

## بهبود سیستم‌های توصیه گر با استفاده از الگوریتم بهینه سازی رقابت استعماری در تخمین رتبه‌های کاربران

الهام کاظمی<sup>۱</sup>، الهام صادقی<sup>۲</sup>

<sup>۱</sup> هیئت علمی دانشگاه آزاد اسلامی واحد آبادان

<sup>۲</sup> کارشناسی ارشد مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد آبادان

### چکیده

با توجه به گسترش روزافزون استفاده از شبکه‌های اجتماعی و گرایش افراد در رده‌های مختلف به سمت این شبکه‌ها، اطلاعات و داده‌های موجود در آن‌ها نیز افزایش یافته است، در سایت‌های شبکه‌های اجتماعی نیاز به پیشنهاد نه تنها برای محصولات و خدمات بلکه برای اشخاص و گروه‌ها هم وجود دارد که به این ترتیب کاربر می‌تواند به شخص یا مد حرفه‌ای به‌طور بالقوه لینک داده شود. به همین منظور، سیستم‌های توصیه گر در میان حجم عظیمی از انتخاب‌های ممکن برای رسیدن به گزینه‌های مفید و موردعلاقه کاربران، به‌عنوان سیستم‌های تأثیرگذار به‌منظور راهنمایی و هدایت افراد به کار گرفته شدند. درواقع هدف اصلی این سیستم‌ها، شناسایی علایق کاربران و فیلترکردن داده‌های موردعلاقه کاربران از میان حجم انبوهی از داده‌ها است. سیستم‌های توصیه گر سنتی را می‌توان به سه دسته کلی دسته‌بندی کرد، که عبارتند از: مبتنی بر محتوا، پالایش گروهی و ترکیبی. چالش‌های بسیاری در سیستم‌های توصیه گر سنتی وجود دارند که این چالش‌ها عبارتند از: تنگی داده‌ها (پراکندگی)، شروع سرد، حمله افراد بدخواه به سیستم، مقیاس پذیری، کاربر غیر معمول، هم معنایی، عدم شفافیت، عدم تعمیم‌پذیری همسایگی. در این پژوهش یک روش جدید برای بهبود سیستم‌های توصیه گر با استفاده از الگوریتم رقابت استعماری ارائه می‌شود. در این روش پیشنهادی، یک الگوریتم جدید خوشه‌بندی برای دسته‌بندی کاربران و یا آیتم‌ها در سیستم‌های توصیه گر ارائه می‌شود. با توجه به تحقیقات بسیاری که در مورد سیستم‌های توصیه گر صورت گرفته است. هنوز هم چالش‌های بسیاری به‌منظور بررسی و رفع شدن در پیش رو هستند، که می‌توان با ارائه روش‌های نوین تر علاوه بر رفع چالش‌های موجود، میزان کارایی را در چنین سیستم‌هایی افزایش داد. این روش‌ها عبارتند از: خوشه‌بندی گراف کاربران، داده‌کاوی اطلاعات و استخراج قوانین، فازی سازی روابط اعتماد.

**واژه‌های کلیدی:** سیستم‌های توصیه گر، سیستم‌های مبتنی بر محتوا، حمله افراد بدخواه به سیستم، مقیاس پذیری.

**۱- مقدمه**

با توجه به گسترش روزافزون استفاده از شبکه‌های اجتماعی و گرایش افراد در رده‌های مختلف به سمت این شبکه‌ها، اطلاعات و داده‌های موجود در آن‌ها نیز افزایش یافته است، در سایت‌های شبکه‌های اجتماعی نیاز به پیشنهاد نه‌تنها برای محصولات و خدمات بلکه برای اشخاص و گروه‌ها هم وجود دارد که به این ترتیب کاربر می‌تواند به شخص یا مد حرفه‌ای به‌طور بالقوه لینک داده شود. در واقع، با توسعه تجارت الکترونیک<sup>۱</sup> وب‌سایت‌ها، کاربران به سختی مناسب‌ترین آیتم‌ها را از میان مجموعه بزرگی از آیتم‌های موجود (محصولات و خدمات) که وب‌سایت‌ها در اختیار آنها قرار می‌دهند، پیدا می‌کنند. در نتیجه، رشد انفجاری و تنوع اطلاعات موجود بر روی وب، کاربران را در تصمیم‌گیری‌های خود ضعیف کرده است و در دسترس بودن انتخاب‌ها، به جای سودمند بودن، رفاه کاربران را کاهش داده است [۲].

مبحث اصلی در روش مبتنی بر محتوا استفاده از ویژگی‌های یک آیتم برای پیش‌بینی علاقه کاربر است. به‌عنوان مثال، برای توصیه یک کتاب، می‌توان ویژگی‌هایی مانند، نام نویسنده، سبک، کلمات کلیدی و غیره را در نظر گرفت. در روش پالایش گروهی می‌توان از اطلاعات کسب شده از گذشته کاربران در سیستم استفاده کرد. بنابراین به کمک این اطلاعات عمل پیش‌بینی آیتم برای کاربر فعال<sup>۲</sup> انجام خواهد گرفت. فرض اساسی این روش‌ها به این صورت است که می‌توان نظرات کاربران دیگر را انتخاب و جمع‌آوری کرد و یک پیش‌بینی منطقی را از ترجیحات کاربر فعال بدست آورد. به عبارت دیگر، اگر کاربران در مورد کیفیت و یا ارتباط برخی از موارد با یکدیگر تشابه داشته باشند، بنابراین به احتمال زیاد، آنها در مورد موارد دیگر نیز شباهت‌هایی را خواهند داشت [۸]. در روش ترکیبی نیز از ترکیب سایر روش‌ها به‌منظور بهبود عملکرد سیستم و رفع نواقص هر روش موجود می‌توان کمک گرفت [۹].

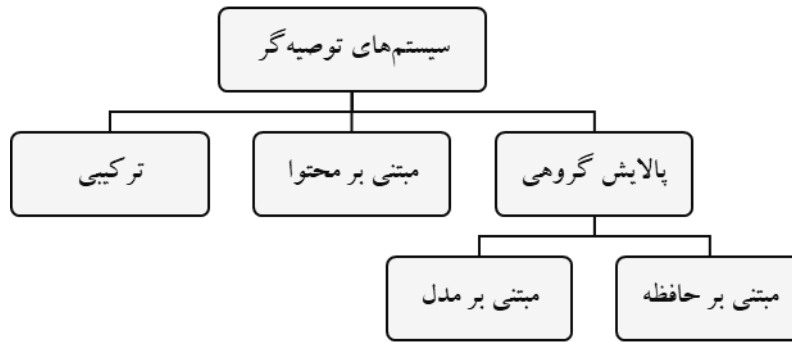
در این مقاله یک روش جدید برای بهبود سیستم‌های توصیه‌گر با استفاده از الگوریتم رقابت استعماری ارائه می‌شود. در این روش پیشنهادی، یک الگوریتم جدید خوشه‌بندی برای دسته‌بندی کاربران و یا آیتم‌ها در سیستم‌های توصیه‌گر ارائه می‌شود. برای این منظور، ابتدا کاربران و یا آیتم‌ها به صورت یک گراف مدل می‌شوند که کاربران و یا آیتم‌ها، رئوس گراف را تشکیل داده و یال وزندار بین دو راس بیانگر میزان شباهت بین آنها می‌باشد. در مرحله‌ی بعد، مراکز اولیه‌ی خوشه‌ها با استفاده از الگوریتم رقابت استعماری بدست می‌آیند. این مراکز اولیه با استفاده از یک مکانیزم ارائه شده با مراکز بهینه‌ی دیگری جایگزین می‌شوند. در نهایت خوشه‌هایی که دارای تعداد کمی عضو می‌باشند با سایر خوشه‌ها ادغام می‌شوند و خوشه‌های نهایی را تشکیل می‌دهند. در نهایت برای انتخاب همسایگی کاربر هدف، از کاربران موجود در خوشه‌ی مربوط به کاربر هدف استفاده می‌شود. در نهایت بر اساس خوشه‌های به دست آمده به کاربران آیتم‌های مورد نظر پیشنهاد می‌شود.

**مبانی نظری****دسته‌بندی سیستم‌های توصیه‌گر**

سیستم‌های توصیه‌گر سنتی را می‌توان با توجه به آنچه در شکل ۱ نشان داده شده است، به صورت زیر دسته‌بندی نمود، که در ادامه توضیحات بیشتر را در مورد هر یک ارائه خواهیم کرد.

