بهره برداری از ژانرهای فیلم و اطلاعات جمعیت‌شناختی کاربران به‌منظور بهبود سامانه‌های توصیهٔ فیلم

صمد محمدی، واهه آغازاریان* و علیرضا هدایتی

گروه مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران

چکیده

سامانه‌های توصیه فیلم ابزارهای کارآمدی هستند که به کاربران کمک می‌کنند فیلم‌های مورد علاقه خود را با بررسی علایق قبلی کاربران پیدا کنند. این سامانه‌ها بر اساس امتیاز کاربران به فیلم‌های گذشته و استفاده از آنها برای پیش‌بینی علایق آنها در آینده ایجاد شده‌اند؛ با این حال، امتیازدهی نامناسبی که کاربران ارائه می‌دهند، منجر به ایجاد مشکلی به نام پراکندگی داده می‌شود. این مشکل موجب کاهش کارایی سامانه‌های توصیه فیلم می‌شود. از سوی دیگر، سایر داده‌های موجود مانند ژانر فیلم‌ها و اطلاعات جمعیت‌شناختی کاربران، نقش حیاتی در کمک به روش‌های توصیه‌کننده برای تولید بهتر توصیه‌ها دارند. این مقاله یک روش توصیه فیلم را با استفاده از ژانرهای فیلم و اطلاعات جمعیت‌شناختی کاربران پیشنهاد می‌کند. همچنین ما مدلی کارآمد جهت ارزیابی پروفایل امتیازدهی کاربر و تعیین کمینه امتیاز مورد نیاز برای تولید یک پیش‌بینی دقیق را پیشنهاد می‌کنیم؛ سپس، امتیازات مجازی مناسب با پروفایل‌هایی که امتیازات نامناسبی دارند ترکیب می‌شوند. این امتیازدهی مجازی با استفاده از شباهت مقادیر بین کاربران به‌دست‌آمده از ژانرهای فیلم و اطلاعات جمعیت‌شناختی کاربران محاسبه می‌شوند؛ علاوه بر این، یک معیار مفید برای تعیین میزان قابل‌اعتماد بودن یک بخش معرفی شده‌است که قابلیت اطمینان امتیازدهی مجازی را تضمین می‌کند؛ درنهایت، امتیازهای ناشناخته برای کاربر هدف براساس پروفایل‌های امتیازدهی توسعه‌یافته پیش‌بینی می‌شوند. آزمایش‌های انجام‌شده بر روی دو مجموعه‌داده توصیه فیلم معروف نشان‌می‌دهد که رویکرد پیشنهادی کارآمدتر از سایر توصیه‌کننده‌های مقایسه شده‌است.

واژگان کلیدی: سامانه‌های توصیه، فیلم، پراکندگی داده‌ها، اطلاعات جمعیت‌شناختی، ژانر

Exploiting movie genres and user demographic information to improve movie recommendation systems

Samad Mohamadi, Vahe Aghazarian* and Alireza Hedayati
Department of Computer Engineering, Islamic Azad University,
Central Tehran Branch, Tehran, Iran

Abstract

Movie recommendation systems are efficient tools to help users find their relevant movies by investigating the previous interests of users. These systems are established on considering the ratings of users provided for movies in the past and using them to predict their interests in the future. However, users mainly provide insufficient ratings leading to make a problem called data sparsity. This problem makes reducing the effectiveness of movie recommendation systems. On the other hand, other available data such as genres of movies and demographic information of users play a vital role in assisting recommenders in order to better produce recommendations. This paper proposes a movie recommendation method utilizing the movies' genres and users' demographic information. In particular, we propose an effective model to evaluate the user's rating profile and determine the minimum number of ratings required to produce an accurate prediction. Then, appropriate virtual

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۴۰۱/۵/۱۹ • تاریخ پذیرش: ۱۴۰۲/۳/۱۲ • تاریخ انتشار: ۱۴۰۲/۱۲/۲۹ • نوع مطالعه: پژوهشی

[DOI: 10.61186/jsdp.20.4.89]

ratings are incorporated into the profiles with insufficient ratings to expand them. These virtual ratings are calculated using similarity values between users obtained by genres of movies and demographic information of users. Furthermore, an effective measure is introduced to determine how much an item is reliable. This measure guarantees the virtual ratings' reliability. Finally, unknown ratings for target user are predicted based on the expanded rating profiles. Experiments performed on two well-known movie recommendation datasets demonstrate that the proposed approach is more efficient than other compared recommenders.

We propose a movie recommender system in this paper by employing the genres of movies and demographic information of users to address the above-mentioned challenges. To this end, first of all, a model is developed in order to determine whether the target user's rating profile is appropriate to produce accurate recommendations or not. In other words, the developed model determines how many ratings are required for each user to generate an accurate prediction with a high probability. This criterion is used to demonstrate that a rating profile contains sufficient ratings for producing reliable recommendations or not. Then, the quality of rating profiles containing insufficient ratings is boosted using an effective profile expansion technique which incorporates some virtual ratings to these profiles. These virtual ratings are calculated using the similarity values between users which are computed according to the genres of movies and demographic information of users. Moreover, the reliability values of users and items are calculated using appropriate reliability measurements to guarantee that the incorporated virtual ratings are reliable. Experimental results on two movie recommendation datasets indicate the superiority of the proposed approach in respect to other models. In the following, we provide a list of the main contributions of this paper:

- **We develop a model in order to evaluate the users' rating profiles and determine how many ratings are required for generating an accurate prediction.**
- **We propose a powerful profile expansion technique which incorporates some virtual ratings to user-item ratings matrix for improving its quality.**
- **Movies' genres and users' demographic information are used as additional data in the proposed movie recommender system.**
- **The reliability measures of users and items are used in the proposed method to guarantee the reliability of calculated virtual ratings.**
- **The proposed method generates a denser user-item ratings matrix than the original matrix which results in alleviating data sparsity problem significantly.**

The remaining parts of this paper are structured as follows: in section 2, related works are investigated, section 3 includes the details of the proposed method, section 4 refers to the discussion of experimental results, and section 5 provides some conclusions about the paper Movie recommendation systems are efficient tools to help users find their relevant movies by investigating the previous interests of users. These systems are established on considering the ratings of users provided for movies in the past and using them to predict their interests in the future. However, users mainly provide insufficient ratings leading to make a problem called data sparsity. This problem makes reducing the effectiveness of movie recommendation systems. On the other hand, other available data such as genres of movies and demographic information of users play a vital role in assisting recommenders in order to better produce recommendations. This paper proposes a movie recommendation method utilizing the movies' genres and users' demographic information. In particular, we propose an effective model to evaluate the user's rating profile and determine the minimum number of ratings required to produce an accurate prediction. Then, appropriate virtual ratings are incorporated into the profiles with insufficient ratings to expand them. These virtual ratings are calculated using similarity values between users obtained by genres of movies and demographic information of users. Furthermore, an effective measure is introduced to determine how much an item is reliable. This measure guarantees the virtual ratings' reliability. Finally, unknown ratings for target user are predicted based on the expanded rating profiles. Experiments performed on two well-known movie recommendation datasets demonstrate that the proposed approach is more efficient than other compared recommenders.

Keywords*:* Recommendation systems, movie, data sparsity, demographic information, genre

۱- معرفی

در سال‌های اخیر تارنماهای مختلف تجاری و غیر تجاری حجم زیادی از اطلاعات و محصولات را در اینترنت منتشر کرده‌اند. این موضوع باعث می‌شود، کاربران جهت دست‌یابی به اطلاعات مناسب با حجم زیادی از اطلاعات روبه‌رو شوند که این موضوع موجب سردرگمی کاربران می‌شود.

سامانه‌های توصیه، با توصیه اطلاعات موردعلاقه کاربران به‌منظور یافتن اطلاعات موردنیاز آن‌ها نقش مهمی را ایفا می‌کنند [۱، ۲]. هدف اصلی این سامانه‌ها بررسی رفتار کاربران در گذشته و استخراج دانش مناسب برای پیش‌بینی علایق آن‌ها در آینده است. به‌طور کلی، سامانه‌های توصیه برای رسیدن به دو هدف اصلی تلاش می‌کنند؛ نخستین هدف، جلب رضایت کاربران در سامانه

[DOI: 10.61186/jsdp.20.4.89]

با ارائه مرتبط‌ترین توصیه‌ها به آن‌ها است؛ زیرا کاربران مایل هستند به‌طور خودکار مرتبط‌ترین توصیه‌ها را از سامانه دریافت تا از اتلاف وقت خود جلوگیری کنند. هدف دوم، افزایش سطح رضایت صاحبان تجارت الکترونیک با افزایش میزان سود از طریق فروش متقابل است [۳، ۴].

اصولاً روش‌های توصیه را می‌توان بر اساس روشی که برای ایجاد توصیه‌ها برای کاربران استفاده می‌کنند به سه رویکرد اصلی دسته‌بندی کرد. این رویکردها شامل فیلترینگ مشارکتی[5] (CF) ، رویکرد مبتنی بر محتوا [۶] و مدل‌های ترکیبی [۷] است .CF یک رویکرد محبوب است که در طراحی توصیه‌کننده‌ها استفاده می‌شود که ایده اصلی آن استفاده از داده‌های امتیازدهی برای پیش‌بینی ترجیحات کاربران است. این رویکرد را می‌توان در دو سناریوی مختلف از جمله CF مبتنی بر حافظه و CF مبتنی بر مدل انجام داد. در CF مبتنی بر حافظه اندازه‌گیری شباهت‌ها مثل ضریب همبستگی پیرسون، کسینوس و جاکارد برای تعیین مقدار شباهت بین کاربران / اقلام و تشکیل مجموعه‌ای از نزدیک‌ترین همسایگان برای کاربر هدف استفاده می‌شود؛ سپس، رتبه‌بندی‌های مجهول با توجه به ترجیحات همسایگان کاربر هدف پیش‌بینی می‌شوند. مدل مبتنی برCF از روش‌های یادگیری ماشین برای ساخت مدلی به‌منظور پیش‌بینی امتیازدهی‌های نامشخص کاربران استفاده می‌کند. مدل‌های یادگیری مختلفی در سامانه‌های توصیه‌ای از جمله رویکردهای خوشه‌بندی، عامل‌بندی ماتریس، تجزیه ارزش منفرد و غیره استفاده شده‌است [۸-۱۰]. مدل‌های مبتنی بر محتوا از محتوای اطلاعات بخش‌ها در فرآیند توصیه استفاده می‌کنند. محتوای اطلاعات به محصولات استفاده‌شده در سامانه توصیه مرتبط است.. برای مثال، در یک سامانه توصیه فیلم، محتوای اطلاعات می‌تواند ژانرهای فیلم‌های ارائه‌شده در سامانه باشد. مدل‌های ترکیبی، ویژگی‌های دو یا چند مدل دیگر را برای ارائه توصیه‌های دقیق‌تری ترکیب می‌کنند.

