

## ارائه یک سیستم توصیه‌گر ترکیبی برای تجارت الکترونیک

مصطفی رجبزاده<sup>۱</sup>، دانش‌آموخته کارشناسی ارشد، رضا رافع<sup>۲</sup>، استادیار

۱- دانشکده فنی مهندسی - دانشگاه اراک - اراک - ایران - mostafa.rajabzadeh@yahoo.com

۲- دانشکده فنی مهندسی - دانشگاه اراک - اراک - ایران - r-rafeh@araku.ac.ir

**چکیده:** رشد روزافزون تجارت الکترونیک در فضای مجازی باعث معرفی محصولات متنوع و زیادی از سوی شرکت‌های فعال در این حوزه شده است. در چنین شرایطی انتخاب مناسب و بهینه محصولات از میان حجم انبوه اطلاعات ارائه‌شده برای کاربران کاری مشکل است. سیستم‌های توصیه‌گر با توجه به ویژگی‌ها، رفتار و علایق کاربران سعی می‌کنند بهترین و مناسب‌ترین موارد را به کاربران خود پیشنهاد دهند. در این مقاله یک سیستم توصیه‌گر ترکیبی با استفاده از دو روش پالایش گروهی و محدودیت‌محور ارائه شده است که با توجه به استفاده از روش محدودیت‌محور مشکل شروع سرد روش پالایش گروهی را برطرف کرده است. پایگاه دانش روش محدودیت‌محور برای کاربران نیز با استفاده از قوانین استخراج‌شده از تعاملات کاربران همسایه کاربر هدف استخراج شده است که از این نظر هم مشکل جمع‌آوری پایگاه دانش را که به‌صورت دستی زمان زیادی نیاز دارد برطرف نموده است. کارایی سیستم ارائه‌شده بر روی مجموعه داده MovieLens در حوزه فیلم ارزیابی شده و نتایج نشان می‌دهد که روش ارائه‌شده کارایی بهتری نسبت به روش پالایش گروهی دارد.

**واژه‌های کلیدی:** سیستم توصیه‌گر، پالایش گروهی، محدودیت، پایگاه دانش.

## Proposing a Hybrid Recommender System for E-commerce

M. Rajabzadeh, MS<sup>1</sup>, R. Rafeh, Assistant Professor<sup>2</sup>

1- Faculty of Engineering, Arak University, Arak, Iran, Mostafa.rajabzadeh@yahoo.com

2- Faculty of Engineering, Arak University, Arak, Iran, r-rafeh@araku.ac.ir

**Abstract:** The rapid growth of e-commerce has led to introducing various products from companies which are active in this area. As a result, choosing an appropriate product has become difficult for a customer. Recommender systems aimed at offering the best and most appropriate items to users based on their interests and their previous activities. This paper propose a hybrid recommender system which combines two approaches: collaborative filtering and constraint satisfaction. The proposed system, on the one hand, solves the cold start problem of the collaborative filtering technique because of using constraints. On the other hand, it solves the problem of constructing the knowledge base, which is needed for constraint-based methods. Instead of constructing the knowledge base manually, which is time consuming, the knowledge base of the constraint-based approach is extracted from interactions among the target user's neighbors. Using the "MovieLens" dataset for evaluation, the experimental results show that the performance of the proposed approach is better than the collaborative filtering method.

**Keywords:** Recommender system, collaborative filtering, constraint, knowledge base

تاریخ ارسال مقاله: ۱۳۹۳/۰۳/۱۰

تاریخ اصلاح مقاله: ۱۳۹۳/۰۶/۱۱

تاریخ پذیرش مقاله: ۱۳۹۳/۰۶/۱۷

نام نویسنده مسئول: رضا رافع

نشانی نویسنده مسئول: ایران-اراک - دانشگاه اراک- دانشکده فنی مهندسی

## ۱- مقدمه

محدودیت محور بررسی شده، در بخش ۶ سیستم توصیه گر ترکیبی ارائه شده در این مقاله شرح داده شده است. در بخش ۷ سیستم ارائه شده مورد ارزیابی قرار گرفته و در بخش ۸ نیز نتیجه گیری شده است.

## ۲- سیستم های توصیه گر

یک سیستم توصیه گر سیستمی است که توانایی تحلیل رفتارهای گذشته و ارائه توصیه هایی برای مسائل جاری را داشته باشد. به عبارت دیگر در سیستم های توصیه گر 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 آن ها در سیستم موجود است) مناسب ترین و نزدیک ترین کالا به سلیقه او شناسایی و به وی پیشنهاد داده شود [۲].

توصیه هایی که از سوی سیستم های توصیه گر ارائه می شوند به طور کلی می توانند دو نتیجه در برداشته باشند:

- کاربر را در تصمیم گیری یاری کنند (مثلاً از میان چندین گزینه پیش رو کدام بهتر است).
- موجب افزایش آگاهی کاربر در زمینه مورد علاقه وی می شوند (مثلاً در حین ارائه توصیه به کاربر موجب می شود تا وی با اقلام و اشیاء جدیدی که قبلاً آن ها را نمی شناخته، آشنا شود) [۲].

