فازی برای کاهش مشکل فوق پیشنهاد می‌دهیم. در این پژوهش از الگوریتم های فراابتکاری فازی در فروشگاه های اینترنتی فیلم استفاده می‌گردد. در این حالت، با وزن‌های مختلف، فیلم ها به خوشه‌های مختلف تخصیص می‌یابد، لذا در هنگامی که می‌خواهیم یک فیلم را به کاربر پیشنهاد دهیم، ممکن است فیلم های با امتیاز بالا درون چند خوشه‌ی مختلف به کاربر تخصیص یابد.

**مروری بر مطالعات انجام‌شده**

سیستم‌های توصیه‌گر فیلم در سال‌های اخیر توجه بسیاری از پژوهشگران را به خود جلب کرده است. با توجه به اینکه هدف اصلی در این سیستم‌ها ارائه پیشنهادهای شخصی‌سازی‌شده و بهینه برای کاربران است، تحقیقاتی متعددی درباره روش‌ها و الگوریتم‌های مختلف برای بهبود دقت و کارایی این سیستم‌ها انجام شده است. در ادامه، بررسی اجمالی از مجموعه مقالات مرتبط با این موضوع ارائه می‌شود.

Sneha و همکاران در سال ۲۰۱۲ در مقاله‌ای با عنوان "یک سیستم توصیه گر شخصی‌سازی‌شده مبتنی بر محصولات با استفاده از تحلیل وب و وب معنایی" به بررسی استفاده از داده‌های وب و تکنیک‌های معنایی برای ارائه پیشنهادهای بهتر پرداختند. نتایج این مطالعه نشان داد که استفاده از تکنیک‌های معنایی می‌تواند دقت سیستم‌های توصیه‌گر را در حوزه محصول بهبود بخشد

Zhang و همکاران در سال ۲۰۱۳ در پژوهش "سیستم توصیه‌گر شخصی‌سازی شده هیبریدی مبتنی بر فازی برای خدمات و محصولات مخابرات" از ترکیبی از الگوریتم‌های فازی و تکنیک‌های هیبریدی استفاده کردند. این روش توانست نیازهای کاربری را به‌طور دقیق‌تری مورد بررسی قرار دهد و رضایت کاربران را افزایش دهد

[۱] Collabrative Filtering (CF)

Zhu و همکاران در سال ۲۰۱۴ با ارائه مقاله "یک الگوریتم حفاظت از حریم خصوصی موثر برای فیلترینگ مشارکتی مبتنی بر همسایگی" تلاش کردند تا رویکردی برای حفظ حریم خصوصی در سیستم‌های توصیه‌گر مبتنی بر همکاری ارائه دهند. روش پیشنهادی با کاهش خطرات ناشی از افشای اطلاعات حساس، موفق به بهبود عملکرد سیستم شد

Pham و همکاران در سال ۲۰۱۱ مقاله‌ای با عنوان "رویکرد خوشه‌بندی برای توصیه‌های فیلترینگ مشارکتی با استفاده از تحلیل شبکه اجتماعی" منتشر کردند. آن‌ها نشان دادند که ادغام اطلاعات شبکه اجتماعی در سیستم‌های توصیه‌گر می‌تواند دقت و کارایی این سیستم‌ها را افزایش دهد

Bokde و Girase در سال ۲۰۱۵ با مقاله "فیلترینگ مشارکتی مبتنی بر آیتم با استفاده از تکنیک‌های کاهش ابعاد در چارچوب Mahout" به بررسی تاثیر کاهش ابعاد بر روی کارایی سیستم‌های توصیه‌گر پرداختند. نتایج این تحقیق نشان داد که استفاده از این روش می‌تواند حجم محاسبات را کاهش داده و دقت توصیه‌ها را حفظ کند

Li و همکاران در سال ۲۰۱۴ با پژوهش خود تحت عنوان "استفاده از الگوریتم‌های ژنتیک و نظریه گراف برای ماتریس‌سازی عامل توصیه دوست آنلاین"، پیشنهاد کردند که ادغام الگوریتم‌های فراابتکاری نظیر الگوریتم ژنتیک در سیستم‌های توصیه‌گر می‌تواند عملکرد این سیستم‌ها را بهبود بخشد

Adomavicius و Zhang در سال ۲۰۱۴ در مقاله "روش صاف‌سازی تکراری برای بهبود پایداری سیستم‌های توصیه‌گر" به بررسی چگونگی افزایش دقت و قابلیت اعتماد سیستم‌های توصیه‌گر پرداختند. روش صاف‌سازی تکراری در این مقاله به عنوان راهکاری موثر شناخته شد

Candillier و همکاران در سال ۲۰۱۳ مقاله‌ای با عنوان "تنوع در سیستم‌های توصیه‌گر: پل زدن شکاف بین کاربران و سیستم‌ها" منتشر کردند. آن‌ها تاکید کردند که تنوع در پیشنهادها می‌تواند رضایت کاربران را افزایش داده و تجربه کاربری بهتری ایجاد کند

Peska و Vojtas در سال ۲۰۱۵ با ارائه مقاله "پیش‌بینی محبوبیت توییت‌های فیلم با استفاده از شباهت‌های k-NN" به استفاده از روش‌های مبتنی بر شباهت پرداختند. پژوهش آن‌ها نشان داد که پیش‌بینی‌های مبتنی بر محتوای توییت‌ها می‌تواند کاربردهای جالبی در توصیه فیلم داشته باشد

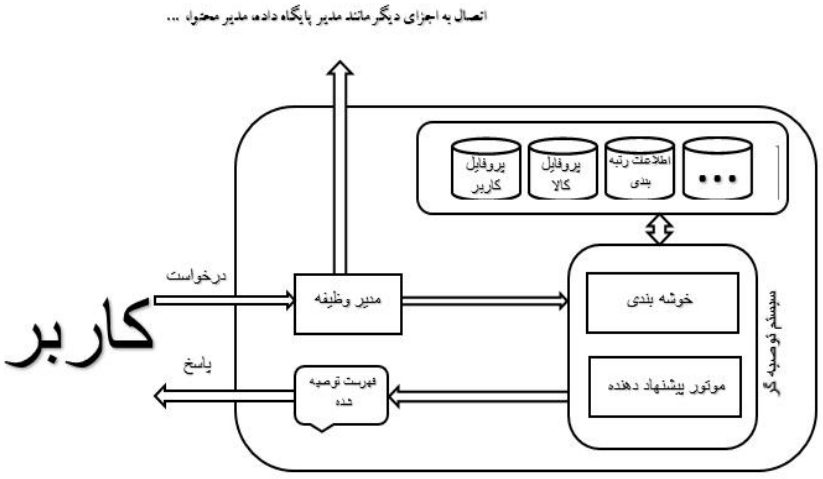
Akbari و همکاران در سال ۲۰۱۳، پژوهشی با عنوان "توصیه‌گر دوست مبتنی بر گراف در شبکه‌های اجتماعی با استفاده از الگوریتم کلونی مصنوعی زنبورها" منتشر کردند. این روش از الگوریتم زنبور عسل برای یافتن دوستان مرتبط در شبکه‌های اجتماعی استفاده کرد و نتایج قابل توجهی در بهبود ارتباطات کاربران ارائه شد.

با توجه به رشد الگوریتم‌های یادگیری ماشین و تحلیل داده‌های کاربر، پژوهش‌های مختلفی از جمله استفاده از محاسبات فازی، کاهش ابعاد داده و الگوریتم‌های فراابتکاری در بهینه‌سازی سیستم‌های توصیه‌گر انجام شده است. این مطالعات پایه‌های محکمی برای طراحی سیستم‌های توصیه‌گر امروزی ارائه می‌دهند.