<sup>1</sup> E-commerce

<sup>2</sup> Target user (Active user)



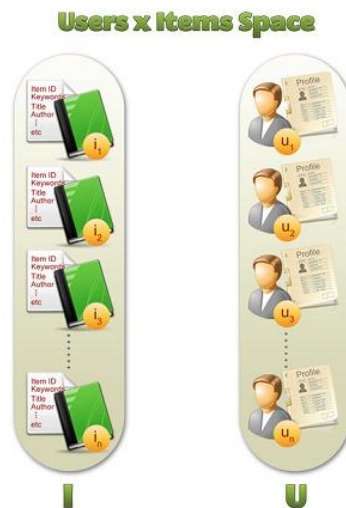
شکل ۱: ساختار سلسله مراتبی از سیستم‌های توصیه‌گر [۱۷]

به‌طور کلی در سیستم‌های توصیه‌گر، محیط توصیه را می‌توان به صورت زیر نشان داد:

اجازه دهید  $U$  مجموعه‌ای از تمام کاربران و  $I$  مجموعه‌ای از تمام آیتم‌هایی باشد که می‌توانند توصیه شوند. فضاهای  $I$  و  $U$  ممکن است در برخی از موارد بسیار بزرگ (بیش از یک میلیون) باشد.  $u$  کاربری است که با عنوان کاربر فعال نیازمند توصیه بوده ( $u \in U$ ) و  $i$  مجموعه‌ای از آیتم‌هایی است که قصد داریم برای کاربر  $u$  پیش‌بینی کنیم ( $i \in I$ ). همچنین  $f$  تابع سودمندی است و اهمیت آیتم  $i$  را برای کاربر  $u$  اندازه‌گیری می‌کند ( $f: U \times I \Rightarrow R$ )، و در آن  $R$  یک مجموعه از مقادیر نامنفی مرتب در یک محدوده خاص می‌باشد. در واقع ما نیاز به انتخاب آیتم  $i' \in I$  داریم که برای هر کاربر  $u$  این تابع سودمندی را به حداکثر مقدار خود برساند. این مسئله را می‌توان براساس رابطه زیر نشان داد:

$$\forall u \in U, i'_u = \operatorname{argmax}_{i \in I} f(u, i) \quad (1)$$

راه‌های مختلفی برای محاسبه تابع سودمندی وجود دارد. این تابع می‌تواند توسط کاربر تعریف شده و یا توسط یک نرم افزار محاسبه گردد. رتبه کاربر یک سه تایی  $(u, I, r)$  از آن است که در آن  $r$  ارزش تعیین شده توسط کاربر  $u$ ، به یک آیتم خاص  $i$  است. معمولاً این مقدار زیر مجموعه‌ای ثابت از اعداد حقیقی و یا یک متغیر باینری است. در فضای  $U$ ، هر کاربر توسط مشخصاتی که شامل ویژگی‌های مختلف مانند شناسه کاربر، سن، جنسیت، درآمد، و غیره است نشان داده شده است. یک مشخصه ساده تنها می‌تواند حاوی ID کاربر باشد. همچنین هر آیتم در فضای آیتم‌های  $I$  توسط مجموعه‌ای از ویژگی‌ها، نشان داده شده است [56]. برای مثال در هنگام توصیه کتاب‌ها، هر کتاب را می‌توان با شناسه، عنوان، نویسنده، و غیره نشان داد. شکل ۲ کاربران و آیتم‌ها را در فضای  $U \times I$  نشان می‌دهد.

شکل ۲: کاربران و آیتم‌ها در فضای  $U \times I$  [۱۰]

مشکل سیستم‌های توصیه‌گر این است که تابع سودمندی  $f$  تنها در بخشی از فضای  $U \times I$  تعریف شده است [۱۹].

### چالش‌های موجود در سیستم‌های توصیه‌گر

چالش‌های بسیاری در سیستم‌های توصیه‌گر سنتی وجود دارند که در مقالات متفاوت به آن‌ها اشاره شده است. در ادامه برخی از چالش‌ها موجود در این سیستم‌ها شرح داده خواهد شد:

تنکی داده‌ها (پراکندگی)

تعداد کاربران و آیتم‌ها در سیستم‌های توصیه‌گر تجاری بسیار زیاد است. حتی کاربرانی که بسیار فعال می‌باشند، در رتبه‌بندی‌های خود تنها تعداد کمی از کل آیتم‌های موجود در یک پایگاه داده را رتبه‌بندی می‌کنند. همچنین آیتم‌های بسیار محبوب توسط تعداد کمی از کل کاربران موجود در پایگاه داده رتبه‌بندی می‌شوند. این مشکل، معمولاً به‌عنوان تنکی داده‌ها شناخته می‌شود و تاثیر منفی عمده‌ای را بر اثربخشی سیستم‌های پالایش گروهی دارد. چرا که در هنگام تنکی داده‌ها، این امکان وجود دارد که نتوان شباهتی را بین دو کاربر تعریف کرد، و در نتیجه ارائه روش پالایش گروهی بدون تاثیر خواهد بود. حتی زمانی که ارزیابی تشابه امکان پذیر باشد، به دلیل اطلاعات ناکافی این ارزیابی تشابه را نمی‌توان قابل اعتماد دانست [۲۲].

شروع سرد<sup>۳</sup>

شروع سرد بر اهمیت مسئله تنکی داده‌ها تأکید دارد. این مشکل برای کاربران و آیتم‌هایی که به تازگی به سیستم اضافه شده‌اند، وجود می‌آید. بنابراین یک کاربر جدید، تنها با رتبه‌دهی به یک تعداد کافی از آیتم‌ها، قبل از بکارگیری سیستم توصیه‌گر، قادر به دریافت توصیه‌های قابل اعتماد و دقیق خواهد بود [۲۳].

<sup>3</sup> Cold start

حمله افراد بدخواه به سیستم<sup>۴</sup>

در سیستم‌های پالایش گروهی، حمله افراد بدخواه به سیستم، به‌عنوان یک چالش اساسی مطرح می‌گردد. چرا که در چنین سیستم‌هایی، افراد می‌توانند با ایجاد مشخصات جعلی از کاربر فعال، توصیه‌ها را تحت تأثیر قرار داده و نتایج را به نفع خود تغییر دهند [۲۴].

مقیاس پذیری<sup>۵</sup>

سیستم‌های مبتنی بر معیار شباهت، پیچیدگی محاسباتی بالایی دارند. از سوی دیگر، در روش‌های پالایش گروهی مبتنی بر حافظه به‌منظور پیش‌بینی سلايق کاربر، باید کل مجموعه داده‌ای را مورد بررسی قرار داد. بنابراین، با افزایش اطلاعات ثبت شده در مجموعه داده‌ها، پیچیدگی محاسباتی به شکلی نمایی افزایش می‌یابد که منجر به کاهش کارایی سیستم می‌گردد [۲۵].

کاربر غیر معمول<sup>۶</sup>

همانطور که گفته شد وظیفه مهم یک سیستم توصیه‌گر بهبود کیفیت توصیه‌ها برای کاربر فعال است. اگر یک کاربر اعتماد خود را به یک سیستم توصیه‌گر از دست دهد، به گونه‌ای که سیستم قادر به ارائه توصیه‌های مناسب به او نباشد، استفاده مجدد او از این سیستم بعید خواهد بود. ممکن است در سیستم‌های توصیه‌گر افرادی وجود داشته باشند که دارای سلايق غیر معمولی در مقایسه با سایر اعضای جامعه باشند. بنابراین سیستم‌های توصیه‌گر پالایش گروهی، به چنین افرادی توصیه‌هایی با کیفیت پایین ارائه خواهند کرد. برخی از نویسندگان از واژه Grey sheep برای تشخیص این افراد از کاربران دیگر استفاده می‌کنند [۲۶].

هم‌معنایی<sup>۷</sup>

در زندگی واقعی، نام محصولات مختلف می‌تواند به اشیاء مشابهی اشاره کند. سیستم‌های توصیه‌گر قادر به پیدا کردن این ارتباط پنهان نیستند. به عبارت دیگر می‌توان گفت، اگر دو محصول مشابه اسمی متفاوتی داشته باشند، سیستم توصیه‌گر موجود هیچ هم‌معنایی و ارتباطی را در بین این دو محصول پیدا نخواهند کرد [۲۷].

عدم شفافیت<sup>۸</sup>

یکی از کاستی‌های عمده در سیستم‌های توصیه‌گر، عدم شفافیت آنها می‌باشد، که در آن کاربران می‌خواهند این موضوع را درک کنند که چرا به آنها تنها آیت‌هایی خاص پیشنهاد می‌شود. در این راستا، در [۲۸] مکانیسمی به‌منظور توضیح درباره منابع ارائه شده، که کاربران موجود در آن می‌توانند استدلال پشت فرایند توصیه را درک کنند.