سیستم های توصیه گر کاربردهای فراوانی دارند که برخی از آن ها به شرح زیر است [۱]:

- تجارت الکترونیک: برای توصیه محصولات و خدمات مختلف.
- اینترنت بنگاهی: برای پیدا کردن افراد خبره در یک زمینه خاص و یا افرادی که در رویارویی با شرایط مشابه، تجاربی کسب کرده و راه حل هایی یافته اند (درون سازمانی).
- کتابخانه دیجیتال: برای پیدا کردن کتاب، مقاله و اقلام دیجیتالی دیگر.
- کاربردهای پزشکی: برای انتخاب پزشک متناسب با شرایط (مکان، نوع بیماری و زمان) بیمار، انتخاب دارو.
- مدیریت ارتباط با مشتری: برای ارائه راهکارهایی جهت حل مشکلات تولیدکننده و مصرف کننده در زنجیره تأمین.

## ۳- سیستم های توصیه گر مبتنی بر پالایش گروهی

یکی از پرکاربردترین روش های توسعه سیستم های توصیه گر استفاده از متد پالایش گروهی است. این متد یک فرآیند اتوماتیک در جهت رتبه بندی محتوای سیستم برای کاربر هدف است که این کار را بر اساس ارزیابی کاربران دیگر انجام می دهد. این نوع سیستم ها از مقایسه شباهت خصوصیات کاربر هدف با گروه دیگری از کاربران که زاویه دید

از جمله کارهایی که امروزه روی شبکه اینترنت رشد و توسعه یافته است خرید اینترنتی یا در معنای جامع تر تجارت الکترونیک است. فروشندگان زیادی بدون محدودیت مکانی و زمانی اقدام به عرضه کالا می کنند. کاربران هم خود را در مقابل خبر، فیلم، کتاب و صدها هزار اقلام دیگر می بینند و با توجه به سلیقه و نیاز خود نمونه ای از آن ها را انتخاب می کنند. اما زمانی که کاربران با حجم عظیمی از اطلاعات روبه رو می شوند مشکل انتخاب به وجود می آید و جست و جو میان این فضا زمان زیادی را از آن ها می گیرد. سیستم های توصیه گر<sup>۲</sup> سیستم های تأثیرگذار در راهنمایی و هدایت کاربر، در میان حجم عظیمی از انتخاب های ممکن برای رسیدن به گزینه مفید و مورد علاقه وی هستند [۱]. بسیاری از کاربران اینترنتی از روش های مختلفی به صورت خود آگاه یا ناخود آگاه با یک سیستم توصیه گر در ارتباط بوده اند. برای مثال زمانی که یکی از دوستان شما پیشنهاد خواندن یک کتاب جدید را به شما می دهد، شما برای سفارش این کتاب به یک کتاب فروشی آنلاین مراجعه می کنید، عنوان کتاب مورد نظر خود را وارد کرده، لیستی از کتاب های یافت شده به شما نشان داده می شود. در قسمت دیگری از وبسایت کتاب فروشی آنلاین در بخشی با عنوان (مشترسانی که این کتاب را خریده اند همچنین علاقه مند هستند) لیستی از کتبی نمایش داده می شود که با توجه به خریدهای قبلی مشتریان احتمالاً شما نیز به آن علاقه مند هستید. در صورتی که یکی از کاربران قدیمی تر این کتاب فروشی آنلاین باشید به محض ورود به وبسایت لیستی از کتاب های پیشنهادی به شما نشان داده می شود [۲].

سیستم توصیه گر با تحلیل رفتار کاربر خود، اقدام به پیشنهاد مناسب ترین اقلام (اطلاعات، کالا و خدمات) به وی می نماید. این گونه سیستم ها برای مواجهه با مشکلات ناشی از حجم فراوان و روبه رشد اطلاعات ارائه شده اند و به کاربران خود کمک می کنند تا در میان حجم عظیم اطلاعات سریع تر به هدف خود نزدیک شوند.

در این مقاله یک سیستم توصیه گر ارائه خواهد شد که از ترکیب دو روش محدودیت محور و پالایش گروهی استفاده می کند که در آن پایگاه دانش روش محدودیت محور با استفاده از قوانین استخراج شده از تعاملات کاربران همسایه کاربر هدف و به صورت خودکار استخراج می شود. جمع آوری پایگاه دانش به صورت دستی نیازمند زمان و هزینه زیاد است که در روش پیشنهادی این مشکل برطرف شده است. ارائه پیشنهادها به کاربر هدف بر اساس پایگاه دانش استخراج شده انجام می شود و به این ترتیب روش پیشنهادی مشکل شروع سرد روش پالایش گروهی را نیز ندارد.

در ادامه در بخش ۲ نمونه هایی از کارهای انجام شده در زمینه سیستم های توصیه گر بررسی شده است. در بخش ۳ به تعریف سیستم های توصیه گر پرداخته شده و کاربردهای آن بیان شده است. در بخش های ۴ و ۵ سیستم های توصیه گر پالایش گروهی و

پیاده‌سازی روش ارائه‌شده نشان داده است که سیستم از نظر خطای پیش‌بینی و کارایی عملکرد بهتری نسبت به روش پالایش گروهی محض از خود نشان داده است.

#### ۴ - سیستم‌های توصیه‌گر مبتنی بر پایگاه دانش

در مواردی که تعداد تعاملات قبلی کاربران با سیستم کم باشد سیستم‌های مبتنی بر پالایش گروهی نمی‌توانند بهینه عمل کنند. همچنین با گذشت زمان علائق کاربران با توجه به تغییر سبک و شرایط زندگی دچار تحولاتی می‌شود که از این نظر سیستم‌های پیشنهادگر مبتنی بر محتوا نیز نمی‌توانند برای کاربری که به‌طور مثال آخرین بار شش سال قبل کامپیوتر خریده است و اکنون دوباره قصد این کار را دارد دقیق عمل کنند. از طرف دیگر، کاربران سیستم‌هایی که دارای محصولات پیچیده مانند اتومبیل هستند دوست دارند نیازهای خود را در مورد نمونه‌هایی مانند قیمت و رنگ به‌طور صریح بیان کنند [۶].