یکی از منابع اصلی مورد استفاده در سامانه‌های توصیه، ماتریس امتیازدهی کاربر بخش است. کارایی سامانه‌های توصیه به‌طور مستقیم به کیفیت ماتریس امتیازدهی کاربر بخش بستگی دارد. به‌عبارت دیگر، اگر کاربران امتیازهای کافی را در سامانه ارائه دهند، منجر به تولید توصیه‌های دقیق‌تری می‌شود [۱۱، ۱۲]. با این حال، کاربران اغلب تمایل دارند که امتیازها را به چند بخش تخصیص دهند درحالی‌که به‌طورمعمول بخش‌های بسیار زیادی در سامانه وجود دارد. این موضوع به‌عنوان یکی از رایج‌ترین مشکلات در سامانه‌های توصیه شناخته می‌شود که به آن پراکندگی داده می‌گویند. گفتنی است که مشکل پراکندگی داده‌ها، تأثیر زیادی در کاهش اثربخشی سامانه‌های توصیه دارد. به‌منظور کاهش مشکل پراکندگی داده‌ها، پژوهش‌های زیادی با هدف توسعه رویکردهای مناسب انجام شده‌است [۱۳-۱۵]. یکی از ایده‌های اصلی این رویکردها استفاده از داده‌هایی مانند روابط اعتماد بین کاربران [۱۶، ۱۷]، اطلاعات برچسب [۱۸] و بررسی اطلاعات [۱۹] در فرآیند توصیه است. در یک سامانه توصیه فیلم، برخی از داده‌های اضافی مربوط به فیلم‌ها و کاربران در سامانه وجود دارد که می‌توان از آنها برای مقابله با مشکل پراکندگی داده‌ها استفاده‌کرد. این داده‌ها شامل ژانر فیلم‌ها و اطلاعات جمعیت‌شناختی کاربران است. با این وجود، نحوه به‌کارگیری چنین داده‌های اضافی در سامانه‌های توصیه فیلم یک مسئله چالش‌برانگیز است [۲۰، ۲۱].

ما در این مقاله یک سامانه توصیه‌کننده فیلم را با استفاده از ژانر فیلم‌ها و اطلاعات جمعیت‌شناختی کاربران برای رفع چالش‌های یادشده پیشنهاد می‌کنیم. در مرحله نخست، مدلی برای تعیین اینکه آیا پروفایل امتیازدهی کاربر هدف برای ارائه توصیه‌های دقیق مناسب است یا خیر، طراحی شده‌است. به‌عبارت دیگر، این مدل تعیین می‌کند که برای هر کاربر چه میزان امتیازی لازم است تا یک پیش‌بینی دقیق با احتمال بالا ایجاد کند. این معیار، یک پروفایل امتیازدهی را جهت امتیازهای کافی برای ارائه توصیه‌های قابل‌اعتماد بررسی می‌کند؛ سپس، کیفیت پروفایل‌های امتیازدهی حاوی امتیازهای ناکافی با استفاده از یک روش مؤثر گسترش پروفایل که برخی امتیازهای مجازی را به این پروفایل‌ها اضافه می‌کند، افزایش می‌یابد. این امتیازدهی مجازی براساس شباهت مقادیر بین کاربران با توجه به ژانرهای فیلم و اطلاعات جمعیت‌شناختی کاربران محاسبه و علاوه‌بر این، مقادیر قابلیت اطمینان کاربران و بخش‌ها با استفاده از اندازه‌گیری قابلیت اطمینان مناسب محاسبه می‌شوند تا اطمینان حاصل شود که امتیازدهی‌های مجازی گنجانده‌شده قابل‌اعتماد هستند. نتایج آزمایش‌ها روی دو مجموعه‌داده پیشنهادی فیلم نشان‌دهنده برتری رویکرد پیشنهادی نسبت به مدل‌های دیگر است. در

[DOI: 10.61186/jsdp.20.4.89]

ادامه، فهرستی از دستاوردهای اصلی این مقاله را ارائه می‌کنیم:

ما مدلی را به‌منظور ارزیابی پروفایل‌های امتیازدهی کاربران و تعیین میزان امتیازها برای ایجاد یک پیش‌بینی دقیق را ایجاد می‌کنیم.

ما یک روش قدرتمند توسعه پروفایل را پیشنهاد می‌کنیم که برخی از امتیازدهی‌های مجازی را با ماتریس امتیازدهی کاربر-اقلام ترکیب می‌کند. ژانر فیلم‌ها و اطلاعات جمعیت‌شناختی کاربران به‌عنوان داده‌های اضافی در سامانه پیشنهادی فیلم استفاده می‌شود. در روش پیشنهادی، معیارهای قابلیت اطمینان کاربران و بخش‌ها برای تضمین قابلیت اطمینان امتیازدهی مجازی محاسبه‌شده استفاده می‌شود.

روش پیشنهادی یک ماتریس امتیازدهی بخش کاربر را که متراکم‌تر از ماتریس اصلی است، ایجاد می‌کند؛ که این ماتریس منجر به کاهش قابل توجه مشکل پراکندگی داده‌ها می‌شود.

ساختار بخش‌های باقی‌مانده این مقاله به شرح زیر است: در بخش ۲، کارهای مرتبط مورد بررسی قرار می‌گیرند؛ بخش ۳ شامل جزئیات روش پیشنهادی است؛ بخش ۴ به بحث در مورد نتایج آزمایش‌ها اشاره و بخش ۵ نتیجه‌گیری‌هایی در مورد مقاله ارائه می‌کند.

۲- آثار مرتبط

پژوهش‌های زیادی برای توسعه سامانه‌های توصیه‌کننده فیلم انجام شده‌است [۲۲-۲۴]. در [۲۲] یک روش توصیه فیلم ترکیبی با استفاده از فیلتر مشترک، فیلتر مبتنی بر محتوا، و یک سامانه خبره فازی برای ارائه فهرستی از فیلم‌های مرتبط برای کاربران پیشنهاد شده‌است. در [۲۳]، یک شبکه عصبی رمزگذار عمیق برای توسعه یک رویکرد توصیه فیلم اجتماعی استفاده شده‌است. این سامانه بر اساس یک مدل ترکیبی بنا شده‌است که از فیلتر مشارکتی، فیلتر مبتنی بر محتوا و اطلاعات اجتماعی کاربران استفاده می‌کند. در [۲۴]، نویسندگان بر روی بهبود یک سامانه توصیه‌کننده فیلم براساس خوشه‌بندی داده‌ها و مفاهیم هوش‌مصنوعی محاسباتی تمرکز کردند. برای این منظور، الگوریتم خوشه‌بندی k-means با رویکرد بهینه‌سازی جستجوی فاخته بهبودیافته است. ایندیرا و همکاران [۲۵] برخی از الگوریتم‌های یادگیری ماشین را برای پیشنهاد یک روش توصیه فیلم مؤثر به کار گرفت. در این روش در مرحله نخست، نوفه‌های موجود در مجموعه‌داده ورودی حذف می‌شود که این عملیات منجر به ایجاد یک مجموعه‌داده قابل‌اعتمادتر، سپس، یک رویکرد انتخاب ویژگی مبتنی بر مدل تحلیل مؤلفه‌های اصلی برای انتخاب زیرمجموعه مناسبی از ویژگی‌ها برای اجرای رویکرد خوشه‌بندی k-means استفاده می‌شود. درنهایت، فهرستی از فیلم‌های مرتبط برای هر کاربر بر اساس خوشه‌های به‌دست‌آمده تولید می‌شود. جان و همکاران [۲۶] نشان داد که ترجیحات کاربران و ویژگی‌های متوالی آنها نقش مهمی در افزایش دقت سامانه‌های توصیه‌کننده فیلم دارد؛ بنابراین، آنها با در نظر گرفتن سه شاخص برای توصیف الگوهای مختلف ترجیحات کاربران، مدلی به نام UMIS پیشنهاد کرده‌اند؛ علاوه بر این، یک رویکرد یادگیری عمیق به مدل UMIS برای تولید توصیه‌های فیلم اعمال شده‌است. در [۲۷]، یک روش توصیه فیلم بر اساس یک پروفایل مثبت کاربر و یک پروفایل منفی کاربر پیشنهاد شده‌است که پروفایل مثبت حاوی مجموعه فیلم‌هایی است که کاربر دوست دارد و پروفایل منفی حاوی مجموعه فیلم‌هایی است که کاربر دوست ندارد. هدف اصلی این روش تهیه مجموعه‌ای از توصیه‌هایی برای هر کاربر است که در آن فیلم‌ها بیشترین شباهت را با پروفایل مثبت و همچنین بیشترین تفاوت را با پروفایل منفی دارند. روی و همکاران [۲۸] به حل مشکل شروع سرد مربوط به فیلم‌های جدید پرداخت که امتیاز تخصیص داده‌شده کافی برای آنها وجود ندارد. برای این منظور، آنها یک رویکرد توصیه فیلم ترکیبی را با استفاده از ژانر فیلم‌ها برای محاسبه مقادیر شباهت بین فیلم‌ها و ارائه توصیه‌هایی برای کاربران پیشنهاد کرده‌اند. اورتگا و همکاران [۲۹] یک روش مبتنی بر فاکتورگیری ماتریس را برای سامانه‌های توصیه‌گر معرفی کرده‌اند، که قادر است نه‌تنها مقادیر پیش‌بینی‌ها را بلکه می‌تواند مقادیر قابلیت اطمینان را نیز محاسبه کند. [۳۰] به‌منظور ارزیابی اعتماد شبکه‌های کاربران و شناسایی کاربران ناکارآمدی که تأثیر منفی بر محاسبه امتیازهای ناشناخته دارند، مفهوم قابلیت اطمینان را در یک سامانه توصیه‌گر آگاه به اعتماد افزوده‌است. در [۳۱]، از دو مؤلفه کارآمد قابلیت اطمینان کاربران و تأثیر انتشار برای ایجاد یک مدل ترجیح اجتماعی برای سامانه‌های توصیه‌گر استفاده می‌شود. [۳۲] با بررسی تضادهای احتمالی در پروفایل‌های همسایگان کاربر هدف، یک معیار قابلیت اطمینان را به‌منظور محاسبه قابلیت اطمینان پروفایل‌های کاربران ارائه داده است. در این روش، علایق

[DOI: 10.61186/jsdp.20.4.89]

کاربران با استفاده از الگوریتم خوشه‌بندیk-means برای یافتن پروفایل‌های غیرقابل‌اعتماد در همسایگی کاربران گروه‌بندی می‌شوند. جیانگ و همکاران [۳۳] از یک رویکرد تصادفی‌هایپرگراف با درنظرگرفتن مدل‌های وزن‌دهی روی یال‌ها و رئوس برای ارائه توصیه‌هایی از طریق یک فرآیند امتیازدهی استفاده‌کرد؛ علاوه بر این، روابط اعتماد بین کاربران و معیار قابلیت اطمینان برای ساختن‌هایپرگراف استفاده می‌شود که منجر به بهبود دقت پیش‌بینی می‌شود. در [۳۴]، نویسندگان به حملات شیلینگ در سامانه‌های توصیه‌گر، با پیشنهاد یک رویکرد فاکتورگیری ماتریس بر اساس معیار قابلیت اطمینان که نشان‌می‌دهد یک پیش‌بینی تا چه حد قابل‌اعتماد است، پرداخته‌اند. بر این اساس، حملات شیلینگ را می‌توان با بررسی امتیازهای پیش‌بینی‌شده که مقادیر قابلیت اطمینان آنها غیرعادی است، شناسایی کرد. نویسندگان این مقاله به‌طور تجربی نشان داده‌اند که مدل پیشنهادی آنها قادر به شناسایی تعداد زیادی از حملات موجود در سامانه‌های توصیه‌گر است.