عدم تعمیم‌پذیری همسایگی<sup>۹</sup>

یکی دیگر از مسائل موجود در سیستم‌های توصیه‌گر کشف روابط پنهان (رابطه تعدی) در بین کاربران است، به این صورت که اگر کاربر یک مشابه کاربر دو باشد و سپس کاربر دو نیز مشابه با کاربر سه باشد، در نتیجه این امکان وجود خواهد داشت که کاربر یک نیز مشابه با کاربر سه باشد، که در سیستم‌های توصیه‌گر چنین ارتباطی مورد توجه قرار نمی‌گیرد.

<sup>4</sup> Attack

<sup>5</sup> Scalability

<sup>6</sup> Unusual user (Grey sheep)

<sup>7</sup> Synonymy

<sup>8</sup> Lack of Transparency

<sup>9</sup> Lack of generalizability of the neighborhood

### سیستم توصیه گر مبتنی بر خوشه بندی

یکی از مهم‌ترین مشکلات و چالش‌های موجود در سیستم‌های توصیه‌گر، انتخاب همسایگی برای کاربر هدف به منظور پیش‌بینی رتبه‌های این کاربر می‌باشد. سیستم‌های توصیه‌گر پالایش گروهی که بر اساس کاربران همسایه‌ی کاربر هدف، رتبه‌ها را پیش‌بینی می‌کند، باید قادر باشند که کاربران همسایه‌ی کاربر هدف را به خوبی شناسایی نمایند. در این گونه سیستم‌ها، با انتخاب درست همسایگی کاربران، می‌توان برای کاربر هدف رتبه‌هایی با دقت بالا را پیش‌بینی نمود. انتخاب و تعیین کاربران همسایه در سیستم‌های توصیه‌گر با چالش‌هایی روبه‌رو می‌باشد. یکی از مشکلاتی که در این زمینه وجود دارد، مشکل تنگی داده ۱۰ و کاربران شروع سرد ۱۱ می‌باشد. در صورتی که یک کاربر هدف به تعداد کمی از آیتم‌ها رتبه داده باشد، محاسبه‌ی شباهت بین این کاربر و سایر کاربران موجود در سیستم با مشکل روبه‌رو خواهد شد. در نتیجه، انتخاب همسایگی مناسب برای این کاربر هدف کار راحتی نخواهد بود.

یکی از روش‌هایی که به منظور رفع مشکل انتخاب همسایگی در سیستم‌های توصیه‌گر مورد استفاده قرار گرفته است، روش خوشه‌بندی می‌باشد. خوشه‌بندی باعث می‌شود که کاربران و یا آیتم‌های مشابه در یک خوشه قرار بگیرند. بنابراین، می‌توان از کاربران موجود در خوشه‌ی کاربر هدف برای پیش‌بینی رتبه‌ی مورد نظر استفاده کرد. اگرچه روش خوشه‌بندی می‌تواند راه‌حلی برای رفع مشکل انتخاب همسایگی در سیستم‌های توصیه‌گر باشد، اما این روش نیز با محدودیت‌ها و مشکلاتی مواجه است. تعیین تعداد درست خوشه‌ها، یک موضوع اصلی در روش‌های مبتنی بر خوشه‌بندی می‌باشد. زیرا کارایی این روش‌ها بستگی به تعریف اولیه‌ی تعداد خوشه‌ها دارد و در صورتی که تعداد مناسبی برای خوشه‌ها تعیین نشود، این روش‌ها نمی‌توانند کارایی خوبی داشته باشند. مشکل دیگر در مورد نامناسب بودن خوشه‌های تولید شده می‌باشد. نتایج ضعیف خوشه‌بندی، ممکن است باعث کاهش دقت پیش‌بینی و همچنین کاهش نرخ پوشش رتبه‌ها شود. این مشکل بیشتر زمانی اتفاق می‌افتد که در طی فرایند خوشه‌بندی، خوشه‌هایی بدست می‌آیند که تعداد اعضای این خوشه‌ها خیلی کم است و لذا این خوشه‌ها نمی‌توانند همسایگی‌های مناسبی را برای کاربران موجود در این خوشه‌ها فراهم کنند. از طرف دیگر، بسیاری از سیستم‌های توصیه‌گر مبتنی بر خوشه‌بندی، فقط از معیار شباهت بین کاربران و یا آیتم‌ها برای خوشه‌بندی آنها استفاده می‌کنند. این کار باعث می‌شود که روش خوشه‌بندی نتواند خوشه‌بندی را به خوبی برای کاربران شروع سرد و همچنین در داده‌های با تنگی بالا انجام دهد. بنابراین، استفاده از فاکتورهای دیگر در کنار معیار شباهت، مانند روابط اعتماد، می‌تواند به این روش‌ها برای دسته‌بندی بهتر کاربران و یا آیتم‌های موجود در سیستم کمک کند.

### روش پیشنهادی

در این تحقیق، یک روش جدید مبتنی بر حافظه<sup>۱۲</sup> به منظور افزایش کارایی سیستم‌های توصیه‌گر مبتنی بر اعتماد معرفی می‌شود. هدف اصلی این روش پیشنهادی، استفاده از معیار قابلیت اطمینان به منظور افزایش دقت پیش‌بینی رتبه‌ها در

<sup>10</sup> Data Sparsity

<sup>11</sup> Cold Start Users

<sup>12</sup> Memory-Based

سیستم‌های توصیه‌گر مبتنی بر اعتماد می‌باشد. بدین منظور، یک معیار قابلیت اطمینان جدید معرفی شده است که از این معیار برای ارائه‌ی یک مکانیزم بازسازی شبکه‌ی اعتماد کاربران در سیستم‌های توصیه‌گر مبتنی بر اعتماد استفاده شده است. ساختار روش پیشنهادی در شکل (۳-۱) نشان داده شده است که دارای شش مرحله شامل: (۱) خوشه بندی و ساخت شبکه‌ی اعتماد، (۲) پیش‌بینی رتبه‌ی اولیه، (۳) محاسبه‌ی معیار قابلیت اطمینان، (۴) بازسازی شبکه‌ی اعتماد، (۵) پیش‌بینی رتبه‌ی نهایی، و (۶) پیشنهاد Top-N آیت‌م موردعلاقه‌ی کاربر هدف، می‌باشد. در ادامه، هر کدام از مراحل روش پیشنهادی به صورت مفصل بیان شده‌اند.

با توجه به شکل زیر، روش پیشنهادی مبتنی بر معیار قابلیت اطمینان از چهار فاز کلی تشکیل شده است. ابتدا در فاز اول، پیش پردازش اولیه داده‌ها انجام می‌گیرد. در فاز دوم شبکه اعتماد ساخته می‌شود و یک پیش‌بینی اولیه از رتبه‌ی آیت‌م مورد نظر برای کاربر هدف بر اساس شبکه‌ی اعتماد اولیه‌ی تشکیل شده انجام می‌گیرد. سپس بر اساس این رتبه‌ی اولیه، معیار ارائه شده برای قابلیت اطمینان محاسبه می‌گردد. در صورتی که معیار محاسبه شده برای قابلیت اطمینان رتبه‌ی مورد نظر از یک حد آستانه‌ی مشخص پایین‌تر باشد، روش ارائه شده برای بازسازی شبکه‌ی اعتماد کاربر هدف اجرا شده و شبکه‌ی اعتماد جدیدی برای این کاربر بدست خواهد آمد. سپس در فاز دوم، بر اساس شبکه‌ی اعتماد جدید به دست آمده، رتبه‌ی نهایی برای آیت‌م مورد نظر بدست می‌آید. در نهایت بر اساس رتبه‌های پیش‌بینی شده برای کاربر هدف، N آیت‌م با رتبه‌ی بالاتر به عنوان آیت‌م‌های موردعلاقه‌ی کاربر هدف به وی پیشنهاد می‌شوند.



شکل ۳: ساختار روش پیشنهادی مبتنی بر خوشه‌بندی

### خوشه‌بندی

در این بخش، یک روش خوشه‌بندی جدید و مبتنی بر تئوری گراف و الگوریتم رقابت استعماری به منظور دسته‌بندی کاربران و یا آیتم‌های موجود در سیستم ارائه شده است. این روش خوشه‌بندی ارائه شده در واقع فاز اول روش پیشنهادی مبتنی بر خوشه‌بندی را تشکیل می‌دهد که شامل سه مرحله می‌باشد. در مرحله اول، از الگوریتم تغییریافته‌ی یافتن چگال‌ترین زیرگراف که در شکل 4 آورده شده است، برای یافتن مراکز اولیه‌ی خوشه‌ها استفاده می‌شود. این الگوریتم، یک زیرمجموعه از کاربران و یا آیتم‌ها با بیشترین عدم وابستگی<sup>۱۳</sup> بر اساس گراف ورودی ایجاد می‌کند و به صورت خودکار مراکز اولیه‌ی خوشه‌ها را تشکیل می‌دهد. بنابراین، در روش خوشه‌بندی پیشنهاد شده نیازی به تعیین تعداد خوشه‌های ورودی توسط کاربر نمی‌باشد.