سیستم‌های توصیه‌گر دانش‌محور در جهت برطرف کردن این چالش‌ها به وجود آمده‌اند که اطلاعات اضافی در مورد کالاها را در پایگاه دانش خود جمع‌آوری کرده، سعی دارند با استفاده از تعامل با کاربر پیشنهادهای دقیقی ارائه دهند [۶]. سیستم‌های دانش‌محور به دو دسته محدودیت‌محور<sup>۵</sup> و موردمحور<sup>۶</sup> تقسیم‌بندی شده‌اند. در هر دو روش فرآیند ارائه پیشنهادها یکسان است. کاربر نیازهای خود را بیان می‌کند و سیستم سعی می‌کند یک راه‌حل مناسب بیابد. اگر راه‌حلی وجود نداشته باشد سیستم نزدیک‌ترین راه‌حل‌های مشابه با خواسته کاربر را به او معرفی می‌کند [۶]. به‌عنوان مثال در سیستم‌های مبتنی بر محدودیت اگر سیستم راه‌حلی پیدا نکند سعی می‌کند با نادیده گرفتن محدودیت‌های نرمی که جریمه کم‌تری دارند پیشنهادهای خود را ارائه دهد.

در [۷] یک سیستم توصیه‌گر ترکیبی ارائه شده است که در آن از ترکیب دو روش پالایش گروهی<sup>۷</sup> و محدودیت‌محور استفاده شده است. ایده کلی این روش پیدا کردن یک جواب بر این اساس است که علاوه بر ارضای محدودیت‌های سخت، کالاهایی را پیشنهاد دهد که جریمه کم‌تری از لحاظ شکستن محدودیت‌های نرم ایجاد کند و از لحاظ نرخ پیش‌بینی‌شده از طریق روش پالایش گروهی دارای رتبه بالاتری باشند.

#### ۵ - سیستم توصیه‌گر ترکیبی پیشنهادی

سیستم‌های توصیه‌گر مبتنی بر پالایش گروهی با استفاده از تعاملات قبلی کاربران با سیستم اقدام به شناسایی و محاسبه شباهت بین کاربران کرده و بر اساس شباهت محاسبه‌شده، مواردی که احتمالاً موردعلاقه کاربر هستند را به وی پیشنهاد می‌دهند. از جمله مزایای سیستم‌های پالایش گروهی بی‌نیازی از جمع‌آوری و نگه‌داری اطلاعات خاص است. اما از مهم‌ترین معایب این روش می‌توان به مشکل شروع

مشابهی با او داشته‌اند در راستای ارائه پیشنهادها شخصی‌سازی‌شده برای کاربر هدف استفاده می‌کنند [۳].

ایده کلی در این روش این است که اگر کاربران  $x$  و  $y$  کالا-هایی را در گذشته به‌صورت مشابه نرخ‌گذاری کرده باشند، در آینده نیز رفتار مشابهی برای کالاهای دیگر از خود نشان می‌دهند [۳]. در این روش یک مجموعه داده<sup>۲</sup> از تاریخچه رفتار کاربران در گذشته موردنیاز است.

یکی از مشکلات روش پالایش گروهی شروع سرد<sup>۴</sup> است. این مشکل برای کاربران تازه‌وارد و یا کاربرانی که تعداد نرخ‌گذاری آن‌ها اندک است (سابقه کمی از آن‌ها در سیستم وجود دارد) پیش می‌آید که دسته‌بندی این نوع کاربران با توجه به غیرقابل‌پیش‌بینی بودن آن‌ها در یک گروه مناسب بسیار دشوار است [۴].

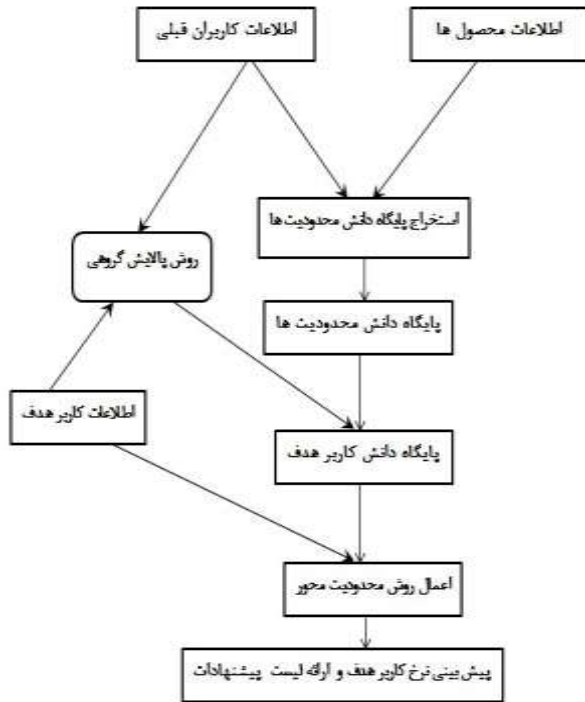
سیستم‌های توصیه‌گر مبتنی بر پالایش گروهی دو فاز یادگیری و پیش‌بینی دارند. با استفاده از یک مجموعه داده که نتیجه بازخورد کاربران با سیستم است (همان نرخ‌گذاری‌ها)، سیستم یاد می‌گیرد و در مرحله بعد پیش‌بینی می‌کند [۴].