۳- روش پیشنهادی

در این بخش، جزئیات روش پیشنهادی توصیه فیلم که مبتنی بر روش گسترش پروفایل مبتنی بر ژانر و اطلاعات جمعیت‌شناختی است، ارائه شده‌است. ما روش پیشنهادی را به‌عنوان GDPE[1] نام‌گذاری می‌کنیم که شامل پنج مرحله است: (۱) ارزیابی پروفایل امتیازدهی (۲) محاسبه قابلیت اطمینان کاربران (۳) محاسبه قابلیت اطمینان‌بخش‌ها (۴) توسعه پروفایل امتیازدهی و (۵) ارائه توصیه‌ها. یک طرح کلی از روش GDPE پیشنهادی در بخش‌های فرعی زیر بحث GDPE در شکل (۱) ارائه شده و جزئیات هر مرحله از مدل پیشنهادی در بخش‌های زیر بیان شده‌است.

۳-۱- ارزیابی پروفایل امتیازدهی

تعداد امتیازهای موجود در پروفایل امتیازدهی کاربر، نقش مهمی‌در تولید پیش‌بینی‌های دقیق برای فیلم‌های دیده‌نشده دارد. عملکرد سامانه‌های توصیه‌گر به امتیازهای کافی موجود در پروفایل‌های امتیازدهی کاربران بستگی دارد. بر این اساس، باید دریافت که آیا تعداد امتیازها در پروفایل امتیازدهی برای پیش‌بینی دقیق کافی است یا خیر. در این بخش، یک روش کارآمد برای بررسی پروفایل امتیازدهی کاربر بر اساس امتیازهای تخصیص‌یافته به فیلم‌ها بیان شده‌است. همچنین این روش تعیین می‌کند که برای پیش‌بینی دقیق امتیازهای ناشناخته چه میزان امتیاز لازم است؛ بنابراین، تعداد امتیازها در یک پروفایل امتیازدهی با تعداد امتیازهای مورد نیاز مقایسه می‌شود و اگر تعداد امتیازهای موجود کمتر از حداقل مقدار محاسبه‌شده برای یک پروفایل امتیازدهی باشد، به‌عنوان یک پروفایل امتیازدهی غیرقابل‌اعتماد در نظر گرفته‌می‌شود که باید با روش گسترش پروفایل امتیازدهی پیشنهادی بهبود یابد.

فرض کنید که U و I به‌ترتیب مجموعه کاربران و بخش‌های تنظیم‌شده در سامانه را نشان می‌دهند. همچنین امتیازهایی که کاربران به فیلم‌ها اختصاص می‌دهند در محدوده [minR, maxR] است. احتمال تخصیص یک امتیاز r∈ [minR, maxR] توسط کاربر u به‌صورت زیر تعریف می‌شود:

$$\alpha_{u,r} = \Pr[r_{u,i} = r] = \frac{n_{u,r}}{|I_u|} \qquad (۱)$$

که در آن $r_{u,i}$ امتیاز فیلم i است که توسط کاربر u ساخته شده‌است، تعداد امتیازهای اختصاص داده‌شده توسط کاربر u در سطح r با $n_{u,r}$ نشان داده‌می‌شود، و I_u به مواردی اشاره دارد که توسط کاربر امتیازدهی شده‌اند. برای محاسبه امتیاز واقعی ارائه‌شده توسط کاربرu ، قانون بیشینه به‌صورت زیر تعریف می‌شود:

$$l_u = \max_{r\in[minR,maxR]}\{\alpha_{u,r}\} \qquad (۲)$$

که امتیاز درست برای کاربر u توسط l_u نشان داده‌شده‌است؛ سپس طبق رابطه (۳) می‌توان کمترین مقدار امتیازهایی را که برای پیش‌گویی درست برای کاربر u مورد نیاز است، محاسبه کرد.

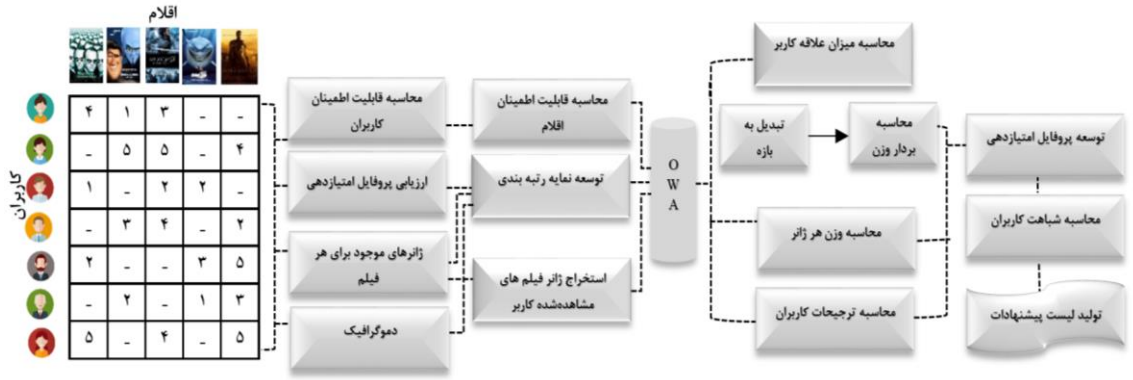
$$n'_u = \left(2\left(\alpha_{u,l_u} + \tilde{\alpha}_u\right)\left(\alpha_{u,l_u} - \tilde{\alpha}_u\right)^{-2} - 2 + \frac{4\left(\alpha_{u,l_u} - \tilde{\alpha}_u\right)^{-1}}{3}\right) \times \ln(maxR - minR)\delta^{-1} \qquad (۳)$$

در رابطه بالا σ یک مقدار آستانه از قبل تعریف شده‌است که میزان قابلیت اطمینان را در تولید یک پیش‌گویی درست تعیین می‌کند و دومین بزرگ‌ترین مقدار $\alpha_{u,r}$ توسط $\tilde{\alpha_u}$ نشان داده شده که به‌صورت زیر به‌دست‌آمده است:

$$\tilde{\alpha}_u = \max_{r\in[minR,maxR]}\{\alpha_{u,r} | r \neq l_u\} \qquad (۴)$$

[1] Genre and Demographic information based Profile Expansion

[DOI: 10.61186/jsdp.20.4.89]



(شکل-۱) طرح کلی مدل پیشنهادی GDPE

(Figure -1). Outline of the proposed GDPE model

با درنظرگرفتن رابطه ۳، اگر $|I_u| \geq n'_u$ ، سپس $\Pr[\hat{I}_u = I_u] \geq 1-\delta$ [35] که در آن $\hat{I}_u$ رتبه پیش‌بینی‌شده برای کاربر u است. رابطه ۳ از لحاظ نظری در موارد زیر اثبات می‌شود:

اصل ۱: $\Pr[r_{(u,i)} = r] = \alpha_{(u,r)}$ تابع جرم احتمال(pmf) از $r_{(u,i)}$ را نشان‌می‌دهد، به‌طوری که $r \in [minR, maxR]$ ، $i \in I_u$ و $u \in U$.

اثبات اصل ۱. فرض کنید که احتمال امتیازدهی ارائه‌شده توسط کاربر u در سطوح مختلف با $\rho = (\rho_{(u,minR)}, \ldots, \rho_{(u,maxR)})$ نشان داده‌شده‌است که از توزیع دیریکله پیروی می‌کند؛ همچنین ρ مثالی را نشان‌می‌دهد که به‌طور تصادفی از طریق دیریکله $(\alpha_{(u,r)})$ ترسیم می‌شود؛ بنابراین، $\Pr[r_{(u,i)} = r|\rho] = \rho_{(u,r)}$ را می‌توان به‌عنوان pmf شرطی $r_{(u,i)}$ در نظر گرفت.

که یک ویژگی اساسی توزیع دیریکله را به‌عنوان $\Pr[r_{(u,i)} = r] = \int \Pr[\rho] \Pr[r_{(u,i)} = r|\rho] \, d\rho = \int \Pr[\rho] \, \rho_{(u,r)} \, d\rho = \alpha_{(u,r)}$ [36] نشان‌می‌دهد.

قضیه ۱. فرض کنید $X = X_1 + X_2 + \cdots + X_n$ و $\sigma^2 = Var[X] = \sum_{i=1}^{n} var[x_i]$ ، که در آن $X_1, X_2, \ldots, X_n$ متغییرهای تصادفی هستند که مقادیر آنها در محدوده [۰،۱] است؛ بنابراین، دو حالت را می‌توان برای هر $\tau \geq 0$ به‌دست‌آورد، به‌طوری‌که $\Pr[X \geq E[X] + \tau] < \exp\left(\frac{\tau^2}{2\left(\sigma^2 + \frac{\tau}{3}\right)}\right)$ و $pr[x \leq E[x] - \tau] < \exp\left(-\frac{\tau^2}{2(\sigma^2 + \frac{\tau}{3})}\right)$ است[37].