<sup>13</sup> Maximally Independent



فرض کنید گراف  $G = (V, E)$ ، یک گراف غیرجهتدار باشد. برای یک زیرمجموعه  $S \subseteq V$ ، مجموعه‌ی یال‌ها را با  $E(S)$  نشان می‌دهیم که به صورت  $E(S) = E \cap S^2$  تعریف می‌شود. چگالی زیرمجموعه‌ی  $S$  که با  $\rho(S)$  نشان داده می‌شود، به صورت  $\rho(S) = \frac{|E(S)|}{|S|}$  تعریف می‌گردد، که  $|S|$  اندازه‌ی زیرمجموعه‌ی  $S$  را مشخص می‌کند. درجه‌ی یک راس  $i \in S$  را با  $deg_S(i)$  نشان می‌دهند و به صورت  $deg_S(i) = |\{j | (i, j) \in E(S)\}|$  تعریف می‌شود. ماکزیمم چگالی گراف  $G$  به صورت  $\rho^*(G) = \max_{S \subseteq V} \{\rho(S)\}$  تعریف می‌شود [۱۱]. در مورد گراف‌های وزندار، چگالی گراف به صورت  $\rho(S) = \frac{\sum_{e \in E(S)} w_e}{|S|}$  تعریف می‌شود، که  $w_e$  وزن یال  $e \in E(S)$  می‌باشد. برای پارامتر  $k > 0$  و یک گراف غیرجهتدار  $G$ ، ماکزیمم چگالی گراف بزرگتر از حد آستانه به صورت  $\rho_{\geq k}^*(G) = \max_{S \subseteq V, |S| \geq k} \rho(S)$  تعریف می‌شود [۱۱].

مسئله‌ی یافتن چگال‌ترین زیرگراف برای یک گراف داده شده، یک مسئله‌ی NP-hard می‌باشد. بنابراین، چندین روش تقریبی برای مسئله‌ی یافتن چگال‌ترین زیرگراف در یک زمان قابل قبول، ارائه شده‌اند [۱۱]. در [۱۱]، چندین الگوریتم تقریبی برای مسئله‌ی یافتن چگال‌ترین زیرگراف، برای هر دو نوع گراف جهتدار و غیرجهتدار ارائه شده است. یکی از این الگوریتم‌های پیشنهاد شده می‌تواند چگال‌ترین زیرگراف را با اعمال محدودیت حداقل اندازه‌ی زیرگراف  $(\rho_{\geq k}^*(G))$ ، به دست بیاورد.

### Algorithm 1. The modified dense subgraph finding

**Require:**  $G = (V, E)$ ,  $k > 0$ , and  $r > 0$

**Algorithm:**

1:  $S, \tilde{S} \leftarrow V$ ;

2: **while**  $S \neq \emptyset$  **do**

3:  $\tilde{A}(S) \leftarrow \{i \in S | deg_S(i) \geq 2 * \rho(S)\}$ ;

4: **Let**  $A(S) \subseteq \tilde{A}(S)$ , **with**  $|A(S)| = r * |\tilde{A}(S)|$ ;

5:  $S \leftarrow S \setminus A(S)$ ;

6: **if**  $|S| \geq k$  **and**  $\rho(S) < \rho(\tilde{S})$  **then**

7:  $\tilde{S} \leftarrow S$ ;

8: **end if**

9: **end while**

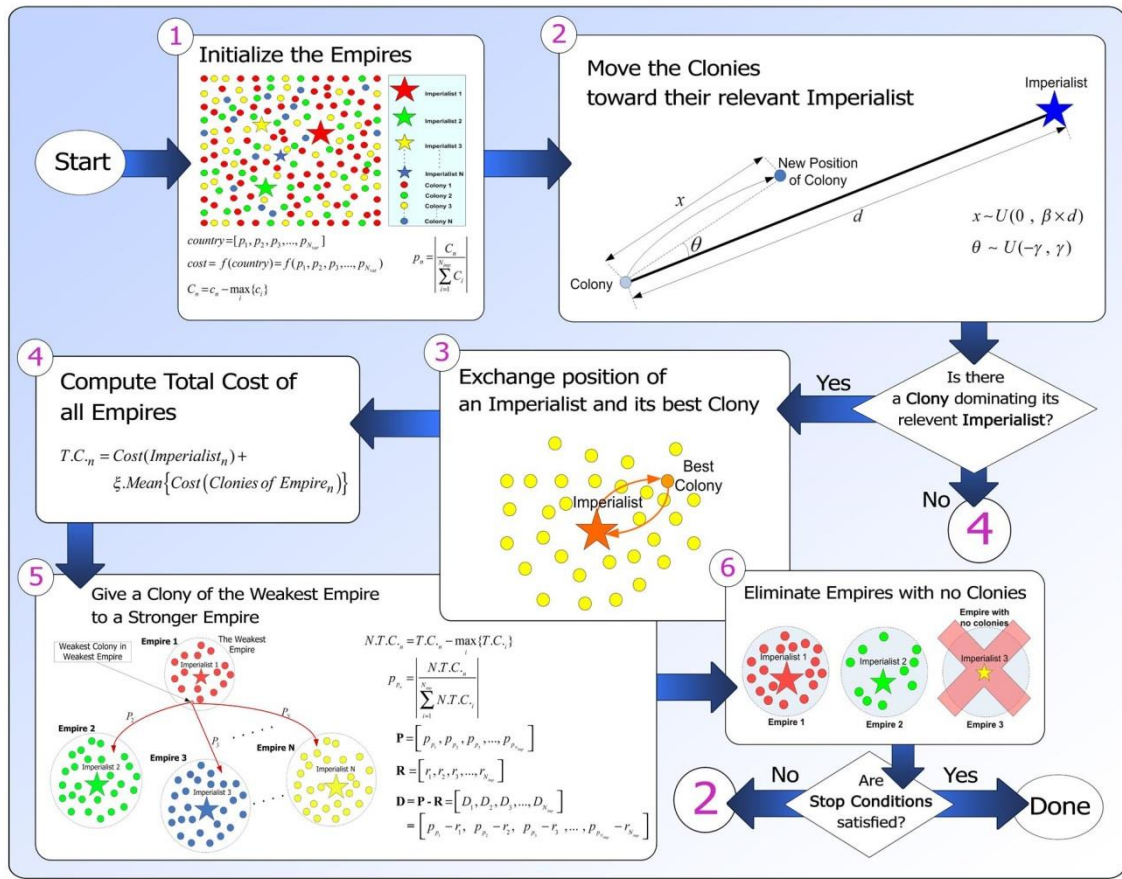
10: **return**  $\tilde{S}$ ;

شکل ۴: شبه‌کد الگوریتم تغییر یافته‌ی مبتنی بر محدودیت برای یافتن چگال‌ترین زیرگراف.

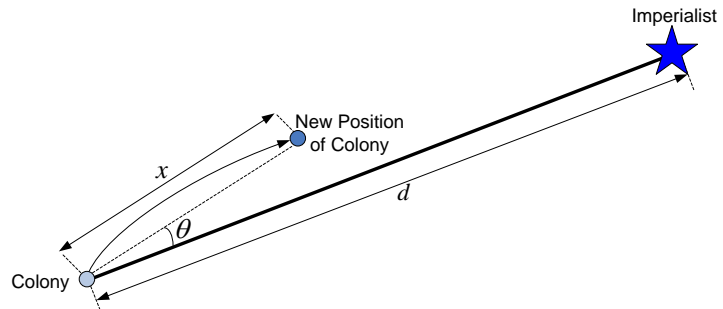
در روش پیشنهاد شده، از الگوریتم مبتنی بر محدودیت معرفی شده در [۱۱] برای یافتن چگال‌ترین زیرگراف استفاده شده است. از زیرگراف بدست آمده به عنوان مجموعه‌ی مراکز اولیه‌ی الگوریتم خوشه‌بندی ارائه شده استفاده می‌شود. بر اساس الگوریتم‌های خوشه‌بندی، مراکز خوشه‌ها باید تا حد امکان نسبت به معیار شباهت از هم دور باشند. بنابراین، هدف اصلی این مرحله، یافتن یک زیرمجموعه از رئوس  $\tilde{S} \subseteq V$  با حداقل تعداد رئوس  $k$  است که میانگین چگالی آنها مینیمم باشد. به همین دلیل، در روش پیشنهادی، الگوریتم ارائه شده در [۱۱] برای یافتن چگال‌ترین زیرگراف، به گونه‌ای تغییر داده شده است که