یکی از مهم‌ترین مسائل در این زمینه کاری محاسبه شباهت بین دو کاربر است که با استفاده از نرخ‌گذاری کاربران روی کالاها و همچنین تعاملات دیگر کاربران با سیستم محاسبه می‌شود [۴].

در [۴] از یک روش پالایش گروهی جدید برای ارائه پیشنهادها به کاربران استفاده شده است. در اکثر روش‌های پالایش گروهی از متدهایی مانند ضریب پیرسون و روش‌های مبتنی بر کسینوس برای به دست آوردن میزان شباهت دو کاربر استفاده می‌شود. اما در این پژوهش فاصله کاربران محاسبه شده است. این فاصله با استفاده از هفت روش مختلف (فاصله اقلیدسی، فاصله جیبشیف، فاصله گاور، فاصله سورنسن، فاصله کانبرا، فاصله لورنتزین و فاصله سیتی‌بلاک) اندازه‌گیری شده و بررسی شده است. با توجه به فاصله محاسبه‌شده از دو الگوریتم نزدیک‌ترین همسایه و  $k$ -means برای مشخص شدن کاربران مشابه به کاربر موردنظر استفاده شده است که نتیجه هرکدام به‌طور جداگانه بررسی شده است. بعد از مشخص شدن همسایگان کاربر هدف سیستم نرخ کاربر برای کالاهایی که هنوز نرخ‌گذاری نکرده است با استفاده از الگوریتم رأی‌گیری به این صورت پیش‌بینی می‌شود که نرخ کاربر برای کالا  $i$  = محبوب‌ترین نرخ که همسایگان او به کالا  $i$  داده‌اند. این روش در نرم‌افزار MATLAB پیاده‌سازی شده و نسبت به روش پالایش گروهی محض نتایج بهتری از لحاظ خطای پیش‌بینی داشته است.

نیازهای کاربران ممکن است در طول زمان تغییر کند و این مسئله‌ای است که اکثر الگوریتم‌ها از آن غافل مانده‌اند. در [۵] یک سیستم توصیه‌گر با استفاده از روش پالایش گروهی ارائه شده است که با در نظر گرفتن عامل زمان و ترتیب خریده‌ها به کاربران پیشنهادهایی ارائه می‌کند. در واقع روش ارائه‌شده به آخرین تعاملاتی که کاربران با سیستم داشته‌اند اهمیت بیشتری داده است. نتایج حاصل از

آن است. در این روش لازم است خبره‌هایی در حوزه موردنظر تمامی دانش موردنیاز جهت ارائه پیشنهاد به کاربر هدف را در ابتدای پیاپی سازی سیستم تعریف کنند و در ادامه نیز آن را با تغییرات محیط به روزرسانی کنند [۱۰].



شکل ۱: طرح کلی سیستم ترکیبی از روش پالایش گروهی و محدودیت محور

در سیستم ترکیبی ارائه شده برای غلبه بر این مشکل اساسی، در ابتدای راه اندازی سیستم توصیه گر ترکیبی اقدام به استخراج پایگاه دانش محدودیت های سیستم شده است. برای استخراج محدودیت ها به صورت خودکار باید الگوی محدودیت ها را تعریف کرد تا بر اساس آن ها محدودیت ها استخراج گردد. هر الگوی محدودیت به صورت رابطه ای بین ویژگی های کاربران و کالاها بیان می شود و به صورت رابطه شرطی زیر نمایش داده می شود:

$$if < condition > then < consequent >$$

در این رابطه  $< Condition >$  شرطی است که به مقدار یک ویژگی یا ویژگی هایی از کاربران اشاره می کند و  $< Consequent >$  نیز به یک مقدار خاص از یک ویژگی از کالاها اشاره می کند.

بعد از تعریف کردن الگوهای محدودیت ها تعاملات قبلی کاربران را پیمایش کرده و بر اساس الگوهای تعریف شده محدودیت های مربوط به کاربر استخراج و ذخیره می شود.

اعمال روش پالایش گروهی: با توجه به حجم اطلاعات حاصل از تعاملات قبلی کاربران با سیستم، حجم پایگاه دانش محدودیت های تولید شده می تواند بسیار زیاد باشد. اگر این پایگاه دانش محدودیت با حجم زیاد به عنوان منبع ارائه پیشنهاد در روش محدودیت محور استفاده شود سیستم با دو مشکل عمده روبرو می شود: در مرحله اول به دلیل حجم بالای پایگاه دانش محدودیت ها ارائه پیشنهاد به هر

سرد برای کاربران و موارد تازه و همچنین پراکندگی داده ها، زمانی که حجم موارد نسبت به تعداد کاربران زیاد باشد اشاره کرد [۸].

سیستم های توصیه گر محدودیت محور یکی از انواع سیستم های توصیه گر دانش محور محسوب می شوند که برخلاف سیستم های پالایش گروهی جهت ارائه پیشنهادها به کاربران به اطلاعات و تعاملات قبلی کاربران نیازی ندارند و بر اساس ویژگی های موارد و نیازهای کاربران عمل می کنند و سعی در پیشنهاد موردی دارند که بیشترین تطابق را با نیازهای کاربر داشته باشد. به این ترتیب سیستم های محدودیت محور با مشکلاتی مانند شروع سرد و یا پراکندگی داده ها روبرو نیستند، اما در عوض نیاز به یک پایگاه دانش بزرگ و کامل دارند که جمع آوری آن امری دشوار و پرهزینه است [۸].