اثبات رابطه ۳. برای نشان‌دادن اینکه $\Pr[\hat{I}_u \neq I_u] \leq \delta$ ، از قضیه ۱ استفاده می‌کنیم. بنابراین، معادله زیر را می‌توان با درنظرگرفتن استدلال‌های احتمال اولیه تعریف کرد:

رابطه(۵) $\Pr[\hat{I}_u \neq I_u] = \Pr[\cup_{r \neq I_u} \{\hat{I}_u = r\}] \leq \sum_{r \neq I_u} \Pr[\hat{I}_u = r] \leq \sum_{r \neq I_u} \Pr[n_{u,r} \geq n_{u,I_u}]$

علاوه بر این، مجموعه‌ای از متغیرهای تصادفی نشان داده شده توسط $R^r_{(u,i)}$ به‌صورت زیر به‌دست می‌آید:

رابطه(۶)
$$R^r_{u,i} = \begin{cases} 1, with\ probability\ Pr[r_{u,i} = r] \\ 0, with\ probability\ Pr[r_{u,i} = I_u] \\ 1/2, otherwise \end{cases}$$

که $r_{(u,i)}$ از طریق اصل ۱، $r \in [minR, maxR]$، $u \in U$ و $i \in I_u$ محاسبه می‌شود. می‌توان ثابت کرد که اگر $R^r_u = n_{(u,r)} + (I_u - n_{(u,I_u)}) / 2$ آنگاه $R^r_u = \sum_{(i \in I_u)} R^r_{(u,i)}$ (بنابراین می‌توان معادله زیر را تعریف کرد:

رابطه (۷)
$$R^r_u \geq \frac{I_u}{2} \Leftrightarrow n_{u,r} + \frac{I_u - n_{u,r} - n_{u,I_u}}{2} \geq \frac{I_u}{2} \Leftrightarrow n_{u,r} \geq n_{u,I_u}$$

علاوه بر این، می‌توان چنین استنباط کرد که $\Pr[n_{(u,r)} \geq n_{(u,I_u)}] = \Pr[R^r_u \geq I_u/2]$ است. بر این اساس برای اثبات رابطه ۳ باید نشان دهیم که $\Pr[R^r_u \geq \frac{l_u}{2}] \leq \frac{\sigma}{maxR - minR}$، $\forall r \neq I_u$ است.

به این معنی که $\Pr[\hat{I}_u \neq I_u] \leq \sum_{(r \neq I_u)} \Pr[n_{(u,r)} \geq n_{(u,I_u)}]$ با فرض اینکه: $\Pr[Rur \geq Iu/2] \leq \delta$

$\mathrm{Var}[R^r_u] = \frac{(\alpha_{u,l_u} + \alpha_{u,r} - (\alpha_{u,l_u} - \alpha_{u,r})^2) l_u}{4}$ واریانس است از $E[R^r_u] = \frac{l_u}{2} - \frac{(a_{u,l_u} - a_{u,l_r}) l_u}{2}$ به انتظار اشاره دارد. بنابراین، قضیه ۱ برای $\tau = \frac{(a_{u,l_u} - a_{u,l_r}) l_u}{2}$ اعمال می‌شود تا معادله زیر به‌دست آید:

(۸)

$$\Pr\left[R^r_u \geq \frac{I_u}{2}\right] = \Pr[R^r_u \geq E[R^r_u] + \tau] \leq \exp\left(-\frac{\tau^2}{2\left(\mathrm{Var}[R^r_u] + \frac{\tau}{3}\right)}\right)$$

$$= \exp\left(-I_u \times \left(\frac{2(\alpha_{u,I_u} + \alpha_{u,r})}{(\alpha_{u,I_u} - \alpha_{u,r})^2} - 2 + \frac{4}{3(\alpha_{u,I_u} - \alpha_{u,r})}\right)^{-1}\right)$$

$$\leq \exp\left(-I_u \times \left(\frac{2(\alpha_{u,I_u} + \tilde{\alpha}_u)}{(\alpha_{u,I_u} - \tilde{\alpha}_u)^2} - 2 + \frac{4}{3(\alpha_{u,I_u} - \tilde{\alpha}_u)}\right)^{-1}\right) \leq \frac{\delta}{(\mathrm{maxR} - \mathrm{minR})}$$

[DOI: 10.61186/jsdp.20.4.89]

رابطه ۸ به اثبات رابطه ۳ می‌انجامد. گفتنی است که روش پیشنهادی برای تعیین این‌که یک پروفایل امتیازدهی چقدر مؤثر است، استفاده می‌شود. برای هر کاربرu ، اگر $|I_u| < n_u'$ باشد، این پروفایل به‌عنوان یک پروفایل غیرقابل‌اعتماد در نظر گرفته‌می‌شود، که باید گسترش یابد؛ بنابراین، n_u' محاسبه‌شده را می‌توان در روش پیشنهادی گسترش پروفایل برای تعیین این‌که آیا یک پروفایل امتیازدهی باید گسترش یابد یا نه، استفاده‌کرد.

۳-۲- محاسبه قابلیت اطمینان کاربران

در این بخش برای محاسبه قابلیت اطمینان آیتم‌ها، یک معیار قابلیت اطمینان را با استفاده از سه فاکتوری که در بخش بعدی استفاده شده‌است، معرفی می‌کنیم. بدین‌منظور در مرحله نخست، این فاکتورها تعریف و برای محاسبه مقدار قابلیت اطمینان، ترکیب‌شده‌اند. برای محاسبه قابلیت اطمینان کاربر n، تعداد امتیازها در پروفایل کاربر u به‌عنوان نخستین فاکتور در نظر گرفته شده‌است. باید توجه شود که تخصیص امتیازهای بیشتر توسط کاربر منجر به افزایش قابلیت اطمینان پیش‌بینی‌ها در سامانه‌های توصیه‌گر می‌شود؛ بنابراین یک ارتباط مستقیم بین مقدار قابلیت اطمینان برای یک کاربر و تعداد امتیازها در پروفایل آن وجود دارد. رابطه (۹) نخستین فاکتور قابلیت اطمینان کاربر u را به‌صورت زیر محاسبه می‌کند:

$$f_i(I_u) = 1 - \frac{\bar{u}}{\bar{u}+|I_u|} \qquad (۹)$$

که در آن I_uبه آیتم‌هایی اشاره دارد که توسط کاربر u امتیازدهی شده‌اند، و$\bar{u}$ نشان‌دهنده میانه مقادیر برای$|I_u|$ است؛ علاوه‌بر امتیازهای اختصاص‌داده‌شده توسط کاربر u، امتیازهایی که توسط کاربران در مجموعه نزدیک‌ترین همسایگان خود آورده شده‌است، تأثیر قابل‌توجهی بر قابلیت اطمینان کاربر u دارد؛ بنابراین، ما این معیار مهم را دومین عامل قابلیت اطمینان کاربر u در نظر می‌گیریم که تأثیر مثبتی بر مقدار قابلیت اطمینان دارد؛ زیرا درنظرگرفتن امتیازهای بیشتر در نتایج فرآیند توصیه منجر به افزایش قابلیت اطمینان پیش‌بینی‌ها می‌شود. مجموعه نزدیک‌ترین همسایگان با استفاده از مقادیر شباهت بین کاربر هدف و سایر کاربران ساخته می‌شود. زیرمجموعه‌ای از بیش‌تر کاربران مشابه به‌عنوان نزدیک‌ترین مجموعه همسایگان انتخاب می‌شود. برای این منظور، مقدار شباهت بین کاربران u و v با استفاده از تابع ضریب همبستگی پیرسون به‌صورت زیر اندازه‌گیری می‌شود:

(۱۰)

$$sim(u,v) = \frac{\sum_{i\in I_{u,v}}(r_{u,i}-\bar{r}_u)\,(r_{v,i}-\bar{r}_v)}{\sqrt{\sum_{i\in I_{u,v}}(r_{u,i}-\bar{r}_u)^2}\sqrt{\sum_{i\in I_{u,v}}(r_{v,i}-\bar{r}_v)^2}}$$

که امتیاز بخش i که توسط کاربر u اختصاص داده‌شده‌است با $r_{u,\,i}$ نشان داده‌می‌شود، میانگین امتیازهای ارائه‌شده توسط کاربر u با $\bar{r}_u$ نشان داده‌می‌شود، و $I_{u,\,v}$ مجموعه‌ای از آیتم‌های رایج برای کاربرu و v را نشان‌می‌دهد. پس از محاسبه مقادیر شباهت، نزدیک‌ترین همسایگان با انتخاب N کاربر با بالاترین مقادیر شباهت تعیین می‌شوند. سپس فاکتور دوم با استفاده از تعداد امتیازهای ارائه شده توسط نزدیک‌ترین همسایگان به‌صورت زیر تعریف می‌شود:

$$f_{i_{NN}}(I_{NN_u}) = 1 - \frac{\bar{\imath}_{NN}}{\bar{\imath}_{NN}+|I_{NN_u}|} \qquad (۱۱)$$

که در آن NN_u نشان دهنده نزدیک‌ترین همسایگان کاربر uاست، I_{NNu} به مجموعه آیتم‌هایی اشاره دارد که توسط نزدیک‌ترین همسایگان امتیازدهی شده‌اند، $|I_{NNu}|$تعداد امتیازهای تولید شده توسط نزدیک‌ترین همسایگان را نشان‌می‌دهد. و $\bar{\imath}_{NN}$دلالت بر میانه مقادیر $|I_{NNu}|$ دارد. سومین فاکتور مورد استفاده در قابلیت اطمینان کاربران در رابطه با مقادیر شباهت بین کاربر هدف و نزدیک‌ترین همسایگان تعریف شده‌است. مقادیر تشابه بیشتر منجر به افزایش قابلیت اطمینان پیش‌بینی‌ها در سامانه‌های توصیه‌گر می‌شود. بنابراین، این عامل با توجه به جمع مقادیر شباهت با استفاده از رابطه زیر تعریف می‌شود:

$$f_s(S_u) = 1 - \frac{\bar{s}}{\bar{s}+S_u} \qquad (۱۲)$$

$$S_u = \sum_{v\in NN_u} sim(u,v) \qquad (۱۳)$$

و $\bar{s}$ به میانه مقادیر S_u اشاره دارد.