یک زیرگراف با کمترین چگالی بدست آورد. شبه‌کد الگوریتم تغییر یافته‌ی مبتنی بر محدودیت برای یافتن چگال‌ترین زیرگراف در شکل (۵-۱) آورده شده است. این الگوریتم بر روی گراف ورودی  $G$  اجرا شده و چگالی آن را محاسبه می‌کند. در مرحله‌ی بعد، رئوسی که درجه‌ی آنها بیشتر از یک حد آستانه‌ی  $\theta$  باشد، به عنوان رئوس کاندید تشخیص داده می‌شوند که می‌توانند از گراف ورودی حذف شوند. رئوس کاندید با  $\tilde{A}(S)$  نشان داده می‌شوند و مقدار آستانه به صورت  $\theta = 2 * \rho(S)$  تعریف می‌گردد. سپس تعدادی از رئوس موجود در مجموعه‌ی کاندید که با  $A(S)$  نشان داده می‌شوند، از مجموعه‌ی کاندید حذف می‌شوند. در صورتیکه زیرگراف بدست آمده غیرتهی باشد، الگوریتم بر روی گراف باقیمانده اجرا می‌شود. در نهایت، مجموعه‌ی رئوس به مجموعه‌ی  $S \setminus A(S)$  کاهش می‌یابد و الگوریتم تضمین می‌کند که زیرگراف بدست آمده حداقل دارای  $k$  راس می‌باشد.

مراکز اولیه‌ی خوشه‌ها که در مرحله‌ی قبل به دست آمده‌اند، ممکن است مراکز بهینه‌ی نهایی نباشند. برای رفع این مشکل، در مرحله‌ی دوم الگوریتم خوشه‌بندی، با استفاده از الگوریتم رقابت استعماری این مراکز بهینه سازی می‌شوند. الگوریتم رقابت استعماری، همانند سایر روش‌های بهینه‌سازی تکاملی، با تعدادی جمعیت اولیه شروع می‌شود. در این الگوریتم، هر عنصر جمعیت، یک کشور نامیده می‌شود. کشورها به دو دسته مستعمره و استعمارگر تقسیم می‌شوند. هر استعمارگر، بسته به قدرت خود، تعدادی از کشورهای مستعمره را به سلطه خود درآورده و آن‌ها را کنترل می‌کند. سیاست جذب و رقابت استعماری، هسته اصلی این الگوریتم را تشکیل می‌دهند. مطابق سیاست جذب که به صورت تاریخی، توسط کشورهای استعماری همچون فرانسه و انگلیس، در مستعمراتشان اعمال می‌شد، کشورهای استعمارگر با استفاده از روش‌هایی همچون احداث مدارس به زبان خود، سعی در از خود بی خود کردن کشور مستعمره، با از میان بردن زبان کشور مستعمره و فرهنگ و رسوم آن داشتند. در ارائه این الگوریتم، این سیاست با حرکت دادن مستعمرات یک امپراطوری، مطابق یک رابطه خاص صورت می‌پذیرد. شکل ۵ این حرکت را نشان می‌دهد.



شکل ۵: شمای کلی الگوریتم رقابت استعماری [۱۲]



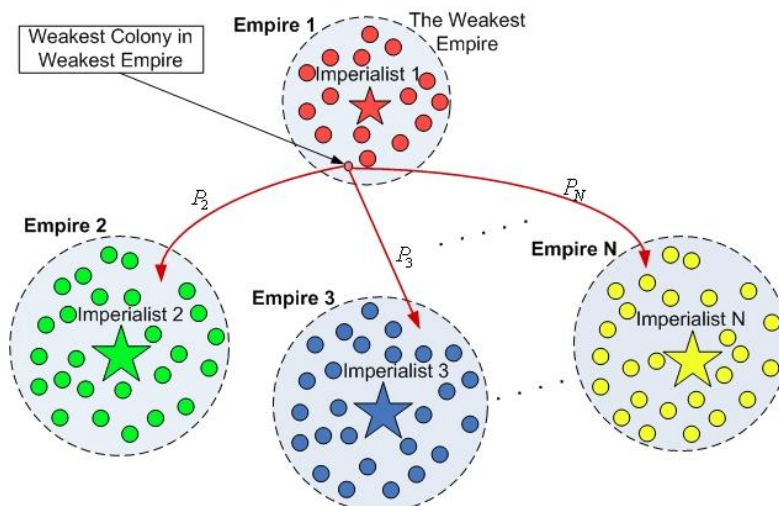
شکل ۶: حرکت مستعمرات به سمت امپریالیست (سیاست جذب)

اگر در حین حرکت، یک مستعمره، نسبت به استمارگر، به موقعیت بهتری برسد، جای آن دو با هم عوض می‌شوند. در ضمن، قدرت کل یک امپراطوری به صورت مجموع قدرت کشور استعمارگر به اضافه درصدی از قدرت میانگین مستعمرات آن تعریف می‌شود. یعنی

$$T.C._n = Cost(imperialist_n) + \xi \cdot mean\{Cost(colonies\ of\ empire_n)\}$$

همانگونه که قبلاً نیز بدان اشاره شد، رقابت استعماری، بخش مهم دیگری از این الگوریتم را تشکیل می‌دهد. در طی رقابت استعماری، امپراطوری‌های ضعیف، به تدریج قدرت خود را از دست داده و به مرور زمان با تضعیف شدن از بین می‌روند. رقابت

استعماری باعث می‌شود که به مرور زمان، به حالتی برسیم که در آن تنها یک امپراطوری در دنیا وجود دارد که آن را اداره می‌کند. این حالت زمانی است که الگوریتم رقابت استعماری با رسیدن به نقطه بهینه تابع هدف، متوقف می‌شود. شکل 6 زیر شمای کلی رقابت استعماری را نشان می‌دهد.



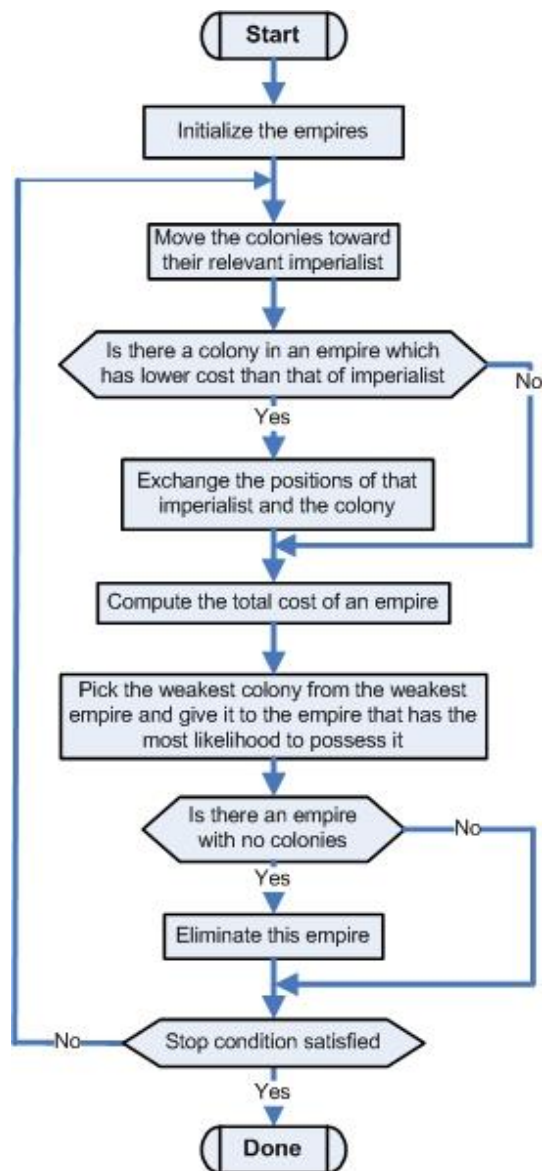
شکل ۷: شمای کلی رقابت استعماری [۱۲]

شکل 8 فلوچارت الگوریتم رقابت استعماری را نشان می‌دهد. همانند دیگر الگوریتم‌های تکاملی، این الگوریتم، نیز با تعدادی جمعیت اولیه تصادفی که هر کدام از آنها یک "کشور" نامیده می‌شوند؛ شروع می‌شود. تعدادی از بهترین عناصر جمعیت (معادل نخبه‌ها در الگوریتم ژنتیک) به عنوان امپریالیست<sup>۱۴</sup> انتخاب می‌شوند. باقیمانده جمعیت نیز به عنوان مستعمره<sup>۱۵</sup>، در نظر گرفته می‌شوند. استعمارگران بسته به قدرتشان، این مستعمرات را با یک روند خاص که در ادامه می‌آید؛ به سمت خود می‌کشند. قدرت کل هر امپراطوری، به هر دو بخش تشکیل دهنده آن یعنی کشور امپریالیست (به عنوان هسته مرکزی) و مستعمرات آن، بستگی دارد. در حالت ریاضی، این وابستگی با تعریف قدرت امپراطوری به صورت مجموع قدرت کشور امپریالیست، به اضافه در صدی از میانگین قدرت مستعمرات آن، مدل شده است.