## روش پیشنهادی

افزایش چشمگیر حجم اطلاعات دیجیتال موجود، منابع الکترونیکی و خدمات آنلاین، منجر به ایجاد مشکلی بنام سرریز اطلاعات شده است - جایی که مردم برای پیدا کردن اطلاعات مورد نیاز خود از مجموعه ای از گزینه های گوناگون دچار مشکل می شوند. امروزه مسئله این نیست که چگونه اطلاعات صحیح برای تصمیم گیری دریافت می شود، بلکه، چگونگی تصمیم گیری درست از مقدار عظیمی از اطلاعات است. این امر باعث ایجاد و توسعه سیستم های فیلترینگ اطلاعات مانند سیستم های توصیه گر (RS) شده است. RS به طور کاربردی از نظرات گروهی استفاده می کند تا به افرادی که در این گروه هستند کمک کند، به طور مؤثرتر محتوای مورد علاقه از یک فضای جستجوی بالقوه را شناسایی کنند. برخی از حوزه های کاربردی برای RS عبارتند از: توصیه هایی برای CD های موسیقی و DVD ها، فیلم ها و کتاب ها.

شکل ۱ چارچوبی کلی را برای سیستم های توصیه گر که شامل سه جزء اصلی است را نشان می دهد: مدیر وظیفه، سیستم توصیه گر و پایگاه های داده. مدیر وظیفه شامل رابط کاربری است که درخواست را از کاربر دریافت می کند و تصمیم می گیرد تا اقدامات مناسب را انجام دهد. درخواست در دامنه تجارت الکترونیک شامل ایجاد اولین خرید و تبدیل شدن به یک مشتری جدید، خرید محصولات بیشتر، لغو سفارشات و غیره است. سیستم توصیه گر از الگوریتم های مختلف جهت خوشه بندی اطلاعات و محاسبه شباهت ها و ارائه پیشنهاد به کاربر استفاده می کند که به طور کامل شرح داده می شود. اجزای پایگاه داده شامل پروفایل های مشتری، پروفایل های کالا، اطلاعات رتبه بندی و غیره می باشد. مشخصات مشتری شامل ویژگی های مختلف کاربر، مانند سن، جنس، درآمد، وضعیت تاهل و غیره است. مشخصات کالا با مجموعه ای از ویژگی ها تعریف شده است؛ به عنوان مثال، در یک برنامه درخواست فیلم، هر کالا (فیلم) را می توان با شناسه، عنوان، ژانر، کارگردان، سال انتشار، بازیگران مشهور و غیره نشان داد. همچنین، مولفه اطلاعات ارزیابی شامل رتبه بندی هایی است که به کالاها توسط مشتریان اختصاص داده می شود . در سیستم ما، کاربران بر اساس ترجیحات خود خوشه بندی می شوند و بر اساس محتویات آنها خوشه بندی می شوند. ساختارهای این خوشه ها در مولفه ساختار خوشه ای نگهداری می شوند که به صورت پویا با استفاده از ماژول خوشه بندی پویا در طول زمان به روز می شود.



*شکل ۱. چارچوب کلی سیستم توصیه گر*

شروع سیستم

همانطور که قبلا گفته شد، در ابتدای کار سیستم با مشکلی به نام شروع سرد(هم برای کاربر و هم برای کالا) روبرو خواهد بود که در مرحله اول باید این مشکل را حل نماییم. مشکل سرد برای کالا و کاربر بدین صورت خواهد بود که:

۱. برای کاربر: سیستم هیچگونه شناختی در مورد کاربر اعم از سن، جنسیت، میزان تحصیلات، حوزه مورد علاقه و ... را ندارد و باید یک پروفایل کاربری برای این کاربر ایجاد نماید.

۲. برای کالا: با توجه به اینکه برای خرید کالا توسط کاربران نیاز این مورد است که کالا در خوشه ای خاص قرار گرفته و به کاربران نمایش داده شود تا امتیاز دریافت کند لذا این امر الزامیست که کالای جدید به کاربران (جدید و قدیم) نمایش داده شود تا بتوان توسط امتیازگیری ها آنها را خوشه بندی کرده و در نمایش های بعدی با مشخصات بهتری نمایش داده شود.

تاکنون روش های مختلفی برای مقابله با مشکلات مذکور ارائه شده است که روش پیشنهادی ما بدین گونه خواهد بود:

۱. برای کاربر: برای ایجاد پروفایل کاربر ما نیاز به دانستن یک سری اطلاعات صریح و ضمنی( ذکر شده در جدول ۱-۱ ) از کاربر خواهیم داشت که برای بخش داده های صریح می توان از فرم های ثبت نام و یا چک باکس های سریع موجود در سیستم استفاده کرد و همچنین برای بخش داده های ضمنی با گذر زمان از تعداد بازدید ها و یا خرید محصولات استفاده می نماییم.

۲. برای کالا: جدیدترین محصولات در قسمتی خاص به همان نام (جدید ترین محصولات) به کاربران (جدید و قدیمی) معرفی و نمایش داده می شود تا از این طریق بتواند از کاربران امتیاز دریافت کرده و جایگاه خود را در خوشه های عضوتغییر دهد.

در شروع کار با ورود کاربر به سیستم، جمع آوری اطلاعات جهت ساخت پروفایل کاربر آغاز می شود. داده های صریح با استفاده از فرم های ثبت نام و همچنین چک باکس های سریع موجود در سیستم از کاربر دریافت می شود که شامل سن، جنسیت، شغل، سطح تخصص، حوزه مورد علاقه و ... می باشد. داده های ضمنی نیز با گذر زمان و طبق بازدید ها و خرید های مشتری دریافت و به پروفایل کاربر اضافه می شود.

سیستم بجز گرفتن داده ها جهت تکمیل اطلاعات و ساخت پروفایل کاربر وظیفه نمایش همزمان کالاهای موجود در سیستم به مشتری را دارد. این امر وظیفه سیستم توصیه گر خواهد بود که همزمان با دریافت اطلاعات از کاربر باید کالاهای مورد نظر کاربر را به آن نمایش دهد تا بتوان از این طریق رضایت کاربر و فروشنده را جلب کند. سیستم توصیه گر از هنگام دریافت اطلاعات تا زمان پیشنهاد محصول مراحل مختلفی را طی می کند که انجام این مراحل زمانبر است. ما سعی داریم با پیشنهاد سیستم توصیه گر مورد نظر خود زمان طی شدن این مراحل را کاهش داده و با این روش همچنین بتوانیم دقت سیستم را جهت پیشنهاد محصولات به کاربران افزایش دهیم.

سیستم توصیه گر

فیلم ها بخشی از زندگی تقریبا همه افراد هستند تا بتوانند نیازهای سرگرمی خود را برآورده کنند و از این رو بخش بزرگی از صنعت سرگرمی را تشکیل می دهند. وب سایت های متعددی ابرداده های فیلم را میزبانی می کنند و کاربران را قادر به جستجو و پیدا کردن فیلم های مورد علاقه خود می کنند. پایگاه داده فیلم اینترنتی[۲] (IMDB) یک سایت محبوب است که تقریبا هر فیلم ساخته شده در آن موجود است. این یک پایگاه داده آنلاین[۳] عالی برای پیدا کردن اطلاعات دقیق در مورد فیلم ها، سریال ها و ویدئو ها است. IMDB حاوی اطلاعاتی مانند ژانر، کلمات کلیدی، سال، زبان، رتبه بندی کاربر و بسیاری از ویژگی های دیگر مربوط به آن فیلم ها و ویدیو ها است. با این حال، IMDB حاوی مقدار زیادی از داده ها است که باید تجزیه و تحلیل شود.