برای محاسبه قابلیت اطمینان کاربرu ، فاکتورهای تعریف شده با استفاده از مکانیزم مبتنی بر وزن ترکیب می‌شوند. با بررسی رابطه‌های ۹، ۱۱، و ۱۲، می‌توانیم نتیجه بگیریم که فاکتورهای اول و دوم به‌ترتیب با توجه به تعداد آیتم‌های امتیازدهی شده توسط کاربر u و تعداد امتیازهای تولید شده توسط نزدیک‌ترین همسایه‌ها، به‌طور مستقل محاسبه می‌شوند. بنابراین، برای محاسبه قابلیت اطمینان کاربرu ، وزن فاکتور اول و دوم برابر با ۱ تنظیم می‌شوند. فاکتور سوم به مقدار فاکتور دوم بستگی

[DOI: 10.61186/jsdp.20.4.89]

دارد. همچنین هرچه مقدار فاکتور دوم بزرگ‌تر باشد، مقدار فاکتور سوم را افزایش می‌دهد و بالعکس. بر این اساس، برای محاسبه قابلیت اطمینان کاربر هدفu ، وزن عامل سوم به‌عنوان مقدار عامل دوم تعیین می‌شود. بنابراین، قابلیت اطمینان کاربر هدف u با استفاده از میانگین هندسی فاکتورهای تعریف شده به‌صورت زیر محاسبه می‌شود:

(۱۴)

$$UR_u = \left[f_i(I_u).f_{i_{NN}}(I_{NN_u}).f_s(S_u)^{f_{i_{NN}}(I_{NN_u})}\right]^{\frac{1}{2+f_{i_{NN}}(I_{NN_u})}}$$

که UR_u به قابلیت اطمینان کاربر u اشاره دارد.

۳-۳- محاسبه قابلیت اطمینان آیتم‌ها

در این بخش، قابلیت اطمینان آیتم‌ها با استفاده از یک معیار با توجه به سه فاکتور مختلف ارزیابی می‌شود. بدین‌منظور این فاکتورها تعریف شده‌اند و سپس ترکیبی از آنها به‌عنوان قابلیت اطمینان آیتم‌ها در نظر گرفته می‌شوند. فاکتور نخست با توجه به تعداد امتیازهای ایجادشده برای یک آیتم توسط کاربران تعیین می‌شود. تخصیص امتیازهای بیشتر به یک آیتم منجر به افزایش قابلیت اطمینان آن می‌شود؛ بنابراین فاکتور نخست تأثیر مثبت بر روی قابلیت اطمینان آیتم دارد. برای تعریف فاکتور نخست از رابطه زیر استفاده می‌شود:

$$f_i(I_i) = 1 - \frac{\bar{\iota}}{\bar{\iota}+|I_i|} \quad (۱۵)$$

که $|I_i|$تعداد امتیازهای ایجادشده را برای آیتم i نشان‌می‌دهد و$\bar{\iota}$ به میانه مقادیر $|I_i|$ اشاره دارد. دومین فاکتور مورد استفاده در محاسبه قابلیت اطمینان آیتم، انحراف استاندارد امتیازهایی است که توسط کاربران به آیتم هدف i اختصاص می‌یابد. هدف اصلی این فاکتور، اندازه‌گیری تفاوت نظرات کاربران در مورد کالای مورد نظر است که در آن، مقدار بالاتر این فاکتور باعث کاهش مقدار قابلیت اطمینان می‌شود؛ بنابراین، این فاکتور بر مقدار قابلیت اطمینان آیتم هدف، تأثیر منفی می‌گذارد و با استفاده از رابطه (۱۶) به‌صورت زیر محاسبه می‌شود:

$$f_{sd}\left(stdev(I_i)\right) = \frac{max - stdev(I_i)}{max - min} \quad (۱۶)$$

که در رابطه بالا، $stdev(I_i)$ مقدار انحراف استاندارد را برای امتیازهای ایجاد شده برای آیتم i را نشان‌می‌دهد، و حداکثر و حداقل مقدار $stdev(I_i)$ به‌ترتیب با max و min نشان داده‌شده‌است.

برای تعریف فاکتور سوم، مقادیر قابلیت اطمینان کاربران که با استفاده از رابطه ۱۴ محاسبه شده‌است، استفاده‌شده‌اند. برای محاسبه این فاکتور، از جمع مقادیر قابلیت اطمینان مربوط به کاربرانی که برای آیتم هدف i امتیازدهی کرده‌اند استفاده شده‌است. علت استفاده از فاکتور سوم این است که تخصیص امتیازدهی به آیتم هدف توسط کاربران قابل‌اعتماد، منجر به افزایش مقادیر قابلیت اطمینان آن می‌شود. بنابراین، از معادله زیر برای تعریف فاکتور سوم آیتم هدف استفاده می‌شود:

$$f_c(C_i) = 1 - \frac{\bar{c}}{\bar{c}+C_i} \quad (۱۷)$$

$$C_i = \sum_{u \in U_i} UR_u \quad (۱۸)$$

و $\bar{c}$ به مقدار میانه همه مقادیر محاسبه‌شده C_i اشاره دارد، U_iبه زیرمجموعه‌ای از کاربران اشاره دارد که آیتم هدف i را امتیازدهی کرده‌اند، و UR_u قابلیت اطمینان کاربر u است که با رابطه (۱۴) محاسبه شده‌است. در نهایت، قابلیت اطمینان آیتم‌ها با توجه به ادغام سه فاکتور به‌دست آمده است. برای این منظور، ما باید وزن این فاکتورها را توسط یک مکانیزم تعیین کنیم. تعداد امتیازهای اختصاص داده شده به آیتم هدف به عوامل دیگر بستگی ندارد، بنابراین وزن فاکتور اول بر روی ۱ تنظیم می‌شود. فاکتور دوم و سوم به‌طور مستقیم به مقدار فاکتور اول بستگی دارد. بنابراین، وزن فاکتور دوم و سوم با مقدار فاکتور اول تنظیم می‌شود. بر این اساس، میانگین هندسی سه فاکتور به‌عنوان قابلیت اطمینان آیتم هدف i در نظر گرفته می‌شود:

$$IR_i = \left[f_i(I_i).f_{sd}\left(stdev(I_i)\right)^{f_i(I_i)}.f_c(C_i)^{f_i(I_i)}\right]^{\frac{1}{1+2f_i(I_i)}} \quad (۱۹)$$

که IR_i دلالت بر قابلیت اطمینان آیتم هدف i اشاره می‌کند. این اندازه‌گیری قابلیت اطمینان در تکنیک گسترش پروفایل رتبه‌بندی پیشنهادی برای تعیین قابل اعتمادترین آیتم‌ها به‌منظور افزودن به پروفایل‌های امتیازدهی غیرقابل‌اعتماد استفاده می‌شود. که این منجر به افزایش قابلیت اطمینان تکنیک پیشنهادی با ارائه ماتریس امتیازدهی کاربر-آیتم به‌منظور فرآیند توصیه کارآمدتر می‌شود.

۳-۴- گسترش پروفایل امتیازدهی

در این بخش، یک تکنیک بسط پروفایل امتیازدهی کارآمد به‌منظور بهبود کیفیت ماتریس امتیازدهی کاربر-آیتم ارائه داده‌شده‌است. به‌طور خاص، هر پروفایل

[DOI: 10.61186/jsdp.20.4.89]

امتیازدهی نامعتبر که دارای امتیازهای ناکافی است، با ترکیب امتیازهای مجازی اضافی گسترش می‌یابد. به‌منظور تضمین قابلیت اطمینان این امتیازهای مجازی، برای هر کاربر u که پروفایل امتیازدهی غیرقابل‌اعتماد I_u دارد، آیتم‌های موجود در مجموعه $\bar{I}_u=I-I_u$ بر اساس مقادیر قابلیت اطمینان آن‌ها که با استفاده از رابطه ۱۹ محاسبه‌شده‌اند، مرتب شده‌است. و زیرمجموعه‌ای از قابل اعتمادترین آیتم‌ها در پروفایل کاربر u در نظر گرفته‌می‌شود. شایان‌ذکر است که حداقل تعداد امتیازهای مورد نیاز یعنی n_u'که با استفاده از رابطه ۳ به‌دست آمده است، برای تعیین اینکه چه تعداد امتیاز مجازی باید به پروفایل امتیازدهی اضافه شود استفاده شده‌است. بر این اساس، تعداد امتیازهای مجازی که باید در نظر گرفته شود جهت گسترش پروفایل امتیازدهی کاربر u برابر با $|n_u'-|I_u|$ است. برای محاسبه امتیازهی مجازی، ابتدا از اطلاعات جمعیت‌شناختی کاربران و ژانر فیلم‌ها برای به‌دست آوردن مقادیر شباهت بین کاربران استفاده شده‌است. مقدار شباهت بین کاربران u و v بر اساس اطلاعات جمعیت‌شناختی با میانگین گیری وزنی بر روی ویژگی‌های جمعیت‌شناختی مانند سن، جنسیت و شغل با استفاده از معادله زیر محاسبه شده‌است.

$$sim(u,v)_{demo}=\frac{\sum_{j=1}^{|D|}x_j d_j}{\sum_{j=1}^{|D|}d_j} \qquad (۲۰)$$

که در آن $sim\ (u, v)_{demo}$ نشان دهنده مقدار شباهت بین کاربران u و v بر اساس اطلاعات جمعیت‌شناختی است، D مجموعه‌ای از ویژگی‌های جمعیت‌شناختی، d_j وزن jامین ویژگی جمعیت‌شناختی، و $x_j=1$ اگرj امین ویژگی جمعیت‌شناختی برای هر دو کاربر u و v برابر باشد. در غیر این صورت $x_j=0$ است. برای محاسبه مقادیر شباهت بر اساس ژانر فیلم‌ها، ژانر فیلم‌هایی را که توسط کاربران امتیازهی شده‌اند را در نظر گرفته‌ایم. بنابراین، مقدار شباهت بین کاربران u و v بر اساس ژانر فیلم‌ها با استفاده از رابطه (۲۱) محاسبه شده‌است.

$$sim(u,v)_{genre}=\frac{\sum_{j=1}^{g}G_{u,j}G_{v,j}}{\sqrt{\sum_{j=1}^{g}G_{u,j}^2}\sqrt{\sum_{j=1}^{g}G_{v,j}^2}} \qquad (۲۱)$$

که $sim\ (u, v)_{gener}$ مقدار شباهت بین کاربران u و v را بر اساس ژانر فیلم‌ها را نشان‌می‌دهد، $G_{(|U|\times g)}$ماتریس ژانر کاربر است، gعداد ژانرهای فیلم است، $G_{(u,\ j)}$ نشان‌می‌دهد که آیا کاربر u به فیلم‌ها در ژانر j امتیاز داده است. لازم به ذکر است که اگر کاربر u حداقل به یک فیلم در ژانر j امتیاز داده باشد، $G_{(u,\ j)}=1$ در غیر این صورت، $G_{(u,\ j)}=0$ است. رابطهٔ (۲۲) میانگین هارمونیک مقادیر تشابه مبتنی بر جمعیت و ژانر را به‌صورت زیر محاسبه کرده است:

(۲۲)

$$sim(u,v)_{demo_{genre}}=\begin{cases}\frac{2\times sim(u,v)_{demo}\times sim(u,v)_{genre}}{sim(u,v)_{demo}+sim(u,v)_{genre}} \\ if\ sim(u,v)_{demo}\neq 0 \\ and\ sim(u,v)_{genre}\neq 0\\ sim(u,v)_{demo}\ if\ sim(u,v)_{demo}\neq 0 \\ and\ sim(u,v)_{genre}=0\\ sim(u,v)_{genre}\ if\ sim(u,v)_{demo}=0 \\ and\ sim(u,v)_{genre}\neq 0\\ 0\ if\ sim(u,v)_{demo}=0 \\ and\ sim(u,v)_{genre}=0\end{cases}$$

که در آن $sim\ (u, v)_{demo\text{-}gener}$ مقدار شباهت ترکیب شده بین کاربران u و v را بر اساس اطلاعات جمعیت‌شناختی و ژانرها نشان داده است .در نهایت، امتیاز مجازی $VR_{(u,\ i)}$برای کاربر u و هر آیتم $i\in\bar{I}_u$ با توجه به مقادیر تشابه ترکیبی به‌صورت زیر محاسبه شده‌است:

$$VR_{u,i}=\frac{\sum_{v\in NN_u}sim(u,v)_{demo\_genre}\times r_{v,i}}{\sum_{v\in NN_u}sim(u,v)_{demo\_genre}} \qquad (۲۳)$$

که در آن r (v, i) به امتیاز مربوط به کاربر u و آیتم i اشاره دارد و NN_u نشان‌دهنده نزدیک‌ترین همسایه کاربر uاست که با انتخاب N کاربر با بالاترین مقادیر شباهت به‌دست می‌آید. پس از محاسبه امتیازهای مجازی، با گنجاندن این امتیازهای مجازی، می‌توان پروفایل‌های امتیازدهی را گسترش داد. برای هر کاربرu ، اگر $|I_u|<n_u'$ ، امتیازدهی مجازی در مجموعه$\bar{I}_u$ با استفاده از رابطه (۲۳) محاسبه شده‌است. سپس قابلیت اطمینان آیتم‌ها که با استفاده از رابطه (۱۹) محاسبه شده‌است، برای مرتب‌سازی آیتم‌های مجموعه$\bar{I}_u$ استفاده شده‌است. در نهایت، پروفایل کاربر u با افزودن تعداد $|n_u'-|I_u|$ امتیازدهی مجازی با بالاترین قابلیت اطمینان گسترش پیدا کرده است. این فرآیند برای همه کاربران با پروفایل‌های غیرقابل‌اعتماد تکرار می‌شود که منجر به ایجاد یک ماتریس امتیازدهی کاربر- آیتم جدید با استفاده از روش پیشنهادی گسترش پروفایل رتبه‌بندی شده‌است.

۳-۵-توصیه

برای ایجاد یک فهرست توصیه برای کاربر هدف، ابتدا مقادیر شباهت نهایی بین کاربران با استفاده ازپروفایل‌های امتیازدهی توسعه‌یافته محاسبه می‌شود و

[DOI: 10.61186/jsdp.20.4.89]

سپس امتیازهای ناشناخته پیش‌بینی می‌شوند. برای این منظور، از رابطه ۱۰ برای محاسبه این مقادیر شباهت به‌عنوان ضریب همبستگی پیرسون استفاده شده‌است. همچنین امتیاز ناشناخته $P_{(u,i)}$ برای کاربر u و آیتم i به‌صورت زیر پیش بینی می‌شود:

$$P_{u,i} = \bar{r}_u + \frac{\sum_{v \in NN_u} sim(u,v) \times (r_{v,i} - \bar{r}_v)}{\sum_{v \in NN_u} sim(u,v)} \quad (۲۴)$$

که در آن sim (u, v) به مقدار شباهت نهایی بین کاربران uو v طبق رابطه ۱۰ بر اساس پروفایل‌های امتیازدهی توسعه یافته اشاره دارد و NN_u مجموعه‌ای حاوی N کاربر را نشان‌می‌دهد که به‌عنوان نزدیک‌ترین همسایه کاربر u در نظر گرفته می‌شوند. پس از محاسبه امتیازهای مجهول با استفاده از رابطه ۲۴، فهرست توصیه‌های top_N برای کاربر هدف ایجاد می‌شود. الگوریتم ۱ شبه کد مدل GDPE را نشان‌می‌دهد.

۴- آزمایش‌ها

در این بخش، میزان مؤثربودن روش پیشنهادی با ارزیابی عملکرد آن در مقایسه با سایر رویکردهای توصیه از طریق انجام آزمایش‌های گسترده بررسی شده‌است. بر این اساس، رویکردهای فاکتورگیری ماتریس احتمالی (۳۸) [PMF]، فاکتورگیری ماتریس احتمالی غیرخطی (۳۹)[NLPMF]، فیلتر مشارکتی چندسطحی(۴۰) [MLCF]، بسط پروفایل آیتم سراسری (۴۱) [IGPE]، تحلیل معنایی نهفته احتمالی مبتنی بر محبوبیت (۴۲) [PPLSA]، مدل فاکتورگیری ماتریس بر اساس داده‌های بازپیوندی(۴۳)[MF-LOD]، شباهت مبتنی بر همبستگی پروفایل کاربر (UPCSim) برای سامانه‌های توصیه‌کننده فیلم [۴۴]، پروفایل مثبت و پروفایل منفی (PP/NP) برای سامانه‌های توصیه‌کننده فیلم [۲۷]، و فیلتر ترکیبی شباهت ژانر(۲۸) [GSHF] به‌عنوان رقبا در آزمایش‌ها استفاده می‌شوند.

۴-۱- مجموعه‌داده‌ها

مجموعه‌داده‌های MovieLens 100Kو MovieLens 1M دو مجموعه‌داده توصیه فیلم هستند که توسط پروژه پژوهشی GroupLens جمع آوری شده‌اند و در این مقاله برای انجام آزمایش‌ها استفاده‌شده‌اند. در مجموعه‌داده MovieLens 100K، تعداد صدهزار امتیاز وجود دارد که توسط ۹۴۳ کاربر به ۱۶۸۲ فیلم اختصاص داده‌شده‌است درحالی‌که در مجموعه‌داده MovieLens 1M ،تعداد ۱۰۰۰۲۰۹ امتیاز وجود دارد که توسط ۶۰۴۰ کاربر به ۳۹۰۰ فیلم داده‌شده‌است. امتیازها در این مجموعه‌داده‌ها در محدوده [۵،۱] با گام ۱ است که در آن ۱ و ۵ به‌ترتیب کمترین و بیشترین علاقه را نشان می‌دهند. این مجموعه‌داده‌ها شامل اطلاعات جمعیت‌شناختی کاربران از جمله سن، جنسیت، شغل و کد پستی و همچنین ۱۸ ژانر مختلف فیلم است.

۴-۲- معیارهای ارزیابی

برای نشان‌دادن کارایی روش‌های پیشنهادی، از چهار معیار ارزیابی: میانگین خطای مطلق(MAE) ، دقت، یادآوری و F1 در آزمایش‌ها استفاده می‌شود MAE معیاری را نشان می‌دهد که در آن میزان خطا بین امتیازهای پیش‌بینی‌شده (p_i) و امتیازهای واقعی (r_i) به‌صورت رابطه زیر می‌شود:

$$MAE = \frac{1}{|Te|} \sum_{i=1}^{|Te|} |r_i - p_i| \quad (۲۵)$$

که |Te| نشان‌دهنده تعداد امتیازهایی است که توسط توصیه‌کنندگان پیش بینی شده‌است.

دقت و یادآوری دو معیاری هستند که کیفیت فهرست توصیه‌های تولید شده توسط روش‌های توصیه را نشان می‌دهند. نرخ آیتم‌های مرتبط تولید شده توسط روش‌های توصیه در فهرست توصیه‌ها و نرخ آیتم‌های مرتبط در مجموعه‌آزمایشی توصیه‌شده به کاربران به‌ترتیب به‌عنوان معیارهای دقت و یادآوری هستند. این معیارها را می‌توان به‌صورت زیر محاسبه کرد:

(۲۶)

$$precision = \frac{|\{recommended\ items\ that\ are\ relevant\}|}{top\_N}$$

(۲۷)

$$recall = \frac{|\{recommended\ items\ that\ are\ relevant\}|}{|\{all\ relevant\ items\}|}$$

معیارF1 به‌عنوان میانگین هارمونیک مقادیر دقت و یادآوری تعریف می‌شود که می‌تواند به‌صورت زیر محاسبه شود:

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (۲۸)$$

۴-۳-آزمایش‌ها

روش GDPE پیشنهادی دارای پارامترهای ورودی است که مقادیر آنها باید قبل از انجام آزمایش‌ها تعیین شود. پارامترهای δ و N در معادله (۳) و (۲۴) برای هر دو مجموعه‌داده Movielens 100K و Movielens 1M روی δ=0.4 و N=90 تنظیم‌شده‌اند. این مقادیر از طریق انجام

[DOI: 10.61186/jsdp.20.4.89]

چندین آزمایش که در بخش ۴-۵ مورد بحث قرار گرفته‌اند، به‌دست می‌آیند. پارامتر d_j وزن ژامین ویژگی جمعیت‌شناختی است که در معادله ۲۰ استفاده شده‌است. ما از اطلاعات سه ویژگی جمعیت‌شناختی برای کاربران شامل سن، جنسیت و شغل استفاده می‌کنیم که وزن آنها به‌ترتیب از چپ به راست ۰.۴، ۰.۳ و ۰.۳ تنظیم شده‌است. گفتنی است که وزن ویژگی‌های جمعیت‌شناختی با انجام یک استراتژی جستجوی حریصانه برای یافتن مقادیر بهینه آنها انتخاب شده‌است. در آزمایش‌ها، ما از سطوح مختلف پراکندگی برای نشان دادن اثربخشی روش پیشنهادی مقایسه شده در کاهش مشکل پراکندگی داده‌ها استفاده کرده‌ایم. بر این اساس، یک سطح پراکندگی θ را با اشاره به مقدار داده‌های در نظر گرفته شده در فرآیند آموزش روش‌های توصیه تعریف کرده‌ایم. سطوح پراکندگیθ بر روی 10%، 20%، 50%، 80%، 90% تنظیم شده‌است که در آن θ=20% به این معنی است که مجموعه آموزشی با انتخاب ۸۰٪ کل داده‌ها تشکیل شده و مجموعه‌آزمون شامل داده‌های باقی‌مانده است. با توجه به این فرض، مقدار بالاتر سطح پراکندگی نشان داده است که مقدار کمتری از داده‌ها در مجموعه آموزشی استفاده شده‌است.