در سیستم توصیه گر ترکیبی پیشنهادی در این مقاله سعی شده است که با بهره بردن از مزایای سیستم های پالایش گروهی و محدودیت محور بتوان ضعف این سیستم ها را از بین برد و پیشنهادهایی بهتر و نزدیک تر به نیازهای کاربران ارائه داد.

در این نحوه ترکیب، ابتدا اطلاعات حاصل از تعاملات قبلی کاربران با سیستم مورد تجزیه و تحلیل قرار گرفته و سپس قواعد و دانش موردنیاز روش محدودیت محور شناسایی و استخراج می گردد. سپس با استفاده از روش پالایش گروهی محدودیت های خاص کاربر هدف استخراج شده و به عنوان پایگاه محدودیت شخصی شده کاربر هدف به روش محدودیت محور داده می شود. روش محدودیت محور نیز با استفاده از پایگاه دانش شخصی شده و مدل کاربر هدف و همچنین اطلاعات کالاها اقدام به ارائه پیشنهاد به کاربر هدف می کند.

طرح کلی روش پیشنهادی در شکل ۱ آمده است که شامل قسمت های زیر است:

پایگاه دانش محدودیت: پایگاه دانشی که محدودیت های سیستم در آن نگهداری می شود. محدودیت ها در واقع قوانینی هستند که ویژگی کاربران و موارد را به یکدیگر مرتبط می کند [۹].

اطلاعات کاربران قبلی: کاربرانی که با یک سیستم تعامل دارند اطلاعاتی را طی این تعامل ها به سیستم می دهند. این اطلاعات می تواند شامل ویژگی های شخصی کاربران مثل سن، قد، جنس و ... باشد یا اطلاعاتی در مورد بازخوردهایی که کاربر به گزینه های موجود در سیستم می دهد. به عنوان مثال کاربر A کالایی با شناسه i را در زمان t خریده و به آن امتیاز ۳ از ۵ داده است.

ویژگی های کاربران: هر کاربر دارای ویژگی ها و نیازمندی هایی است. برای مثال سن و جنسیت یک کاربر جز ویژگی های او محسوب می شود که در هنگام ثبت نام آن ها را به صورت صریح بیان می کند. همچنین کاربر می تواند نیازمندی های خود را به صورت صریح به سیستم اعلام کند. به عنوان مثال کاربر می تواند اعلام کند که به فیلم های با ژانر اکشن علاقه مند است.

استخراج پایگاه دانش محدودیت ها: یکی از مشکلات اساسی استفاده از روش محدودیت محور جمع آوری و به روز کردن پایگاه دانش

در الگوی بالا Age سن کاربر، Genre ژانر فیلم، Release Year سال انتشار فیلم، Gender جنسیت کاربر، Occupation شغل کاربر و Rate نیز متوسط نرخ گذاری کاربر بر روی فیلم‌هایی است که در قسمت شرایط ذکر شده‌اند.

الگوی محدودیت اول مربوط به سن کاربر و ژانر فیلم است که به صورت منطقی بین آن‌ها ارتباطی وجود دارد. به عنوان مثال کاربران با سنین پایین تر به ژانرهای خاصی علاقه دارند که کاربران سنین بالا به آن‌ها علاقه کمتری نشان می‌دهند.

الگوی دوم مربوط به سال انتشار و سن کاربر است. به عنوان مثال کاربران با سنین کم تر از ۳۰ سال علاقه کمتری به نرخ گذاری فیلم‌های قدیمی دارند.

الگوی سوم و چهارم هم مربوط به تأثیر جنسیت و شغل کاربران در انتخاب فیلم‌هایی با ژانر خاص است.

در مرحله بعد شباهت کاربر هدف با سایر کاربران سیستم را پیدا کرده و K تا از شبیه‌ترین کاربران به کاربر هدف را به عنوان کاربران همسایه کاربر هدف در نظر می‌گیریم. محدودیت‌های متناظر به کاربران همسایه به عنوان پایگاه دانش شخصی سازی شده کاربر هدف استخراج می‌شود.

هدف نهایی در حوزه فیلم این است که میزان نرخ گذاری کاربر هدف (Tu) بر روی فیلم مشخص i پیش‌بینی شود تا از طریق نرخ به دست آمده پیش‌بینی شود که کاربر به فیلم علاقه مند است یا خیر. در این مقاله از دو روش برای پیش‌بینی نرخ گذاری استفاده شده است که در ادامه بیان شده است.

### روش اول

محدودیت‌های قابل‌اعمالی که فیلم i را ارضا می‌کنند یا با ویژگی‌های آن منطبق هستند از میان پایگاه دانش محدودیت‌ها استخراج شده و با استفاده از متوسط نرخ محدودیت‌های استخراج شده پیش‌بینی انجام می‌شود. Nu مجموعه کاربران همسایه کاربر هدف، MSCR.nu متوسط نرخ محدودیت‌های ارضاشده کاربر همسایه nu و sim(nu,tu) شباهت کاربر هدف با کاربر همسایه‌اش است.

$$Tu.Rate(i) = \frac{\sum_{nu \in NU} nu.MSCR * sim(nu, Tu)}{\sum_{nu \in NU} sim(nu, Tu)} \quad (2)$$

متوسط نرخ محدودیت‌های ارضاشده هر کاربر همسایه از طریق رابطه زیر به دست می‌آید که در آن C مجموعه محدودیت‌های ارضاشده کاربر همسایه C.Rate.nu نرخ محدودیت ارضاشده و C.count تعداد تکرار محدودیت ارضاشده است.

$$nu.MSCR = \frac{\sum_{c \in C} C.Rate * C.Count}{\sum_{c \in C} C.count} \quad (3)$$

کاربر بسیار زمان بر و سنگین خواهد بود. در مرحله دوم پراکندگی و متفاوت بودن نیازهای کاربران از هم باعث می‌شود پیشنهادها غیربهبینه و نامناسب ارائه شود.