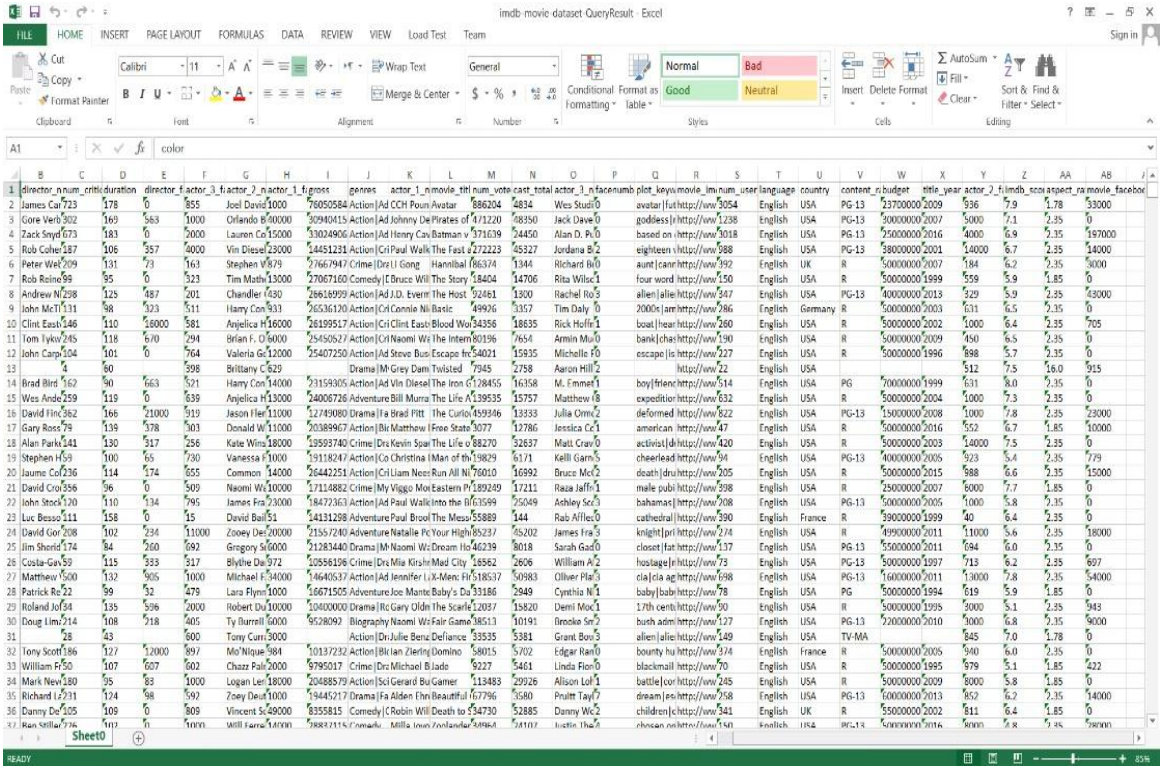
در این پژوهش یک سیستم توصیه گر نوین بر مبنای الگوریتم های فراابتکاری فازی در فروشگاه های اینترنتی فیلم معرفی می‌کنیم که دقت و سرعت توصیه را به خوبی افزایش می‌دهد. برای این کار، سیستم توصیه گر را به مراحل مختلفی تقسیم می کنیم که هر مرحله شامل عملیات گوناگونی جهت بهینه تر کردن سیستم خواهد شد. ما از مجموعه داده های ذکر شده به عنوان ورودی سیستم استفاده می کنیم. در مرحله ۱، با توجه به اینکه تعداد فیلم های موجود در هر ژانر رو به افزایش است و تجزیه و تحلیل اطلاعات آنها زمان زیادی را از ما می گیرد در نتیجه با استفاده از الگوریتم های فراابتکاری شروع به کاهش ابعاد و بهینه سازی آنها می کنیم. بعد از بهینه سازی و کاهش تعداد فیلم ها، در مرحله ۲ فیلم ها را با استفاده از ویژگی های خاصشان خوشه بندی کرده (در اینجا با استفاده از ژانر و کلمات کلیدی) و مرحله دوم بهینه سازی را انجام می دهیم. در مراحل ۳ و۴ موارد مشابه را برای نمایش به کاربر هدف پیدا کرده و به آن توصیه می شوند. این چهار مرحله در زیر به تفصیل توضیح داده شده اند.

دریافت مجموعه داده ها و کاهش ابعاد

همانطور که گفته شد در اینجا ما از مجموعه داده های موجود در فروشگاه اینترنتی فیلم IMDB استفاده خواهیم کرد (شمایی از آن در شکل ۲ آورده شده است). مجموعه داده در هر سطر ویژگی های هر فیلم را بصورت جداگانه در ستون های مختلف، مانند ژانرها در genres، ارائه می دهد. ما از اطلاعات موجود در ستون های زیر استفاده کرده ایم: movie_title، genres، plot_keywords، imdb_score، title_year و language. همانطور که اسمشان نشان می دهد، ستون movie_title شامل اطلاعات نام فیلم، ستون genres شامل اطلاعات ژانرهای فیلم ، ستون plot_keywords شامل اطلاعات کلمات کلیدی موجود در فیلم، ستون imdb_score شامل اطلاعات امتیاز کسب شده از کاربران برای هر فیلم، ستون title_year سال ساخت فیلم و ستون language شامل اطلاعات زبان محاوره ای فیلم است.

[۲] Internet Movie Database
[۳] www.imdb.com

فصلنامه نوآوری در مهندسی و تکنولوژی شماره ۲- دوره ۱- صفحات ۱۰۶-۸۴
Journal of Innovation in Engineering and Technology
ISSN 3060-6098



شکل ۲: مجموعه داده IMDB

با توجه به اینکه تعداد فیلم های موجود در هر ژانر با گذشت زمان رو به افزایش است و این امر باعث می شود که سرعت تجزیه و تحلیل سیستم توصیه گر کاهش پیدا کند بنابراین بهینه سازی و کاهش تعداد این فیلم ها جهت افزایش سرعت سیستم الزامی به نظر می آید. در نتیجه باید از روشهای بهینه سازی برای کاهش تعداد فیلم ها و انتخاب بهترین آنها (در اینجا با توجه به امتیاز فیلم ها) بهره جست که روش پیشنهادی ما الگوریتم فراابتکاری زنبور عسل است. با استفاده از این الگوریتم هم می توان بر مشکل حجم وسیع داده ها غلبه کرد و هم با توجه به اینکه تنها کالاهای نخبه انتخاب و وارد سیستم می شوند در نتیجه کالای پیشنهادی به کاربر حتما از لحاظ کیفیت و امتیاز مورد رضایت بوده و دقت سیستم نیز بالا خواهد رفت. این الگوریتم در دو قسمت اجرا می شود که قسمت اول آن در مرحله ۱ جهت کاهش ابعاد مجموعه داده اولیه و قسمت دوم آن در مرحله ۲ جهت انتخاب بهترین فیلم های موجود در هر خوشه قرار داده شده است.

قبل از اینکه چگونگی اجرای الگوریتم بهینه سازی زنبور عسل در روش پیشنهادی خود را توضیح دهیم ابتدا لازم است توصیفی کلی از چگونگی انجام این روش را ذکر کنیم.

**الگوریتم بهینه سازی زنبور عسل**

در طبیعت و در فصل تابستان هر کلونی زنبور عسل می بایست به جمع آوری غذا بپردازد. در این زمان هر کلونی بخشی از جمعیت بالغ خود را (معمولا ۲۵٪) تحت عنوان زنبورهای پیشاهنگ گزینش کرده و آنها را به منظور جستجوی فضای اطراف کندو می فرستد. آنها به دنبال نواحی از گلها می گردند که دارای نکتار کافی بوده، شهد گیری آنها آسان و دارای مقدار کافی از منابع قندی باشند. این زنبور ها منابع غذا را بررسی کرده و در زمانی که منبع غذا را پیدا می کنند مقداری از نکتار آن منبع را برداشته و بدون هیچ مشکلی به کندو بازمیگردند. زنبور ها نکتاری را که فراهم کرده اند را آزاد می کنند. زنبورهایی که منابع غذایی بهتری را پیدا کرده اند موقعیت این منابع را به سایر زنبور های موجود در کندو اطلاع می دهند. پس از این مرحله، زنبورها مجددا به سوی محل گلها بازگشته و توسط سایر زنبور های کندو همراهی می شوند. تعداد زنبورهای همراهی کننده برای هر پیشاهنگ به کیفیت منبع غذایی یافت شده توسط آن وابسته است. توده گلهایی که دارای منبع غذایی بیشتر هستند و استفاده از منابع آنها ساده تر و با صرف انرژی کمتر می باشد، تعداد بیشتری از زنبور های جدید را به خود جذب می کنند. هدف اصلی کلونی زنبورهای عسل یافتن بهترین منابع غذایی با صرف کمترین انرژی و بیشترین سطح غذا می باشد.