۴-۴-مقایسه با سایر مدل‌ها

نتایج آزمایش‌های انجام‌شده بر روی مجموعه‌داده های Movielens 100K و Movielens 1M در این بخش مورد بحث قرار می‌گیرد، تا برتری مدل ما در مقایسه با سایر توصیه‌کنندگان بررسی شود. نتایج آزمایش‌ها بر اساس معیار MAE در جدول ۱، توانایی مدل توصیه پیشنهادی را در حل مشکل سطوح مختلف پراکندگی داده‌ها نشان داده است. با توجه به اینکه در مدل GDPE، مقادیر MAE برای تمام سطوح پراکندگی کمتر از سایر مدل‌ها است، بنابراین روش پیشنهادی بهترین نتایج را در مقایسه با سایر رقبا به‌دست آورده‌است. در مدل پیشنهادی برای θ = ٪۲۰، معیار MAE بر روی مجموعه‌داده های Movielens 100K و Movielens1M به‌ترتیب مقادیر 0.802 و ۰.۸۲۷ به‌دست آورده‌است. همچنین در روش MF-LOD، معیار MAE با به‌دست آوردن مقادیر 0.849 و ۰.۸۶۱ به‌ترتیب بر روی مجموعه‌داده های Movielens 100K و Movielens 1M ،دومین بهترین عملکرد را دارد. باید توجه کرد که، عملکرد مدل‌های مقایسه شده با افزایش سطح پراکندگی θ از ۱۰٪ به 90% کاهش یافته‌است، زیرا افزایش سطح پراکندگی منجر به در نظر گرفتن داده‌های کمتر در فرآیند آموزش روش‌های توصیه می‌شود. در نتیجه می‌توان نتیجه گرفت که مدل پیشنهادی GDPE نسبت به سایر مدل‌ها، توانایی بالاتری در کاهش مشکل پراکندگی داده‌ها دارد. نتایج آزمایش‌ها برای هر دو مجموعه‌داده Movielens 100K و Movielens 1M در جداول ۲، ۳ و ۴ به‌ترتیب برای معیارهای دقت، یادآوری و F1 گزارش‌شده‌اند. در این آزمایش‌ها، سطوح پراکندگی مختلف از θ=10% تا θ=90% را با گام ۱۰٪ در نظر گرفته‌ایم؛ علاوه‌بر این، تعداد آیتم‌های توصیه‌شده به کاربر هدف برای همه روش‌های توصیه مقایسه شده، روی top_N=10 تنظیم شده‌است. این نتایج نشان‌می‌دهد که مدل پیشنهادی GDPE در مقایسه با سایر رقبا به مقادیر بالاتری از معیارهای دقت، یادآوری و F1 دست یافته‌است. شایان‌ذکر است که این معیارها نشان می‌دهند که، فهرست توصیه‌های به‌دست‌آمده چقدر برای کاربر هدف مناسب است. بنابراین، می‌توان نتیجه گرفت که روش پیشنهادی، فهرست توصیه‌های دقیق‌تر و مناسب‌تری را نسبت به سایر توصیه‌کننده‌ها برای کاربر هدف ایجاد کرده است. روش پیشنهادی برای معیارهای دقت، یادآوری و F1 بر روی مجموعه‌داده Movielens 100K، درمورد سطح پراکندگی θ= ٪۲۰ به‌ترتیب به مقادیر ۰.۶۸۹، ۰.۷۳۴ و ۰.۷۱۰ دست می‌یابد. همچنین برای θ= ٪۲۰، مدل MF-LOD بعد از روش پیشنهادی، بهترین مدل است که در آن مقادیر معیارهای دقت، یادآوری و F1 به‌ترتیب ۰.۶۴۱، ۰.۶۸۳ و ۰.۶۶۱ هستند. مقادیر تمام معیارهای ارزیابی زمانی کاهش می‌یابد که سطح پراکندگیθ از ۱۰٪ به 90% افزایش یابد. نتایج جداول ۴-۲ نشان می‌دهد که روش پیشنهادی نسبت به سایر روش‌ها در کاهش مشکل پراکندگی داده‌ها کاراتر است. از دلایل اصلی این دستاورد، افزودن امتیاز مجازی به پروفایل امتیازدهی کاربر و همچنین استفاده از ژانرهای فیلم و اطلاعات جمعیت‌شناختی کاربران در روش پیشنهادی است.

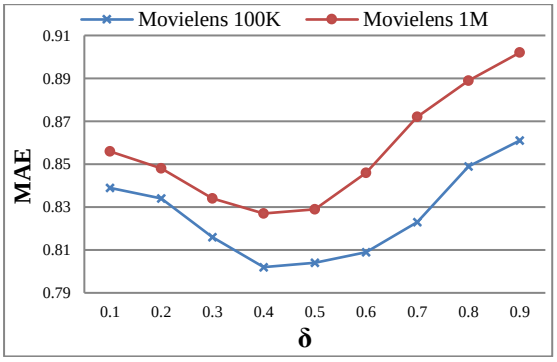
۴-۵-تجزیه و تحلیل پارامتر

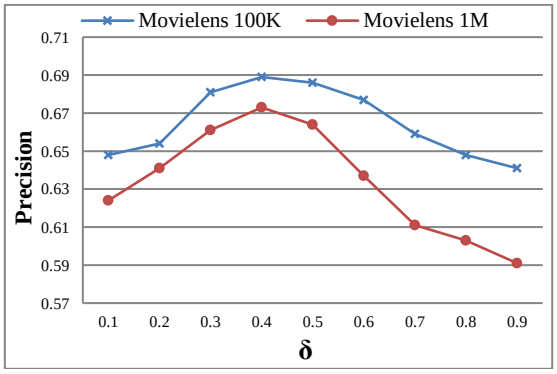
در این بخش، تأثیر مقادیر مختلف پارامترهای ورودی بر کارایی روش GDPE پیشنهادی با توجه به معیارهای ارزیابی مختلف بررسی شده‌است. برای انجام این کار، تعدادی آزمایش بر روی مجموعه‌داده‌های Movielens

[DOI: 10.61186/jsdp.20.4.89]

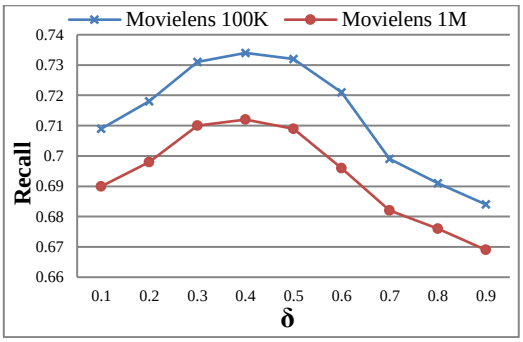
100K و Movielens 1M انجام شده‌است. شکل (۲) مقادیر معیارهای MAE، دقت، یادآوری و F1 را برای GDPE پیشنهادی، بر حسب مقادیر مختلف پارامتر δ برای هر دو مجموعه‌داده Movielens 100K و Movielens 1M نشان داده است. δ یک پارامتر ورودی است که در معادله (۳) استفاده شده که به‌عنوان یک مقدار آستانه از پیش تعریف شده برای تعیین سطح قابلیت اطمینان در تولید پیش بینی‌های دقیق است. گفتنی است که برای سطح پراکندگی در این آزمایش‌ها از θ =20% استفاده کرده‌ایم. نتایج آزمایش‌ها نشان داده‌است که در روش پیشنهادی GDPE، با افزایش δ از 0.1 به 0.4 مقادیر MAE کاهش یافته‌است؛ درحالی‌که، مقادیر MAE زمانی افزایش می‌یابد که δ بیشتر از ۰.۴ باشد. علاوه‌براین، افزایش δ از δ=0.1 به δ=0.4 منجر به افزایش عملکرد GDPE بر اساس معیارهای دقت یادآوری و F1 شده‌است. در حالی که معیارها ذکرشده، برای δ با مقادیری بیش از ۰.۴، کاهش‌یافته‌اند. بر این اساس، می‌توان نتیجه گرفت که δ=0.4 بهترین مقدار است که بهترین عملکرد روش پیشنهادی را از نظر تمام معیارهای ارزیابی بر روی هر دو مجموعه‌داده Movielens 100K و Movielens 1M ایجاد کرده‌است. مقادیر بیشتر از ۰.۴ برای پارامتر δ، منجر به افزایش تعداد امتیاز مجازی اضافه شده به پروفایل‌های امتیاز غیرقابل‌اعتماد شده‌است؛ بنابراین، زمانی که δ از ۰.۴ تجاوز کند، تعداد زیادی از امتیازهای مجازی در پروفایل‌های غیرقابل‌اعتماد گنجانده می‌شوند که منجر به افزایش احتمال افزودن امتیازهای نامربوط و کاهش دقت توصیه روش پیشنهادی می‌شود.

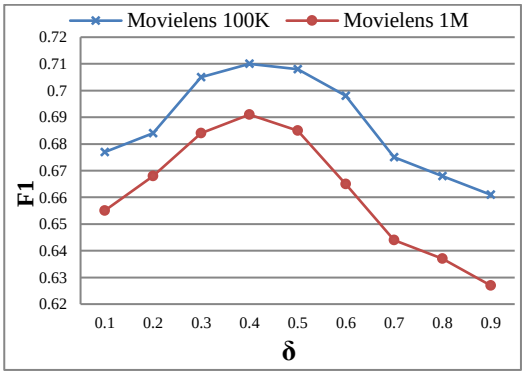
شکل (۳)، تأثیر مقادیر مختلف نزدیک‌ترین همسایه‌ها (N) را بر عملکرد GDPE اندازه گیری شده توسط معیار MAE برای هر دو مجموعه‌داده Movielens 100K و Movielens 1M نشان داده‌است. N در رابطه (۲۴) به‌عنوان تعداد نزدیک‌ترین همسایگان به کار گرفته شده در فرآیند پیش بینی امتیاز، استفاده شده‌است. در این نتایج، مقدار N از N=10 به N=120 با گام ۱۰ تغییر می‌کند. این نتایج نشان‌می‌دهد که مقادیر MAE با افزایش N از ۱۰ به ۹۰ بر روی هر دو مجموعه‌داده کاهش می‌یابد؛ بنابراین، افزایش مقدار پارامتر N از 10 به 90 منجر به افزایش عملکرد مدل پیشنهادی GDPE شده‌است؛ علاوه‌براین، زمانی که N از ۹۰ بیشتر شود، عملکرد روش پیشنهادی کاهش یافته‌است. این امر به این دلیل است که وقتی N از ۹۰ بیشتر شود، احتمال درنظرگرفتن همسایه‌های ناکارآمد در فرآیند توصیه افزایش می‌یابد، که این موضوع باعث کاهش دقت پیش‌بینی‌ها می‌شود. یکی دیگر از پارامترهای مهم مورد استفاده در فرآیند توصیه، top_N است که به تعداد آیتم در نظر گرفته‌شده در فهرست توصیه‌ها اشاره دارد. شکل‌های (۴و۵) مقادیر دقت، یادآوری و F1 روش‌های پیشنهادی برای top_N با مقادیر ۱۵،۱۰،۵ بر روی مجموعه‌داده‌های Movielens 100K و Movielens 1M نشان داده‌اند. همان‌طورکه نشان داده‌شده‌است، براساس طول فهرست توصیه‌ها بر روی دو مجموعه‌داده بالا، مقادیر تمام معیارهای یادشده برای مدل GDPE، بیشتر از سایر مدل‌های مقایسه شده‌است؛ بنابراین، روش پیشنهادی GDPE با توجه به معیارهای دقت، فراخوانی و F1 و همچنین مقادیر مختلف top_N، به‌طور قابل توجهی بهتر از سایر توصیه‌کنندگان عمل کرده‌است. شکل‌های (۴و۵) نشان داده‌اند که، مقادیر دقت همه مدل‌های مقایسه‌شده با افزایش top_N از ۵ به ۱۵ کاهش یافته‌است. همچنین، افزایش مقدار top_N منجر به افزایش مقدار معیار یادآوری برای تمام روش‌های توصیه شده‌است.