با شکل‌گیری امپراطوری‌های اولیه، رقابت امپریالیستی میان آنها شروع می‌شود. هر امپراطوری‌ای که نتواند در رقابت استعماری، موفق عمل کرده و بر قدرت خود بیفزاید (و یا حداقل از کاهش نفوذش جلوگیری کند)، از صحنه رقابت استعماری، حذف خواهد شد. بنابراین بقای یک امپراطوری، وابسته به قدرت آن در جذب مستعمرات امپراطوری‌های رقیب، و به سيطرة در آوردن آنها خواهد بود. در نتیجه، در جریان رقابت‌های امپریالیستی، به تدریج بر قدرت امپراطوری‌های بزرگ‌تر افزوده شده و امپراطوری‌های ضعیف‌تر، حذف خواهند شد. امپراطوری‌ها برای افزایش قدرت خود، مجبور خواهند شد تا مستعمرات خود را نیز پیشرفت دهند.

<sup>14</sup> Imperialist

<sup>15</sup> Colony



شکل ۸: فلوجارت الگوریتم پیشنهادی [۱۲]

با گذشت زمان، مستعمرات، از لحاظ قدرت به امپراطوری‌ها نزدیک‌تر خواهند شد و شاهد یک نوع همگرایی خواهیم بود. حد نهایی رقابت استعماری، زمانی است که یک امپراطوری واحد در دنیا داشته باشیم، با مستمراتی که از لحاظ موقعیت، به خود کشور امپریالیست، خیلی نزدیک هستند.

### ارزیابی روش پیشنهادی

در این بخش قصد داریم، نتایج بدست آمده برای روش پیشنهادی را به منظور ارزیابی روش ارائه شده با روش‌های جدید دیگر مورد بررسی قرار دهیم. به منظور ارزیابی روش ارائه شده از مجموعه داده Epinions استفاده شده است. در این مجموعه داده، با استخراج کاربران و آیتم‌های متفاوت می‌توانیم ارزیابی‌های متفاوتی را بر روی نتایج بدست آمده داشته باشیم. در انتها، به منظور سنجش میزان خطا و پوشش روش پیشنهادی نسبت به سایر روش‌ها، از معیارهای MAE، MAUE، RC و UC کمک گرفته شده که روابط هر یک از آنها در بخش ۳-۴ نشان داده شده است.

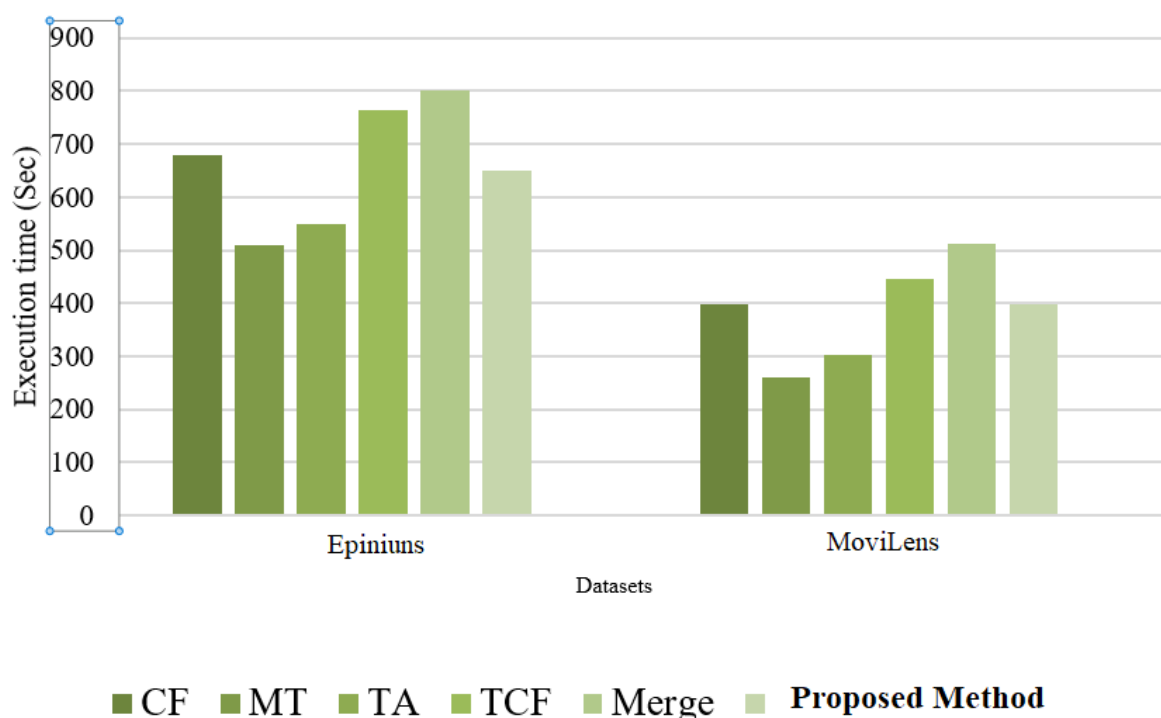
همانطور که گفته شد، به منظور ارزیابی بیشتر، از میان مجموعه داده‌های Epinions، کاربران و آیتم‌های متفاوت استخراج شده‌اند که گروه‌های استخراج شده عبارتند از:

- کاربران Cold: کاربرانی که کمتر از ۵ آیتم را رتبه‌بندی کرده‌اند.
- کاربران Heavy raters: کاربرانی که بیش از ۱۰ آیتم را رتبه‌بندی کرده‌اند.
- کاربران Opinionated: کاربرانی که بیش از ۴ آیتم را رتبه‌بندی کرده‌اند و انحراف معیار استاندارد رتبه‌بندی آنها بیش از ۱/۵ است.
- کاربران Black sheep: کاربرانی که بیش از ۴ آیتم را رتبه‌بندی کرده‌اند و اختلاف بین رتبه‌بندی‌های آنها برای یک آیتم خاص و میانگین رتبه‌های داده شده به آن آیتم بیش از ۱ باشد.
- آیتم‌های Controversial: آیتم‌هایی که انحراف معیار استاندارد رتبه‌بندی آنها بیش از ۱/۵ است.
- آیتم‌های Niche: آیتم‌هایی که کمتر از ۵ بار رتبه‌بندی شده‌اند.

تحلیل زمان اجرای روش‌های پیشنهادی

در این بخش روش‌های پیشنهادی ارائه شده را نسبت به سایر روش‌ها از لحاظ زمان اجرا<sup>۱۶</sup> مورد مقایسه قرار دهیم. لازم به ذکر است، به منظور پیاده سازی روش‌های پیشنهادی ارائه شده، و برای انجام محاسبات از سیستمی با پردازنده core-i5 و حافظه ۴ گیگابایت استفاده کرده‌ایم. علاوه بر این، در آزمایشات از مجموعه داده‌های Epinions و MovLens استفاده شده و میزان زمان صرف شده برای هر مجموعه داده با توجه به روشی که به منظور پیش‌بینی رتبه‌ها به کار گرفته شده نشان داده خواهد شد. همچنین به دلیل بزرگ بودن مجموعه داده Epinions، با توجه به اینکه تعداد کاربران در مجموعه داده MovLens برابر با ۱۹۸۶ است. در مجموعه داده‌های Epinions نیز، تنها از ۱۹۸۶ کاربر به منظور محاسبه زمان اجرای روش‌های متفاوت بر روی مجموعه داده‌های مورد نظر استفاده شده است. همانطور که در شکل ۹ نشان داده شده، زمان اجرای روش‌های متفاوت را بر روی مجموعه داده‌های Epinions و MovLens می‌بینیم.