الگوریتم بهینه سازی زنبورهای عسل از رویه این موجودات به منظور یافتن منابع غذا الهام گرفته شده است. هر پاسخ در فضای حل مساله به عنوان یکی از منابع غذا شناخته می شود. در این مرحله، ما تنها از قسمت اول الگوریتم بهینه سازی استفاده می کنیم. به این صورت که، به تعداد ژانرهای موجود در مجموعه داده زنبور پیشاهنگ به سیستم ارسال می کنیم و به هر کدام دستور می دهیم تا N عدد از بهترین فیلم ها (از نظر امتیاز کاربری) را انتخاب و به کندو (سیستم توصیه گر) اعلام نماید. با این کار بجای اینکه سیستم مجبور باشد با تعداد زیادی از فیلم ها که شاید بعضی از آنها امتیاز بسیار پایینی داشته باشند کار کند و باعث کاهش سرعت آن شود، با تعداد خاصی از فیلم ها در ژانر های مختلف که امتیاز کاربری آنها بالاست و دقت و سرعت سیستم را بالا می آورند کار می کند. با این کار مجموعه داده موجود برای تجریه و تحلیل سیستم را کاهش داده ایم که این کار باعث افزایش سرعت آن می شود و همچنین با توجه به اینکه در هر ژانر بهترین های آنها انتخاب شده است بنابراین دقت سیستم برای پیشنهاد به کاربر هدف نیز افزایش خواهد یافت. در نتیجه در این مرحله ما توانستیم مجموعه کل داده ها که شامل اطلاعات تمام فیلم های موجود در سیستم است را به الگوریتم بهینه سازی داده و مجموعه ای بهینه شده از اطلاعات را برای ورود به مرحله بعد آماده نماییم.

### خوشه بندی سلسله مراتبی

داده های فیلم شامل تعدادی از کلمات کلیدی برای توصیف هر فیلم است. فیلم ممکن است به چند ژانر متعلق باشد. بنابراین، ژانر هر فیلم شامل مجموعه های کلمات کلیدی که برای آن فیلم مشخص شده است می شوند. ما در اینجا برای خوشه بندی کردن فیلم ها با کمک گرفتن از ستون های ژانر و کلمات کلیدی از خوشه بندی سلسله مراتبی استفاده کرده ایم. به عنوان مثال، اجازه دهید فیلم های زیر را با مجموعه های کلیدی و ژانر مرتبط کنیم:

فیلم۱:

- کلمات کلیدی فیلم۱: {کلمه۱، کلمه۲، کلمه۳، کلمه۴، کلمه۵}
- ژانرهای فیلم۱: {ژانر۱، ژانر۲، ژانر۳}

فیلم۲:

- کلمات کلیدی فیلم۲: { کلمه۲، کلمه۳، کلمه۷}
- ژانرهای فیلم۲: { ژانر۱، ژانر۲}

فیلم۳:

- کلمات کلیدی فیلم۳: { کلمه۱، کلمه۶، کلمه۸}
- ژانرهای فیلم۳: { ژانر۲، ژانر۳}

مجموعه کلید واژه های هر ژانر را به صورت زیر تنظیم می کنیم:

مرحله ۱: ترکیب همه کلمات کلیدی موجود در فیلم هایی که ژانر مورد نظر در لیستشان قرار دارد.

- کلمات کلیدی ژانر۱: { کلمه۱، کلمه۲، کلمه۲، کلمه۳، کلمه۳، کلمه۴، کلمه۵، کلمه۷}
- کلمات کلیدی ژانر۲: { کلمه۱، کلمه۱، کلمه۲، کلمه۲، کلمه۳، کلمه۳، کلمه۴، کلمه۵، کلمه۶، کلمه۷، کلمه۸}
- کلمات کلیدی ژانر۳: { کلمه۱، کلمه۱، کلمه۲، کلمه۳، کلمه۴، کلمه۵، کلمه۶، کلمه۸}

مرحله ۲: یک وزن به هر کلمه در مجموعه کلمات کلیدی مرتبط می شود. وزن هر کلمه با توجه به نرمال کردن کل وزن کلید واژه های مجموعه به ۱ بدست می آید. وزن کلید واژه وابسته به تعداد تکرار کلمات کلیدی در مجموعه کلمه کلیدی است.

- کلمات کلیدی ژانر۱: {< کلمه۱، ۰.۱۲۵ >،< کلمه۲، ۰.۲۵ >،< کلمه۳، ۰.۲۵ >،< کلمه۴، ۰.۱۲۵ >،< کلمه۵، ۰.۱۲۵ >،< کلمه۷، ۰.۱۲۵ >}
- کلمات کلیدی ژانر۲: {< کلمه۱، ۰.۱۸۱۸ >،< کلمه۲، ۰.۱۸۱۸ >،< کلمه۳، ۰.۱۸۱۸ >،< کلمه۴، ۰.۰۹۰۹ >،< کلمه۵، ۰.۰۹۰۹ >،< کلمه۶، ۰.۰۹۰۹ >،< کلمه۷، ۰.۰۹۰۹ >،< کلمه۸، ۰.۰۹۰۹ >}
- کلمات کلیدی ژانر۳: {< کلمه۱، ۰.۲۵ >،< کلمه۲، ۰.۱۲۵ >،< کلمه۳، ۰.۱۲۵ >،< کلمه۴، ۰.۱۲۵ >،< کلمه۵، ۰.۱۲۵ >،< کلمه۶، ۰.۱۲۵ >،< کلمه۸، ۰.۱۲۵ >}

مرحله ۳: پس از ترکیب کلی کلمات کلیدی فیلم برای تمام ژانرها یک ماتریس شکل می گیرد که ردیف ها نشانگر کلمات کلیدی هستند و ستون ها ژانر را نشان می دهند.

**جدول۱: ماتریس کلمات کلیدی-ژانر**

| | ژانر۱ | ژانر۲ | ژانر۳ |
|---|---|---|---|

| | | | |
|---|---|---|---|
| کلمه۱ | ۰.۱۲۵ | ۰.۱۸۱۸ | ۰.۲۵ |
| کلمه۲ | ۰.۲۵ | ۰.۱۸۱۸ | ۰.۱۲۵ |
| کلمه۳ | ۰.۲۵ | ۰.۱۸۱۸ | ۰.۱۲۵ |
| کلمه۴ | ۰.۱۲۵ | ۰.۰۹۰۹ | ۰.۱۲۵ |
| کلمه۵ | ۰.۱۲۵ | ۰.۰۹۰۹ | ۰.۱۲۵ |
| کلمه۶ | ۰.۰ | ۰.۰۹۰۹ | ۰.۱۲۵ |
| کلمه۷ | ۰.۱۲۵ | ۰.۰۹۰۹ | ۰.۰ |
| کلمه۸ | ۰.۰ | ۰.۰۹۰۹ | ۰.۱۲۵ |

فاصله و مقادیر ضریب همبستگی پیرسون بین ژانرها پس از تعیین مقادیر کلیدی کلمات برای ژانرها محاسبه می شود. این مقادیر را در جدولی جمع آوری کرده تا بتوان از آنها برای مقایسه استفاده کرد.

با توجه به اینکه ۲۷ ژانر فیلم وجود دارد که در مجموع ۱۹۵۶۱ کلمه کلیدی را شامل می شود، به منظور کشف روابط ژانر، خوشه بندی سلسله مراتبی یک روش مناسب خواهد بود. خوشه بندی سلسله مراتبی اطلاعات را به ساختار سلسله مراتبی بر اساس نزدیکی داده ها با یکدیگر سازماندهی می کند. خوشه بندی متراکم، یک روش است که به طور گسترده ای برای خوشه بندی سلسله مراتبی استفاده می شود، با خوشه های تک سلولی آغاز می شود، که هر کدام شامل یک داده واحد هستند، و در هر مرحله یک سری عملیات ادغام را انجام می دهند تا یک خوشه باقی بماند. نتیجه معمولا توسط یک دندروگرام نشان داده می شود که ساختارهای خوشه بندی بالقوه را نشان می دهد. با بریدن دندروگرام در سطوح مختلف، می توان ساختارهای مختلف خوشه بندی را بدست آورد.

هنگام ترکیب دو خوشه در هر سطح، از الگوریتم فاصله بین دو خوشه به عنوان الگوریتم پیوند کامل استفاده می کنیم. الگوریتم پیوند کامل برای خوشه های کوچک موثر است. همچنین تضمین می کند که تمام اقلام در حداکثر فاصله یکدیگر قرار دارند، به این معنی که از فاصله ای بین اقلام خوشه ها برای تعریف فاصله بین خوشه ای استفاده می کند.