به همین دلیل قبل از استفاده از پایگاه دانش محدودیت‌های استخراج شده اقدام به پالایش آن برای کاربر هدف می‌شود و برای آن منظور از روش پالایش گروهی استفاده می‌شود. ابتدا شباهت کاربر هدف با سایر کاربران از رابطه زیر محاسبه می‌شود:

$$sim(u1, u2) = \frac{2 * |Pu1 \cap Pu2|}{|Pu1| + |Pu2|} \quad (1)$$

در رابطه ۱، Pu1 و Pu2 مجموعه ویژگی‌های کاربران u1 و u2 هستند. بعد از محاسبه فاکتور شباهت، شبیه‌ترین کاربران به کاربر هدف به عنوان همسایه آن انتخاب می‌شوند و تمام محدودیت‌های کاربران همسایه به عنوان پایگاه دانش شخصی سازی شده کاربر هدف در نظر گرفته می‌شود و به عنوان ورودی روش محدودیت‌محور استفاده می‌شود.

اعمال روش محدودیت‌محور: روش محدودیت‌محور با توجه به اطلاعات کاربر هدف، محدودیت‌های قابل‌اعمال از پایگاه دانش شخصی سازی شده را استخراج می‌کند. سپس همه کالاها را با محدودیت‌های قابل‌اعمال مطابقت می‌دهد. اگر کالای بتواند تمام محدودیت‌های قابل‌اعمال را ارضا کند به لیست پیشنهادها اضافه می‌شود. احتمال دارد هیچ کالایی تمام محدودیت‌ها را ارضا نکند که در این صورت کالای به لیست پیشنهادهایی اضافه می‌شود که کمترین محدودیت را نقض کرده باشد.

### ۶- پیاده‌سازی سیستم پیشنهادی

روش ارائه شده در این مقاله روی پایگاه اطلاعاتی movielens پیاده‌سازی شده است. سایت movielens متعلق به گروه تحقیقاتی Grouplens از دانشگاه مینیستوتا است [۳]. هر کاربر می‌تواند در این سایت ثبت نام کرده و در هنگام ثبت نام اطلاعاتی مانند سن، جنسیت، شغل و کد پستی را وارد کند. بعد از آن کاربران به فیلم‌ها امتیاز می‌دهند که این امتیاز میزان علاقه فرد به فیلم مورد نظر را نشان می‌دهد.

طبق الگوریتم پیشنهادی در ابتدا باید پایگاه دانش محدودیت‌های حاصل از تعاملات قبلی کاربران با سیستم استخراج گردد که این امر نیازمند تعریف الگوی محدودیت‌ها در حوزه فیلم است.

چهار الگوی استفاده شده در این روش برای استخراج محدودیت‌ها بر اساس منطق موجود در حوزه فیلم و بر اساس اطلاعاتی که از ویژگی‌های کاربران و فیلم‌ها در اختیار است عبارتند از:

- 1) **If (Age ^ Genre) Then Rate**
- 2) **If (Age ^ Release Year) Then Rate**
- 3) **If (Gender ^ Genre) Then Rate**
- 4) **If (Occupation ^ Genre) Then Rate**

## روش دوم:

استفاده از متوسط‌گیری که در روش اول مورد بررسی قرار گرفت، با توجه به تغییرات کمی که دارد نمی‌تواند به خوبی علایق کاربران را منعکس کند. برای رفع این مشکل نرخ‌گذاری قبلی کاربران شبیه به کاربر هدف در روابط تأثیر داده می‌شود.

$$Tu.Rate(i) = \frac{Tu.R + \sum_{nu \in NU} (nu.Rate(i) - nu.MSCR) * sim(nu, Tu)}{\sum_{nu \in NU} sim(nu, Tu)} \quad (4)$$

متوسط نرخ‌گذاری کاربر هدف بر روی فیلم‌هایی که محدودیت‌های متناظر به فیلم را ارضا می‌کند.  $Nu.Rate(i)$  نرخ‌دهی کاربر همسایه بر روی فیلم مورد نظر است.

## ۷- ارزیابی سیستم پیشنهادی

برای سنجش دقت و کارایی یک سیستم پیشنهاددهنده از دو نوع معیار می‌توان استفاده کرد. دسته اول معیارهای دقت آماری و دسته دیگر معیارهای دقت پشتیبانی تصمیم است.