<sup>16</sup> Execution time



شکل ۹: مقایسه زمان اجرای روش‌های متفاوت با روش‌های پیشنهادی ارائه شده

براساس نتایج بدست آمده در مورد زمان اجرای روش‌های متفاوت، در بین روش‌های ارائه شده، روش MT با توجه به اینکه تنها از کاربران مورد اعتماد کاربر فعال جهت پیش‌بینی رتبه‌ها استفاده می‌کند، زمان اجرای کمتری را در مقایسه با سایر روش‌ها دارد. همچنین، روش پیشنهادی نسبت به روش MT دارای پیچیدگی محاسباتی بیشتر خواهد بود و این موضوع مربوط به پیدا کردن کاربران تحت تسلط در روش پیشنهادی است که به‌عنوان محاسبات اضافی میزان زمان اجرا را در این روش نسبت به روش MT افزایش داده است. علاوه بر این، در روش TA به‌منظور پیش‌بینی رتبه‌ها برای هر کاربر فعال، سایر کاربران به‌عنوان کاربران مورد اعتماد کاربر فعال در نظر گرفته می‌شوند، که این موضوع فرایند پیش‌بینی رتبه را زمانبر کرده و باعث افزایش زمان اجرای این روش نسبت به روش MT می‌گردد. براساس آنچه گفته شد، روش CF به‌منظور محاسبه مقدار رتبه‌های هر کاربر فعال نیاز به محاسبه شباهت کاربر فعال با سایر کاربران دارد، که این مسئله موجب افزایش زمان اجرای این روش نسبت به روش‌های MT، TA و PTARS خواهد شد. با توجه به نتایج بدست آمده برای روش‌های TCF، Merge و روش پیشنهادی که به‌منظور پیش‌بینی رتبه‌ها، ترکیبی از اعتماد و شباهت را به کار می‌گیرند، روش Merge و روش پیشنهادی به جهت محاسبه اطمینان دارای زمان اجرای بیشتری در مقایسه با TCF بوده و روش پیشنهادی به دلیل اینکه، علاوه بر محاسبه اطمینان برای کاربران و آیتم‌ها از مفهوم پرتو تسلط جهت حذف برخی از کاربران استفاده کرده، دارای زمان اجرای بیشتری نسبت به روش Merge می‌باشد.

درواقع نتایج بدست آمده در مورد زمان اجرای روش‌های متفاوت، نشان دهنده این موضوع است که بکارگیری اطلاعات اجتماعی اضافی در روش‌های مبتنی بر اعتماد، علاوه بر افزایش میزان کارایی این سیستم‌ها، زمان اجرا را افزایش خواهد داد. بنابراین به‌منظور ارائه یک روش مناسب، نیازمند این هستیم، که تناسب را در بین میزان کارایی و زمان اجرا حفظ کنیم.

## بحث و نتیجه گیری

براساس مطالعات انجام شده در مورد سیستم‌های توصیه‌گر، روش‌های پالایش گروهی برای سیستم‌های توصیه‌گر مقیاس بزرگ امیدوار کننده هستند، اما موفقیت آنها به‌طور عمده به همسایه‌های مشابه وابسته است، که با توجه به تنگی ماتریس رتبه‌بندی کاربر-آیتم، این فرایند اغلب موفقیت آمیز نیست. در واقع درک روابط بین کاربران و دوستان آنها، دانش ما را در مورد رفتار کاربران و رتبه‌بندی‌های آنها افزایش خواهد داد. براساس مطالعات صورت گرفته، استفاده از اطلاعات اجتماعی برخی از مسائل موجود در سیستم‌ها توصیه‌گر سنتی را رفع خواهد کرد. این موضوع اهمیت استفاده از اطلاعات اجتماعی را به‌منظور بهبود کارایی سیستم‌های توصیه‌گر سنتی نشان می‌دهد. علاوه بر این، با وجود ارائه راهکارهای متفاوت به‌منظور تولید توصیه‌هایی با دقت بالا، هنوز هم چالش‌های زیادی در این زمینه باقی مانده است. بنابراین ارائه یک سیستم مناسب که تا حد امکان قادر به حل مشکلات موجود باشد، می‌تواند به‌عنوان یک مسئله اساسی در سیستم‌های توصیه‌گر سنتی مطرح گردد.

با توجه به آنچه گفته شد، به دلیل وجود برخی مسائل در سیستم‌های توصیه‌گر پالایش گروهی، که تنها از معیارهای شباهت به‌منظور یافتن کاربران مشابه جهت محاسبه رتبه‌ها استفاده می‌کنند، سیستم‌های مبتنی بر اعتماد ارائه شدند. در واقع سیستم‌های توصیه‌گر مبتنی بر اعتماد، از اعتماد به‌عنوان یکی از اطلاعات اجتماعی، جهت رفع مشکلات سیستم‌های توصیه‌گر سنتی استفاده می‌کنند. به عبارت دیگر در این سیستم‌ها تنها از کاربران مورد اعتماد هر کاربر به‌منظور پیش‌بینی رتبه‌ها استفاده می‌شود و با انتشار اعتماد در این سیستم‌ها می‌توان مشکل مربوط به کاربران شروع سرد را در سیستم‌های پالایش گروهی رفع کرد. همچنین در سیستم‌های توصیه‌گر مبتنی بر اعتماد، به دلیل استفاده از کاربران مورد اعتماد برای هر کاربر، مسئله حمله افراد بدخواه از بین خواهد رفت و کاربران بدخواه دیگر نمی‌توانند با کپی برداری از مشخصات کاربر فعال نتایج را به نفع خود تغییر دهند. علاوه بر این، با توجه به نتایج بدست آمده برای روش پیشنهادی اول، دیدیم که در سیستم‌های توصیه‌گر مبتنی بر اعتماد در نظر گرفتن وزن یکسان برای کاربران می‌تواند در دقت چنین سیستم‌هایی تأثیر گذار باشد. بنابراین با تغییر وزن برخی از کاربران مورد اعتماد، میزاد دقت در سیستم‌های توصیه‌گر مبتنی بر اعتماد افزایش خواهد یافت. در نتیجه در نظر گرفتن وزن مناسب برای هر کاربر مورد اعتماد، در شبکه اعتماد ایجاد شده، می‌تواند تأثیر مثبتی را در پیش‌بینی رتبه‌ها برای کاربر فعال داشته باشد. همچنین یکی دیگر از مشکلات موجود در سیستم‌های توصیه‌گر مبتنی بر اعتماد، که در روش پیشنهادی دوم مورد بحث قرار گرفت، مسئله رتبه‌بندی آیتم‌ها برای کاربرانی بود که در شبکه اعتماد آنها هیچ دوست مورد اعتمادی وجود نداشت. در نتیجه، هیچ یک از آیتم‌ها برای چنین کاربرانی رتبه‌بندی نشده و میزان پوشش کاربران و آیتم‌ها در سیستم‌های توصیه‌گر مبتنی بر اعتماد کاهش خواهد یافت. از این رو، در روش‌های مبتنی بر اعتماد، در صورتی که اعتماد بین کاربران را با اطلاعات شباهت و اجتماعی کاربران و آیتم‌ها ترکیب کنیم، به نتایج بهتری دست خواهیم یافت. در واقع یکی از دلایل استفاده از مفهوم شباهت در کنار اعتماد و اطمینان کاربران و آیتم‌ها در روش پیشنهادی دوم، پوشش چنین مسئله‌ای در سیستم‌های توصیه‌گر مبتنی بر اعتماد است. البته لازم به ذکر است، که استفاده از هر گونه اطلاعات اضافی به‌منظور بهبود این سیستم‌ها، زمان اجرا را به‌منظور پیش‌بینی رتبه‌های مورد نیاز برای کاربر فعال افزایش خواهد داد.