اکنون با توجه به مقادیر فاصله و ضریب همبستگی پیرسون بین ژانرها شروع به خوشه بندی آنها نموده تا به تعداد خوشه های مورد نظر دست پیدا کنیم. در اینجا ما دندوگرام بدست آمده (در شکل ۳) را در فاصله ۰.۵ قطع می کنیم که با این کار ۵ خوشه اصلی را برای تعریف به سیستم خواهیم داشت. این ۵ خوشه هر کدام شامل ژانرهای مختلفی بوده که بدین شرح خواهند بود (که می توان با توجه به نظر طراح سیستم توصیه گر این دندوگرام را در سطوح مختلف دیگری برش داد):

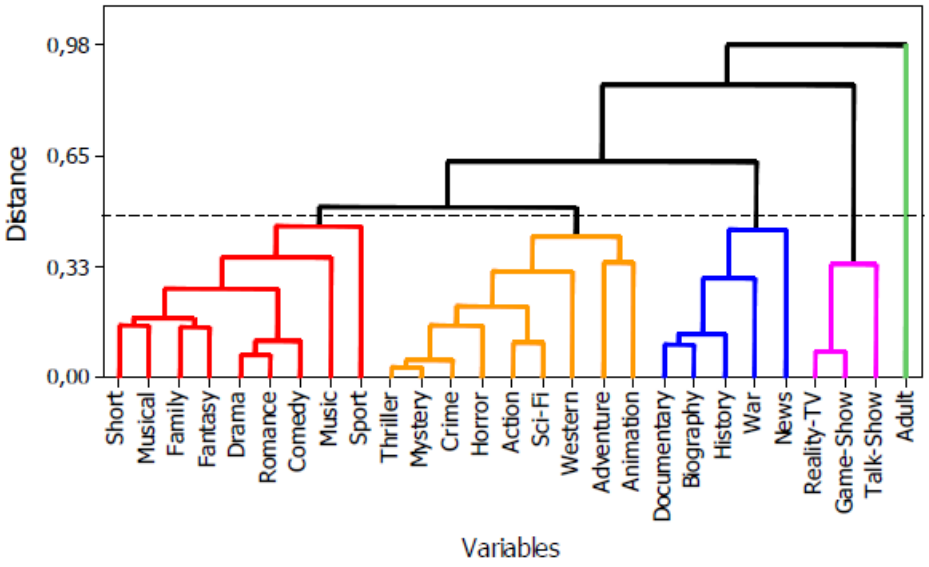
۱- کوتاه، درام، کمدی، رمانتیک، خانوادگی، موزیک، فانتزی، ورزشی، موزیکال

۲- هیجان انگیز، ترسناک، اکشن، جنایی، حادثه ای، افسانه ای، معمایی، انیمیشن، وسترن

۳- مستند، بیوگرافی، تاریخی، جنگی، اخبار

٤- برنامه زنده، برنامه گفتگو، برنامه مسابقه

٥- بزرگسال




*شکل ۳: دندوگرام حاصل از خوشه بندی سلسله مراتبی*

حال، با مشخص شدن ژانرهای موجود در خوشه های بدست آمده با توجه به اینکه فیلم های موجود در مجموعه داده ها هر کدام به تنهایی شامل چند ژانر مختلف می شود در نتیجه باید هر فیلم را بر اساس ژانرهایش بصورت فازی به خوشه های ایجاد شده نسبت دهیم. بدین صورت که با توجه به فرمول (۱) هر فیلم را با توجه به درجه عضویتشان(تعداد ژانرهایش) به خوشه ها اضافه می کنیم.

$$\begin{cases} \cdot \\ \frac{p}{q} \\ ۱ \end{cases} \quad (۱)$$

که در آن سه حالت وجود دارد:

۱. ۰: برای حالتی که ژانر های فیلم در خوشه وجود ندارد.

۲. $p/q$: برای حالتی که تعدادی از ژانرها در خوشه وجود دارد. که در آن $p$ تعداد ژانرهایی خواهد بود که بین ژانرهای فیلم و خوشه مشترک می باشد و $q$ تعداد کل ژانر های فیلم مورد نظر است.

۳. ۱: برای حالتی که تمام ژانرهای فیلم در خوشه موجود می باشد.

درجه عضویت بدست آمده را در امتیاز کاربری هر فیلم ضرب کرده و به عنوان امتیاز نهایی هر فیلم در خوشه ثبت می نماییم. در طول زمان با ورود فیلم های جدید در سیستم آنها را نیز به همین صورت وارد مجموعه داده کرده و به صورت پویا خوشه بندی و امتیاز دهی می کنیم. اکنون با توجه به اینکه تمام فیلم های موجود در سایت به خوشه های مختلف

دسته بندی شدند زمان این رسیده است که مرحله دوم الگوریتم فراابتکاری خود را آغاز نماییم. در این مرحله از بهینه سازی، در هر خوشه از بین هر ژانر بهترین فیلم ها را انتخاب نموده و جهت نمایش به کاربر هدف آماده سازی می کنیم.

**پیدا کردن موارد مشابه**

در روش پیشنهادی ما، پیشنهاد نهایی محصول به کاربر هدف بر اساس علایق کاربر به ژانرهای مختلف فیلم انجام میگیرد. برای اطلاع از اینکه کاربر به چه ژانرهایی علاقه دارد میتوان به دو صورت اقدام کرد: (۱) در هنگام پر کردن فرم نظرخواهی و پرسش این سوال که به چه ژانرهایی علاقه دارد (۲) در هنگام استفاده کاربر از سیستم با توجه به مشاهده لینک ها و زمانی که صرف دیدن آنها می کند می شود فهمید که علاقه آن بیشتر نسبت به چه ژانرهایی خواهد بود، این امر باعث می شود تغییر علاقه کاربر در زمان های مختلف نیز پوشش داده شود و فاکتور زمان در سیستم های توصیه گر نادیده گرفته نشود.

برای پیدا کردن موارد مشابه با علاقه کاربر جهت نمایش به آن باید مراحل مختلفی را بررسی و نتیجه آنها را بدست بیاوریم که شامل مراحل زیر خواهد بود:

۱. در مرحله ۲ مجموعه ای از ژانرها را در خوشه ها قرار دادیم، که ژانر مورد علاقه کاربر هدف در آنها قرار دارد. در حال حاضر مجموعه داده ورودی که شامل تمام فیلم ها با ژانرهای مختلف می شود به تعداد محدودی از فیلم هایی که در هر خوشه قرار دارند کاهش پیدا کرده اند.
۲. ژانرهایی که کاربر هدف در فرم به آنها علاقه داشته و یا در لینک های آن گشت و گذار کرده و به آنها امتیاز داده است را شناسایی می کنیم.
۳. با توجه به ژانرهای مورد علاقه کاربر خوشه های مربوط به ژانرها را پیدا کرده و آنها را انتخاب می کنیم.
٤. در اینجا با استفاده از درجه شباهت *cosine*، شباهت بین فیلم های موجود در خوشه انتخاب شده را با فیلم های مورد علاقه کاربر اندازه گیری می کنیم.

$$\text{Sim (i,j)}=\frac{\sum_{t=1}^{n}(r_{t,i}-r_t)\times(r_{t,j}-\mathrm{r}_t)}{\sqrt{\sum_{t=1}^{n}(r_{t,i}-\mathrm{r}_t)^2}\times\sqrt{\sum_{t=1}^{n}(r_{t,j}-\mathrm{r}_t)^2}} \qquad (۲)$$

۵. در نتیجه شباهت بین فیلم مورد علاقه کاربر را با فیلم های موجود، در مجموعه زیر قرار می دهیم.

$$(۳)\ \mathrm{Sim}^1=\{\mathrm{sim(i,1)}, \ldots, \mathrm{sim(i,i\text{-}1)}, \mathrm{sim(i,i+1)}, \ldots, \mathrm{sim(i,n)}\}$$

٦. عناصر موجود در مجموعه بالا را به ترتیب نزولی مرتب می کنیم و با انتخاب یک مقدار آستانه برای آن، عناصری که از حد آستانه بیشتر هستند را انتخاب کرده و در مجموعه *Sim2* قرار می دهیم.