### معیارهای دقت آماری

این معیارها بر روی مقایسه عددی نرخ‌های پیش‌بینی شده مجموعه داده با مقادیر واقعی وارد شده توسط کاربران متمرکز شده‌اند. در این پژوهش  $MAE^4$  از دسته معیارهای آماری انتخاب شده است که در رابطه زیر نشان داده شده است. در این رابطه  $P(i,j)$  نرخ پیش‌بینی شده و  $R(i,j)$  نرخ‌گذاری واقعی کاربر است. هر چقدر که مقدار  $MAE$  کم‌تر باشد دقت و کارایی روش بالاتر است [۱۰].

$$MAE = \frac{\sum_{(i,j)} P(i, j) - R(i, j)}{n} \quad (5)$$

### معیارهای دقت پشتیبانی تصمیم

این معیارها میزان مؤثر بودن یک الگوریتم پالایش گروهی را در کمک به کاربران برای انتخاب گزینه‌های باکیفیت بالا ارزیابی می‌کنند. از این دسته روشی به نام حساسیت  $ROC^9$  انتخاب شده است که در واقع معیار اندازه‌گیری قدرت تشخیص سیستم‌های پالایش گروهی است. این معیار بر این فرض استوار است که فرآیند پیش‌بینی یک فرآیند باینری است یعنی یک گزینه یا خوب است یا بد.

در این معیار چهار حالت زیر مورد بررسی قرار می‌گیرد:

مثبت درست ( $TP^{10}$ ): حالتی که یک گزینه به کاربر پیشنهاد شده و او هم آن را خریداری کرده است.

مثبت غلط ( $FP^{10}$ ): حالتی است که کاربر یک گزینه را خریداری کرده ولی سیستم آن را به او پیشنهاد نداده است.

منفی غلط ( $FN^{10}$ ): حالتی است که سیستم یک کالا را به کاربر پیشنهاد داده، اما کاربر آن را خریداری نکرده است.

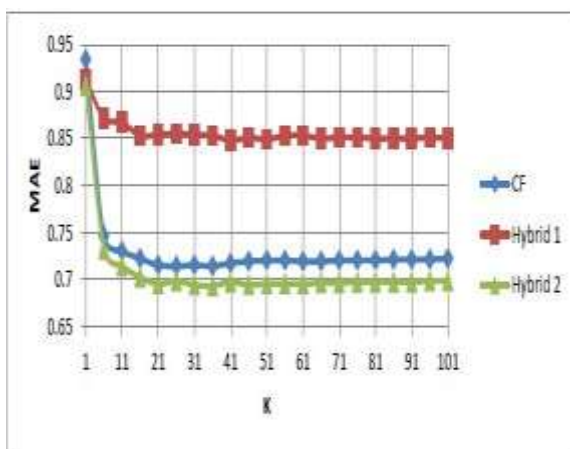
منفی درست ( $TN^{10}$ ): حالتی است که یک گزینه به کاربر پیشنهاد نشده و او هم آن را خریداری نکرده است.

در نتیجه دقت پشتیبانی تصمیم سیستم پیشنهادی از رابطه زیر محاسبه می‌شود:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

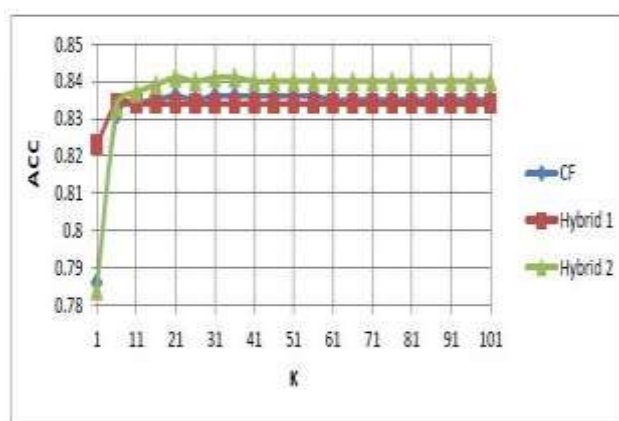
هرچه مقدار این پارامتر بیش‌تر باشد کارایی روش نیز بالاتر است [۱۱]. ارزیابی بر اساس تقسیم مجموعه داده به دو دسته آزمایش و آموزش انجام می‌شود. ۲۰ درصد داده‌ها به‌عنوان مجموعه آزمایش و ۸۰ درصد داده‌ها به‌عنوان مجموعه آموزش انتخاب شده‌اند. نتایج به‌دست‌آمده از ارزیابی سیستم توصیه‌گر ترکیبی پیشنهادی در زیر آمده است. در نمودارها  $k$  تعداد همسایه‌های کاربر هدف،  $CF$  روش پالایش گروهی حساس به زمان  $[\Delta]$ ،  $Hybrid1$  روش ترکیبی پیشنهادی اول و  $Hybrid2$  روش ترکیبی پیشنهادی دوم است.  $MAE$  معیار دقت آماری و  $ACC$  معیار دقت پشتیبانی تصمیم سیستم می‌باشند.

همان‌طور که در شکل‌های ۲ و ۳ مشاهده می‌شود افزایش تعداد همسایه‌های کاربر هدف ( $k$ ) تا ۲۰ موجب افزایش کارایی در تمام روش‌ها می‌شود ولی از ۲۰ همسایه بیش‌تر تأثیری در کارایی مشاهده نمی‌شود. کارایی هر سه روش به ازای  $k$ های کم‌تر از ۱۰ پایین است. هرچه تعداد همسایه‌ها پایین‌تر باشد احتمال تغییر کاربر هدف نسبت به همسایه‌ها نیز بیش‌تر می‌شود. در مقابل اگر  $k$  بیش‌تر از ۱۱ باشد پیش‌بینی علاقه کاربر هدف به فیلم مورد نظر با دقت بیش‌تری انجام می‌شود. کارایی روش ترکیبی اول ( $Hybrid1$ ) نسبت به روش پالایش گروهی ( $CF$ ) کم‌تر است. علت این امر در نحوه پیش‌بینی نرخ‌ها با استفاده از متوسط‌گیری بر روی محدودیت‌های ارضاشده در روش ترکیبی اول است. در عوض روش ترکیبی اول مشکل شروع سرد روش پالایش گروهی را ندارد و محاسبه شباهت کاربران و پیش‌بینی نرخ فیلم‌های جدید نیازمند تعاملات قبلی کاربر نیست.