اگر یک روش بتواند همسایگی‌های خوبی را برای کاربر هدف تعیین کند، قادر خواهد بود رتبه‌های مورد نظر را با دقت بالایی تخمین بزند و همچنین درصد رتبه‌هایی که پیش‌بینی می‌شود نیز افزایش می‌یابد. از جمله روش‌هایی که می‌تواند به سیستم‌های توصیه‌گر در فرآیند انتخاب همسایگی کمک کند، روش‌های مبتنی بر خوشه‌بندی می‌باشند. در واقع، یکی از



مشکلات این معیار، قابلیت و کارایی پائین آن در مواجهه با تنگی ماتریس کاربر-آیتم می‌باشد. بدین منظور، در این پژوهش یک معیار قابلیت اطمینان جدید پیشنهاد شده است که در آن علاوه بر معیار شباهت بین کاربران، از روابط اعتماد نیز استفاده شده است. علاوه بر این، یک مکانیزم جدید برای بازسازی شبکه‌های اعتماد کاربران به منظور بهبود دقت رتبه‌های پیش‌بینی شده با استفاده از معیار قابلیت اطمینان ارائه شده، معرفی شده است. این مکانیزم بازسازی پیشنهاد شده، بر روی شبکه‌های اعتمادی اجرا می‌شود که دارای قابلیت اطمینان پائینی هستند و کاربرانی که دارای قابلیت اطمینان پائین‌تری نسبت به یک حد آستانه باشند، از شبکه‌ی اعتماد کاربر هدف حذف می‌شوند. بنابراین، استفاده از این مکانیزم باعث ایجاد شبکه‌های اعتمادی با کیفیت بالاتر بر اساس معیار قابلیت اطمینان ارائه شده، می‌شود. علاوه بر این، استفاده از شبکه‌های اعتماد با کارایی بالا در سیستم‌های توصیه‌گر مبتنی بر اعتماد باعث افزایش دقت پیش‌بینی رتبه‌ها و در نتیجه ارائه توصیه‌های مطلوب‌تر به کاربران می‌شود. با توجه به اینکه در مکانیزم پیشنهاد شده برای بازسازی شبکه‌ی اعتماد، کاربرانی با قابلیت اطمینان پائین حذف می‌شوند، بنابراین، باید این نکته را در نظر داشت که روش پیشنهادی نباید باعث کاهش قابل توجه نرخ پوشش رتبه‌ها شود.

ارزیابی عملکرد روش پیشنهادی و مقایسه آن با کارهای پیشین عملکرد مناسب روش پیشنهادی را هم از نظر دقت و هم از نظر پیچیدگی محاسباتی نشان داد.

#### پیشنهادات و کارهای آتی

با توجه به تحقیقات بسیاری که در مورد سیستم‌های توصیه‌گر صورت گرفته است. هنوز هم چالش‌های بسیاری به‌منظور بررسی و رفع شدن در پیش رو هستند، که می‌توان با ارائه روش‌های نوین‌تر علاوه بر رفع چالش‌های موجود، میزان کارایی را در چنین سیستم‌هایی افزایش داد. از این رو، می‌توان پیشنهادهایی را به‌منظور بهبود هر چه بیشتر چنین سیستم‌هایی ارائه کرد که در ادامه به برخی از آنها اشاره خواهد شد.

#### • خوشه‌بندی گراف کاربران

در نظر گرفتن مسئله خوشه‌بندی گراف، یکی از مسائل مهم در شبکه‌های اجتماعی است، به این صورت که با یافتن خوشه‌های مناسب در بین کاربران می‌توانیم عمل انتشار را برای هر کاربر فعال تنها در خوشه‌ای که کاربر عضو آن می‌باشد انجام دهیم. علاوه بر این، می‌توان با توجه به خوشه‌های بدست آمده، برای رتبه‌بندی آیتم‌های کاربر فعال، در خوشه مربوطه از کاربرانی کمک بگیریم که از نفوذ بیشتری در بین سایر کاربران برخوردار هستند.

#### • داده‌کاوی اطلاعات و استخراج قوانین

همانطور که گفته شد، استفاده از اطلاعات اجتماعی در سیستم‌های توصیه‌گر منجر به بهبود عملکرد این سیستم‌ها خواهد شد. در نتیجه، ارائه سیستم‌هایی که براساس روش‌های داده‌کاوی بتوانند قوانینی را از روابط کاربران و آیتم‌هایی استخراج کنند، میزان کارایی سیستم‌های توصیه‌گر را افزایش خواهد داد.

#### • فازی سازی روابط اعتماد

یکی از مسائل موجود در سیستم‌های توصیه‌گر، در نظر گرفتن مقدار اعتماد یکسان برای کاربرانی است که در شبکه اعتماد ایجاد شده برای کاربر فعال، در یک عمق خاص قرار دارند. بنابراین با ارائه روش‌ها و توابع عضویت فازی، می‌توان مقدار اعتماد را برای هر کاربر مورد اعتماد براساس توابع عضویت فازی به صورت بهینه محاسبه کنیم.

## منابع:

- [1]. A. I. Schein, et al. Methods and metrics for cold-start recommendation. in Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. 2002.
- [2]. Adomavicius, G. and A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 2005. 17(6): p. 734-749.
- [3]. Bobadilla, J. , et al. , Recommender systems survey. Knowledge-Based Systems, 2013. 46(0): p. 109-132.
- [4]. Breese, J. S. , D. Heckerman, and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, in Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence. 1998. p. 43-52.
- [5]. Burke, R. , Hybrid Web Recommender Systems, in The Adaptive Web, P. Brusilovsky, A. Kobsa, and W. Nejdl, Editors. 2007. p. 377-408.
- [6]. D. Billsus and M. J. Pazzani, User modeling for adaptive news access. User modeling and user-adapted interaction, 2000. 10: p. 147-180.
- [7]. D. Goldberg, et al. , Using collaborative filtering to weave an information tapestry. Communications of the ACM, 1992. 35: p. 61-70.
- [8]. Fields, Ben, et al. "Analysis and exploitation of musician social networks for recommendation and discovery. " Multimedia, IEEE Transactions on 13. 4 (2011): 674-686.
- [9]. G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions, 2005. 17: p. 734-749.
- [10]. I. Soboroff and C. Nicholas. Combining content and collaboration in text filtering. in Proceedings of the IJCAI. 1999.
- [11]. I. Zaihrayeu, P. P. Da Silva, and D. L. McGuinness, "IWTrust: Improving user trust in answers from the web," in *Trust Management*, Springer, 2005, pp. 384–392. DOI: 10.1007/11429760\_27
- [12]. J. A. Konstan, et al. , GroupLens: applying collaborative filtering to Usenet news. Communications of the ACM, 1997. 40: p. 77-87.
- [13]. K. Goldberg, et al. , Eigentaste: A constant time collaborative filtering algorithm. Information Retrieval, 2001. 4: p. 133-151.
- [14]. K. Musial, Recommender system for online social network. LAP Lambert Academic Publishing, 2009.
- [15]. M. Pazzani and D. Billsus, Learning and revising user profiles: The identification of interesting web sites. Machine learning, 1997. 27: p. 313-331.
- [16]. McPherson, M. , L. Smith-Lovin, and J. M. Cook, Birds of a Feather: Homophily in Social Networks. Annu. Rev. Sociol. , 2001. 27(1): p. 415-444.
- [17]. P. Massa and B. Bhattacharjee, Using trust in recommender systems: an experimental analysis. Trust Management, ed: Springer, 2004: p. 221-235.
- [18]. P. Massa and P. Avesani, "Trust metrics in recommender systems," in Computing with social trust, Springer, 2009, pp. 259–285. DOI:10.1007/978-1-84800-356-9\_10

- [19]. P. Resnick, et al. GroupLens: an open architecture for collaborative filtering of netnews. in Proceedings of the 1994 ACM conference on Computer supported cooperative work. 1994.
- [20]. Pazzani, M. J. , A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 1999. 13: p. 393-408.
- [21]. R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, vol. 12, pp. 331-370, 2002.
- [22]. R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. in Proceedings of the fifth ACM conference on Digital libraries. 2000.
- [23]. Schafer, J. B. , et al. , Collaborative Filtering Recommender Systems, in *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Editors. 2007, Springer Berlin Heidelberg. p. 291-324.
- [24]. Schafer, J. B. , J. Konstan, and J. Riedl, Recommender systems in e-commerce, in Proceedings of the 1st ACM conference on Electronic commerce. 1999, ACM. p. 158-166.
- [25]. Steck, Harald. "Item popularity and recommendation accuracy. " Proceedings of the fifth ACM conference on Recommender systems. ACM, 2011.
- [26]. T. Tran and R. Cohen. Hybrid recommender systems for electronic commerce. in Proceedings of Knowledge-Based Electronic Markets, Papers from the AAI Workshop, Technical Report WS-00-04, AAI Press. 2000.
- [27]. Xiwang, Y. , G. Yang, and L. Yong. Bayesian-inference based recommendation in online social networks. in INFOCOM, 2011 Proceedings IEEE. 2011.
- [28]. Y. Matsuo and H. Yamamoto, Community gravity: Measuring bidirectional effects by trust and rating on online social networks, in: *WWW*, (2009), 751–760. doi: 10.1145/1526709.1526810
- [29]. Yang, X. , et al. , A survey of collaborative filtering based social recommender systems. *Computer Communications*, 2014. 41(0): p. 1-10.
- [30]. Zhou, Tao, et al. "Solving the apparent diversity-accuracy dilemma of recommender systems. " *Proceedings of the National Academy of Sciences* 107. 10 (2010): 4511-4515.