**ایجاد توصیه ها و نمایش آنها به کاربر**

در روش فیلترینگ اشتراکی ، ارزش رتبه کالاهای ناشناخته به عنوان مجموع امتیاز سایر موارد مشابه محاسبه می شود. در مرحله ۳ مشابه ترین فیلم ها را در *Sim2* قرار دادیم. گام بعدی استفاده از یک فرمول پیش بینی برای محاسبه امتیازات برای موارد ناشناخته است.

سارور و همکارانش [۴]دو تکنیک برای محاسبات پیش بینی، که مجموع وزن و رگرسیون هستند، را ارائه داده اند. پارامترهای روش رگرسیون برای تصمیم گیری سخت است و ممکن است تاثیر زیادی در نتیجه داشته باشد. در نتیجه، ما برای تعیین پیش بینی از مجموع وزن محاسبه شده استفاده کردیم. اول، ضریب $b$ را توسط فرمول (۴) به عنوان عامل نرمال کننده بدست می آوریم:

$$b=\frac{1}{\sum_{l\epsilon sim2}|sim(i,l)|} \quad (۴)$$

جایی که آیتم l به مجموعه شباهت sim۲ برای آیتم i مربوط می شود.

پس از آن، پیش بینی برای کاربر t بر روی آیتم i توسط (۵) محاسبه می شود:

$$p_{t,i} = r_i + b\sum_{l\in sim2} sim(i,l) \times (r_{t,l} - r_l) \quad (۵)$$

که در آن امتیازات مربوط به امتیاز متوسط آیتم های i و l به عنوان ri و rl تعریف می شوند.

پس از انتخاب همه آیتم های ناشناخته برای کاربر هدف M انتخاب توصیه می شود. با توجه به امتیازات محاسبه شده در بالا، می توان بردار امتیاز برای کاربر هدف را محاسبه نمود. پس از اتمام مراحل چهارگانه سیستم توصیه گر، اکنون نوبت به نمایش فیلم های پیشنهادی سیستم توصیه گر به کاربر هدف می باشد. بردار امتیاز بدست آمده در مرحله ۴ را که برای کاربر بصورت نزولی مرتب شده است را انتخاب کرده و M مورد اول را به او پیشنهاد داده و به کاربر نمایش می دهیم.

**یافته ها**

افزایش چشمگیر حجم اطلاعات منجر به ایجاد مشکلی بنام سرریز اطلاعات شده است - جایی که مردم برای پیدا کردن اطلاعات مورد نیاز خود از مجموعه ای از گزینه های گوناگون دچار مشکل می شوند. سیستم های توصیه گر به طور کاربردی از نظرات گروهی استفاده می کند تا به افرادی که در این گروه هستند کمک کند، به طور مؤثرتر محتوای مورد علاقه از یک فضای جستجوی بالقوه را شناسایی کنند. برخی از حوزه های کاربردی برای سیستم های توصیه گر عبارتند از: توصیه هایی برای CD های موسیقی و DVD ها، فیلم ها و کتاب ها.

فیلم ها به بخش مهمی از زندگی همه افراد تبدیل شده اند تا بتوانند نیازهای سرگرمی آنها را برآورده کنند. وب سایت های متعددی ابرداده های فیلم را میزبانی می کنند و کاربران را قادر به جستجو و پیدا کردن فیلم های مورد علاقه خود می کنند. کاربران اغلب به واسطه اینکه تعداد فیلم ها بصورت پویا در حال تغییر و افزایش است دچار سردرگمی می شوند. بنابراین، جستجوی اطلاعات مناسب که با علاقه مندی های کاربر مرتبط است دشوار خواهد بود. در نتیجه سیستمی مورد نیاز است که بتواند کاربران را به سمت صفحات و یا فیلم های مورد علاقه اش راهنمایی کند.

[۴] Sarwar et al

در این تحقیق یک سیستم توصیه گر نوین بر مبنای الگوریتم های فراابتکاری فازی در فروشگاه های اینترنتی معرفی شده است که دقت و سرعت توصیه را به خوبی افزایش می‌دهد. در این بخش قصد داریم با مقایسه روش پیشنهادی خود (که در بخش قبل به تفصیل بیان گردیده است) با روش های دیگر که قابلیت استفاده از الگوریتم فراابتکاری را ندارند بتوانیم به اثبات صحبت های خود بپردازیم.

## معیارهای سنجش کیفیت روش پیشنهادی

هنگامی که صحبت از ارزیابی الگوریتم می شود این سوال در ذهن مطرح می شود که دقیقا به دنبال ارزیابی چه چیزی هستیم؟ در اکثر اوقات هنگامی که از ارزیابی الگوریتم های سیستم توصیه گر صحبت می کنیم منظور ارزیابی دقت توصیه[۵]، معیار پوشش[۶] و یا سرعت الگوریتم است.

### سرعت پیش بینی:

همانگونه که قبلا گفته شده است برای ارائه پیشنهاد کالا به کاربر باید یک سری عملیات انجام شود (اعم از جمع آوری اطلاعات، خوشه بندی، پیدا کردن تشابه و ...) که انجام آنها مستلزم صرف زمان خواهد بود. در سیستم های توصیه گر هر چه این زمان کمتر باشد بنابراین سرعت سیستم بالاتر خواهد بود. پس سرعت پیش بینی تنها به زمان صرف شده برای انجام توصیه بستگی دارد.

### دقت توصیه:

دقت یا درستی توصیه به معنی تعداد صحیح توصیه است. در هر سیستم توصیه جواب ها در یکی از چهار دسته زیر قرار می گیرند:

۱. جواب صحیح مثبت(TP[۷]): رکوردهایی در این دسته قرار می گیرند که در دسته مثبت می باشند و توصیه گر نیز به درستی آنها را توصیه کرده است.
۲. جواب صحیح منفی(TN[۸]): رکوردهایی در این دسته قرار می گیرند که در دسته منفی می باشند و توصیه گر نیز به درستی آنها را توصیه نکرده است.

[۵] Precision
[۶] Coverage
[۷] True Positive
[۸] True Negative

۳. جواب غلط مثبت(FP[۹]): رکوردهایی در این دسته قرار می گیرند که در دسته منفی می باشند و توصیه گر به اشتباه آنها را توصیه کرده است.

٤. جواب غلط منفی(FN[۱۰]): رکوردهایی در این دسته قرار می گیرند که در دسته مثبت می باشند و توصیه گر به اشتباه آنها را توصیه نکرده است.

دقت توصیه نشان دهنده این است که چه تعداد از پیشنهادهای ارائه شده به کاربر واقعا مورد علاقه او بوده و سیستم به درستی آنها را به کاربر پیشنهاد داده است که از طریق فرمول زیر قابل محاسبه است:

$$\text{Precision:}\frac{TP}{TP+FP} \quad (۶)$$

که همان طور که گفته شد در آن TP: تعداد توصیه هایی را نشان می دهد که مورد علاقه کاربر است و سیستم نیز آنرا به درستی توصیه کرده است و FP: تعداد توصیه هایی را نشان می دهد که مورد علاقه کاربر نیست اما سیستم آنها را به اشتباه پیشنهاد داده است.

**معیار پوشش:**

این معیار بیانگر این است که سیستم تا چه حد توانایی کشف فیلم های مورد علاقه کاربر را دارد و برابر با نسبت توصیه های مربوطه به تمام فیلم هایی است که باید توصیه شود. که برای محاسبه آن از رابطه زیر استفاده می شود:

$$\text{Coverage:}\frac{\text{تعداد توصیه انجام شده}}{\text{تعداد کل فیلم های مورد توصیه}}$$

**محیط و داده های استفاده شده جهت شبیه سازی**

برای ارزیابی روش پیشنهادی خود داده های مورد نیاز را از پایگاه داده فیلم اینترنتی (IMDB) گردآوری نموده ایم. مجموعه داده در هر سطر ویژگی های هر فیلم را بصورت جداگانه در ستون های مختلف، مانند ژانرها در genres، ارائه می دهد. ما از اطلاعات موجود در ستون های زیر استفاده کرده ایم: movie_title، genres، plot_keywords، imdb_score، title_year و language. ما فیلم های با عناوین انگلیسی را که بین سال های ۲۰۰۶ و ۲۰۱۰، در یک دوره پنج ساله بودند را انتخاب کرده ایم که مجموعا ۴۸۴۸۳ عدد فیلم را با ۲۷ ژانر و ۱۹۵۶۱ کلمه کلیدی به نمایش می گذارد.