شکل ۲: نتایج ارزیابی روش‌های ترکیبی و پالایش گروهی با استفاده از معیار MAE

- [2] D. Jannach, M. Zanker, A. Felfering and G. Friedrich, *Recommender System An Introduction*: Cambridge University Press, 2011.
- [3] A. Bahrehmand and R. Rafeh, "Proposing a news metric for collaborative filtering," *Software Engineering and Applications*, pp. 411-416, 2011.
- [4] G. M. Dakhel and M. Mahdavi, "Providing an effective collaborative filtering algorithm based on distance measures and neighbors' voting," *Computer Information Systems and Industrial Management Application*, vol. 3, pp. 241-251, 2013.
- [5] R. Rafeh and A. Bahrehmand, "An adaptive approach to deal with unstable behavior of users in collaborative filtering system," *Information Science*, vol. 3, pp. 205-223, 2012.
- [6] A. Felfering, G. Friedrich, D. jannach and M. Zanker, "Developing constraint-based recommenders," in: *Recommender System Handbook*, Springer, pp. 187-226, 2010.
- [7] M. Zanker, M. Aschinger and M. Jessenitsching, "Constraint-based personalized configuring of product and servic bunndles," *Mass Customization*, 2010.
- [8] M. Zanker, "A Collaborative constraint-based meta level recommender," *RecSys*, vol. 12, pp. 139-145, 2008.
- [9] A. Felfering and R. Burke, "Constraint-based recommender systems: technologies and research issues," *ACM*, vol.4, pp. 147-157, 2008.
- [10] N. Lathia, S. Hailes and L. Capra, "Evaluating collaborative filtering over time," *Advances in Artificial intelligence*, vol. 2, pp. 121-125, 2009.
- [11] J. L. Herlocker and J. A. Konstan, "Evaluating hybrid recommender systems," *Information Systems*, vol. 22, pp. 39-53, 2004.



شکل ۳: نتایج ارزیابی روش‌های ترکیبی و پالایش گروهی با استفاده از معیار ACC

در روش ترکیبی دوم (Hybrid2) از نرخ‌گذاری قبلی کاربران همسایه بر روی فیلم مورد پیش‌بینی استفاده شده است و با استفاده از متوسط نرخ محدودیت‌های ارضاشده این مقدار نرمال شده و بنابراین نسبت به روش پالایش گروهی همان‌طور که در شکل مشخص شده، کارایی بیشتری دارد و مشکل شروع سرد نیز ندارد. در واقع نتایج نشان می‌دهند که اگر در سیستم توصیه‌گر از ویژگی‌های کاربران و کالاها در امر پیشنهاددهی استفاده شود عملکرد سیستم بهتر می‌شود.

## ۸- نتیجه‌گیری و ارائه پیشنهادها برای آینده

هدف اصلی در سیستم‌های توصیه‌گر پالایش حجم وسیع اطلاعات کاربران و کالاها و ارائه پیشنهادها متناسب با علائق کاربر هدف است. در این مقاله با استفاده از ترکیب روش پالایش گروهی و محدودیت‌محور کارایی سیستم‌های توصیه‌گر بهبود داده شده است. سیستم ارائه‌شده با استفاده از پایگاه دانش استخراج‌شده از تعاملات کاربران همسایه کاربر هدف با سیستم، مشکل شروع سرد روش پالایش گروهی را ندارد. ارزیابی‌ها نیز نشان داد که سیستم ترکیبی عملکرد بهتری نسبت به روش پالایش گروهی دارد.

در آینده می‌توان برحسب نیاز حوزه‌های مختلف روش‌های دیگر را باهم ترکیب کرد تا کارایی سیستم در حوزه موردنظر بیشتر شود. مثلاً با ترکیب روش‌های پالایش گروهی و استفاده از قواعد معنایی برای پیدا کردن روابط معنادار کالاها و کاربران دقت سیستم‌ها را بالا برد و با مشکل شروع سرد نیز مقابله کرد. همچنین قصد داریم در آینده فاکتور زمان را برای دقیق‌تر شدن پیشنهادها در محاسبه شباهت کاربران دخالت دهیم.

## مراجع

- [1] 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*, vol. 17, pp. 734-749, 2005.

## زیرنویس‌ها

- <sup>۱</sup> E-Commerce
- <sup>۲</sup> Recommender Systems
- <sup>۳</sup> Data Set
- <sup>۴</sup> Cold Start
- <sup>۵</sup> Constraint based
- <sup>۶</sup> Case based
- <sup>۷</sup> Collaborative Filtering
- <sup>۸</sup> Mean Absolute Error
- <sup>۹</sup> Receiver Operating Characteristic
- <sup>۱۰</sup> True Positive
- <sup>۱۱</sup> False Positive
- <sup>۱۲</sup> False Negative
- <sup>۱۳</sup> True Negative
- <sup>۱۴</sup> Accuracy