[۹] False Positive
[۱۰] False Negative

همانند سایر الگوریتم های یادگیری ماشین[11] در اینجا نیز داده ها به دو دسته آموزش و تست تقسیم می شوند و ارزیابی بر روی داده های تست صورت می گیرد. این شبیه سازی در محیط برنامه Visual Studio ۲۰۱۲ با زبان C# انجام گرفته است.

## ارزیابی نتایج

هدف اصلی روش پیشنهادی ما بهبود دقت و کاهش زمان پاسخ سیستم در کنار پویایی الگوریتم سیستم توصیه گر نسبت به سیستم های دیگر است. در اینجا ما برای بررسی و ارزیابی عملکرد روش پیشنهادی خود آنرا با دو سیستم توصیه گر دیگر که یکی از آنها همین روش بدون استفاده از الگوریتم فراابتکاری و دیگری بدون استفاده از روش خوشه بندی است را مقایسه و نتایج را بررسی می نماییم.

دقت روش پیشنهادی

در این بخش، بررسی مقایسه ای میزان دقت بین روش پیشنهادی و روش های گفته شده بررسی می گردد. به این گونه که تعداد فیلم ها از بین N=۱۰۰ تا N=۹۰۰ متغیر است و از بین این تعداد M=۱۰ فیلم را به کاربر مورد نظر پیشنهاد می دهیم.

**جدول ۲: مقایسه دقت روش پیشنهادی و دو روش دیگر**

| تعداد فیلم های قابل توصیه | معیار دقت | | |
|---|---|---|---|
| | روش پیشنهادی | بدون الگوریتم فراابتکاری | بدون الگوریتم خوشه بندی |
| ۱۰۰ | ٪۹۹ | ٪۹۶ | ٪۹۲ |
| ۲۰۰ | ٪۹۹ | ٪۹۵ | ٪۸۸ |
| ۳۰۰ | ٪۹۸ | ٪۹۶ | ٪۸۶ |
| ۴۰۰ | ٪۹۸ | ٪۹۴ | ٪۸۰ |
| ۵۰۰ | ٪۹۶ | ٪۹۴ | ٪۷۵ |
| ۶۰۰ | ٪۹۵ | ٪۸۹ | ٪۷۲ |
| ۷۰۰ | ٪۹۳ | ٪۸۵ | ٪۶۹ |
| ۸۰۰ | ٪۹۰ | ٪۸۱ | ٪۶۱ |

[11] Machine learning

| | | | |
|---|---|---|---|
| ۹۰۰ | ٪۸۸ | ٪۷۷ | ٪۵۶ |


*شکل ۴: مقایسه دقت روش پیشنهادی و دو روش دیگر*

همانطور که در جدول ۲ می بینید از نتایج به خوبی مشخص است، زمانی که در الگوریتم خود از روش های فراابتکاری و یا خوشه بندی استفاده نمی نماییم با افزایش تعداد فیلم ها به تدریج با کاهش دقت روبرو خواهیم شد. اما به خوبی مشخص است که این کاهش در روش پیشنهادی بسیار اندک بوده است. همانطور که نتایج نشان می دهد دقت روش پیشنهادی در تمام موارد نسبت روش های دیگر بالاتر بوده است که این بیانگر کیفیت مناسب روش پیشنهادی این پژوهش می باشد. نمودار شکل ۴ این مقایسه را به خوبی نشان می دهد.

معیار پوشش

در این بخش، بررسی مقایسه ای میزان پوشش بین روش پیشنهادی و روش های گفته شده بررسی می گردد.

**جدول ۳: مقایسه میزان پوشش روش پیشنهادی و دو روش دیگر**

| تعداد فیلم های قابل توصیه | معیار دقت | | |
|---|---|---|---|
| | روش پیشنهادی | بدون الگوریتم فراابتکاری | بدون الگوریتم خوشه بندی |
| ۱۰۰ | ٪۸۳ | ٪۷۹ | ٪۶۲ |
| ۲۰۰ | ٪۸۵ | ٪۸۰ | ٪۶۷ |
| ۳۰۰ | ٪۸۰ | ٪۷۷ | ٪۵۸ |

| | | | |
|---|---|---|---|
| ۴۰۰ | ٪۷۹ | ٪۷۳ | ٪۶۴ |
| ۵۰۰ | ٪۸۵ | ٪۸۲ | ٪۶۹ |
| ۶۰۰ | ٪۸۲ | ٪۸۰ | ٪۶۰ |
| ۷۰۰ | ٪۷۸ | ٪۷۰ | ٪۷۲ |
| ۸۰۰ | ٪۸۱ | ٪۷۹ | ٪۶۸ |
| ۹۰۰ | ٪۸۵ | ٪۸۲ | ٪۵۶ |


*شکل ۵. مقایسه پوشش روش پیشنهادی و دو روش دیگر*

همانطور که در جدول ۳ می بینید از نتایج به خوبی مشخص است بر خلاف معیار دقت که در الگوریتم ها با افزایش تعداد فیلم ها به تدریج کاهش دقت را داشته اند، در این معیار با افزایش تعداد فیلم ها پوشش به صورت خطی کاهش پیدا نمی کند، زیرا با افزایش تعداد فیلم های قابل توصیه تعداد فیلم های با کیفیت توصیه شده نیز افزایش پیدا کرده و این میزان پوشش تقریبا در یک محدوده مشخص جای می گیرد. همان گونه که نتایج نشان می دهد میزان به دست آمده برای معیار پوشش در روش پیشنهادی در تمام موارد نسبت به دو روش دیگر بالاتر بوده است که این بیانگر کیفیت مناسب روش پیشنهادی این پژوهش می باشد. نمودار موجود در شکل ۵ این مقایسه را به خوبی نشان داده است.

سرعت پیش بینی

در اینجا به منظور بررسی بهبود زمان، زمان صرف شده برای توصیه ۱۰ فیلم به کاربر را با دو حالت دیگر را بررسی می کنیم:

**جدول ۴. مقایسه زمان روش پیشنهادی با دو روش دیگر**

| | |
|---|---|
| در حالت روش پینهادی(میلی ثانیه) | ۲۳ |
| در حالت بدون الگوریتم فراابتکاری(میلی ثانیه) | ۱۰۵ |
| در حالت بدون الگوریتم خوشه بندی(میلی ثانیه) | ۱۶۲ |

همانگونه که از جدول ۴ مشخص است به کمک روش پیشنهادی توانسته ایم زمان متوسط توصیه برای کاربر هدف را تا حد زیادی کاهش دهیم که این امر به دلیل همان کاهش ابعاد صورت گرفته است. در اینجا با توجه به اینکه دیگر نیازی نیست که سیستم برای یافتن فیلم مورد نظر تمام فیلم های موجود در سایت را بررسی نموده و پاسخ کاربر را دهد، با کاهش ابعاد صورت گرفته می تواند تنها با انجام تعداد کمی مقایسه به جواب بهینه خود رسیده و آنرا به کاربر پیشنهاد دهد. که این امر باعث شده است تا سیستم بتواند با سرعتی بالاتر به کاربر جواب دهد.

ما در این بخش سیستم توصیه گر پیشنهادی خود را که در بخش قبل بصورت کامل توضیح داده شده است را با دو حالت سیستم توصیه گر بدون الگوریتم فراابتکاری و سیستم توصیه گری که الگوریتم فراابتکاری را بدون خوشه بندی اعمال می کند مقایسه نمودیم. نتایج بدست آمده از بررسی ها نشان می دهد که استفاده از الگوریتم بهینه سازی فراابتکاری مذکور توانسته با کاهش ابعاد داده های موجود در سیستم زمان صرف شده برای توصیه فیلم به کاربر را کاهش دهد و همچنین با توجه به خوشه بندی سلسله مراتبی فازی که انجام شد دقت توصیه نیز بهبود یافته و توانسته نتایج مطلوبتری را در زمان کوتاهتری به کاربر ارائه دهد. نتایج حاصله بیانگر کیفیت بهتر و مناسبتر روش پیشنهادی ما نسبت به روش های پیشین خواهد بود.

## نتیجه گیری

در این پژوهش ما قصد داشتیم یک سیستم جدید بر مبنای الگوریتم های فراابتکاری فازی در فروشگاه های اینترنتی فیلم معرفی کنیم که دقت و سرعت توصیه را به خوبی افزایش دهد. همانگونه که توضیح داده شده است این اتفاق باید در زمان فیلترسازی انجام پذیرد. ما با استفاده از الگوریتم های بهینه سازی فراابتکاری توانستیم داده های خود را بصورت استاندارد بهینه کرده و کاهش ابعاد دهیم. یک الگوریتم بهینه سازی فراابتکاری یک روش ابتکاری است که میتواند با تغیرهایی کم برای مسائل مختلف بهینه سازی به کار رود. الگوریتم های فراابتکاری بطور قابل ملاحظه ای توانایی یافتن جوابهای با کیفیت بالا را برای مسائل بهینه سازی سخت افزایش میدهد. در طرح پیشنهادی ما از الگوریتم بهینه سازی کلونی زنبور عسل استفاده شد که بصورت جداگانه در دو مرحله اجرا می شود.

مجموعه داده بکار رفته در روش ما، مجموعه داده فیلم است که با توجه به اینکه هر فیلم دارای مشخصه های متمایزی است لذا خوشه بندی این فیلم ها به ما برای رسیدن سریعتر به جواب کمک بسیاری خواهد کرد. البته بخاطر اینکه هر فیلم امکان داشتن معیارهای مشابه با چندین خوشه را دارد بنابراین خوشه بندی آنها بصورت کلاسیک بسیار سخت است. برای حل این مشکل از الگوریتم خوشه بندی سلسله مراتبی فازی استفاده می نماییم. در این حالت، با وزن‌های مختلف، فیلم ها به خوشه‌های مختلف تخصیص می‌یابد. علاوه بر این، ما بجز معیار شباهت و امتیازدهی، از فاکتور زمان نیز به عنوان پارامتر مهمی برای شناسایی حوزه های مورد علاقه کاربر و اندازه گیری تشابه کاربران برای خوشه بندی استفاده کرده ایم. از این رو به نظر می‌رسد نوآوری‌های بکار رفته در این پژوهش ، جهت استفاده از خوشه‌بندی فراابتکاری فازی قادر به افزایش دقت توصیه‌های ارائه شده، می‌باشد. نتایج پیاده سازی روش پیشنهادی و اعمال آن بر روی مجموعه داده های واقعی بیانگر کیفیت مناسب روش پیشنهادی در مقایسه با روش پایه و سایر مطالعات است.

بنابراین، با توجه به اینکه در مرحله استفاده از االگوریتم بهینه سازی فراابتکاری توانسته ایم فیلم های با امتیاز بالا را جهت پیشنهاد به کاربر جدا نموده و آماده توصیه قرار دهیم لذا این امر باعث می شود که سیستم خطای توصیه فیلم های کم امتیاز را انجام ندهد و بصورت کلی با نتایجی که در بخش قبل بدست آورده و بررسی نمودیم میبینیم که روش پیشنهادی ما توانسته دقت و سرعت توصیه فیلم به کاربر را بر اساس علایق آن بهبود داده و کیفیت مطلوبی را از خود نشان دهد.

## منابع


۱. Sneha Y.S, G. Mahadevan, Madhura, A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web", International Journal of Computer Theory and Engineering (IJCTE) Vol. ٤, No. ٢, April ٢٠١٢

۲. Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang and J. Lu, "A hybrid fuzzy-based personalized recommender system for telecom products/services", Inf. Sci., vol. ٢٣٥, pp. ١١٧-١٢٩, ٢٠١٣.

۳. T. Zhu,Y. Ren, W. Zhou, An effective privacy preserving algorithm for neighborhood-based collaborative filtering, Future Generation Computer SystemsVolume ٣٦, Pages ١٤٢–١٥٥, July ٢٠١٤.

۴. M. C. Pham, Y. Cao, R. Klamma, M. Jarke, A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis, Journal of Universal Computer Science, vol. ١٧, ٥٨٣-٦٠٤, no. ٤ ٢٠١١.

۵. D. K. Bokde, S Girase, An Item-Based Collaborative Filtering using Dimensionality Reduction Techniques on Mahout Framework, Information Retrieval (cs.IR), ٢٠١٥.

۶. Q. Li, M. Yao, J. Yang, N. Xu, Genetic Algorithm and Graph Theory Based Matrix Factorization Method for Online Friend Recommendation, Hindawi Publishing Corporation, Scientific World Journal, Volume ٢٠١٤, Article ID ١٦٢١٤٨, ٥ pages, ٢٠١٤.

۷. Adomavicius, G., Zhang, J.: Iterative smoothing technique for improving stability of recommender systems. In: Proceedings of the Workshop on Recommendation Utility Evaluation: Beyond RMSE. CEUR Workshop Proceedings, vol. ٩١٠, pp. ٣–٨ ٢٠١٤.

۸. Candillier, L., Chevalier, M., Dudognon, D., Mothe, J.: Diversity in recommender systems: bridging the gap between users and systems. In: Proceedings of the International Conference on

Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services, pp. ٤٨–٥٣, ٢٠١٣.

٩. L. Peska, P. Vojtas, Biased k-NN Similarity Content Based Prediction of Movie Tweets Popularity, Semantic Scholar Home, pp. ١٠١–١١٠, ٢٠١٥.

١٠. F. Akbari, A. H. Tajfar, A. Farhoodi Nejad Graph-based Friend Recommendation in Social Networks using Artificial Bee Colony, ٢٠١٣ IEEE ١١th International Conference on Dependable, Autonomic and Secure Computing, ٢٠١٣.

١١. N. Lian-ju, D. Hai-yan, an algorithm for friend-recommendation of social networking sites based on SimRank and ant colony optimization, The Journal of China Universities of Posts and Telecommunications, ٢٠١٤.

١٢. Z. Rasheed and M. Shah, "Movie genre classification by exploiting audio-visual features of previews", Proc. the ١٦th International Conference on Pattern Recognition vol.٢, no., pp.١٠٨٦-١٠٨٩ □□□.٢, ٢٠٠٢.

١٣. Z. Rasheed, Y. Sheikh, and M. Shah, "On the use of computable features for film classification," IEEE Transactions on Circuits and Systems for Video Technology, vol.١٥, no.١, pp. ٥٢- ٦٤, Jan. ٢٠٠٥.

١٤. H. Zhou, T. Hermans, A. V. Karandikar, J. M. Rehg, "Movie Genre Classification via Scene Categorization", Proc. ١٠th international conference on Multimedia, pp. ٧٤٧-٧٥٠, ٢٠١٠.

١٥. Austin, E. Moore, U. Gupta, and P. Chordia, "Characterization of movie genre based on music score," IEEE International Conference on Acoustics Speech and Signal Processing, pp.٤٢١-٤٢٤, ٢٠١٠.

١٦. P. Resnick and H. Varian, “Recommender systems,” Communications of the ACM, vol. ٤٠, no. ٣, pp. ٥٦–٥٨, ١٩٩٧.

١٧. C. Haruechaiyasak, C. Tipnoe, S. Kongyoung,C. Damrongrat, and N. Angkawattanawit, “A Dynamic Framework for Maintaining Customer Profiles in E-Commerce Recommender Systems”,Proc. Int’l Conf. e-Technology, e-Commerce and e-Service (EEE'٠٥), pp. ٧٦٨-٧٧١, ٢٠٠٥.

١٨. B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, ”Item-based collaborative filtering recommendation algorithms,” Proceedings of the ١٠th International WWW Conference, pp. ٢٨٥-٢٩٥, ٢٠٠١.

١٩. Y. -H. Chien and E.I. George, ”A Bayesian model for collaborative filtering,” Proceedings of the statistics, ١٩٩٩.