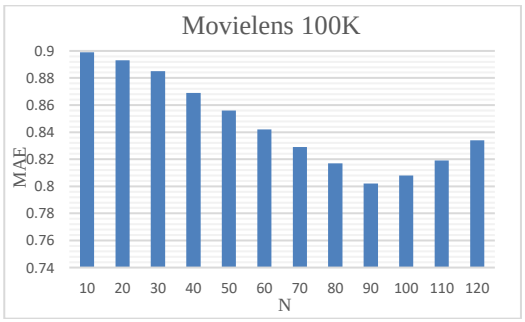
[DOI: 10.61186/jsdp.20.4.89]



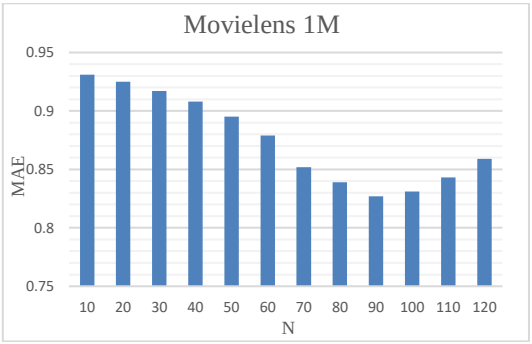


(شکل- ۲) مقادیر معیارهای MAE، دقت، یادآوری و F1 برای روش پیشنهادی بر اساس مقادیر مختلف پارامتر δ. سطح پراکندگی روی ۲۰٪ =θ تنظیم شده‌است.

(Figure –2) Values of MAE, precision, recall and F1 criteria for the proposed method based on different values of the δ parameter. The scattering level is set to θ=20%.



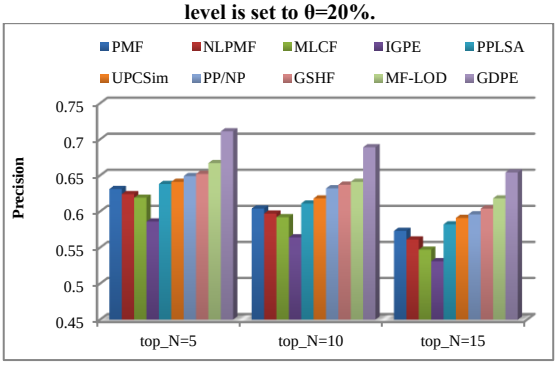
a



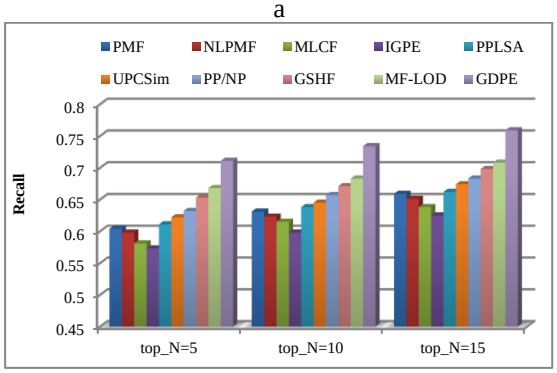
b

(شکل – ۳) مقادیر MAE برای روش پیشنهادی بر اساس تعداد متفاوت نزدیک‌ترین همسایگان (N): (a) مجموعه‌داده Movielens۱۰۰K، (b) مجموعه‌داده Movielens 1M. سطح پراکندگی روی ۲۰٪ =θ تنظیم شده‌است.

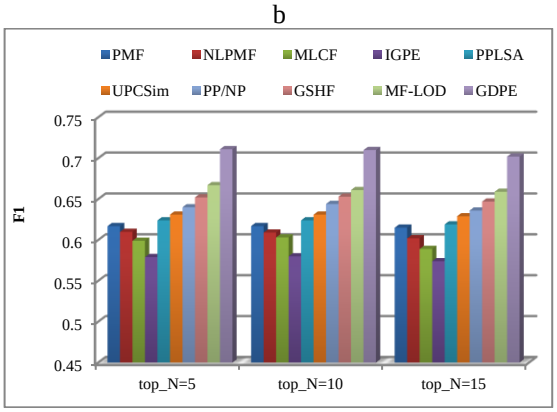
(Figure –3) MAE values for the proposed method based on different number of nearest neighbors (N): (a) Movielens 100K dataset, (b) Movielens 1M dataset. The scattering level is set to θ=20%.



a



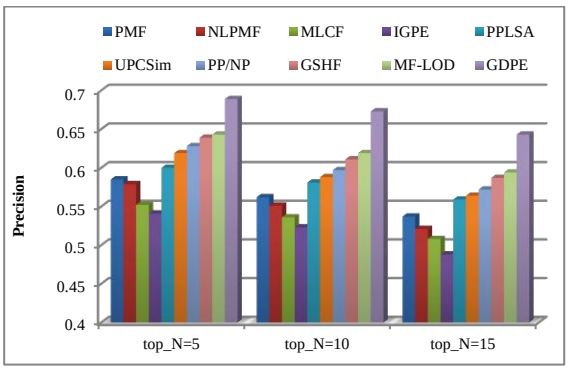
b



c

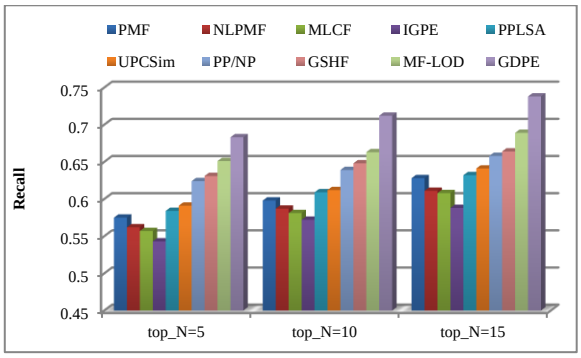
(شکل- ۴) مقایسه عملکرد در مجموعه‌داده **Movielens 100K** برحسب مقادیر مختلف **top_N**: (**a**) معیار دقت، (**b**) معیار یادآوری، و (**c**) معیار **F1**. سطح پراکندگی روی ۲۰٪ =θ تنظیم شده‌است.

Figure 4. Performance comparison on the Movielens 100K dataset in terms of different values of top_N: (a) precision measure, (b) recall measure, and (c) F1 measure. The scattering level is set to θ=20%.

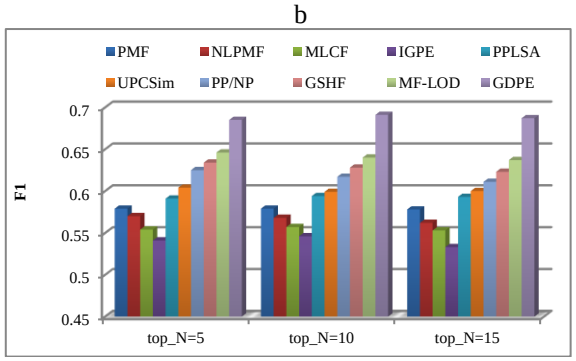


a

[DOI: 10.61186/jsdp.20.4.89]



b



c

(شکل -۵) مقایسه عملکرد در مجموعه‌داده Movielens 1M از نظر مقادیر مختلف top_N: (الف) معیار دقت، (ب) معیار یادآوری، و (ج) معیار F1. سطح پراکندگی روی ۲۰٪ =θ تنظیم شده‌است.

(Figure -5) Performance comparison on the Movielens 1M dataset in terms of different top_N values: (a) precision measure, (b) recall measure, and (c) F1 measure. The scattering level is set to θ=20%.

۵- نتیجه‌گیری

روش‌های پیشنهاد فیلم، کاربردهای زیادی در سامانه‌های دنیای واقعی دارند که هدف اصلی این روش‌ها کمک به کاربران برای یافتن فیلم‌های موردعلاقه خود است. این روش‌ها اغلب از سلایق کاربران در گذشته استفاده می‌کنند که به‌عنوان امتیازهایی که کاربران به فیلم‌های مختلف اختصاص داده‌اند، نشان داده می‌شوند. با این حال، کاربران تمایل دارند که تنها بخش کوچکی از فیلم‌ها را امتیازدهی کنند که این امر، منجر به ایجاد مشکلی به نام پراکندگی داده‌ها می‌شود. ما در این مقاله، یک روش توصیه فیلم را پیشنهاد کرده‌ایم که در آن از یک تکنیک گسترش پروفایل کارآمد با هدف توانمندسازی پروفایل امتیازدهی کاربر از طریق امتیازهای مجازی اضافی استفاده شده‌است. در روش پیشنهادی، یک مدل احتمالی برای تعیین تعداد امتیازهای مجازی به هر پروفایل امتیازدهی استفاده شده‌است. برای محاسبه این امتیازهای مجازی از ژانر فیلم‌ها و اطلاعات جمعیت‌شناختی کاربران استفاده شده‌است. برای نمایش میزان اثربخشی روش پیشنهادی، آزمایش‌های متعددی با استفاده از دو مجموعه‌داده توصیه‌شده فیلم انجام شده‌است و نتایج آنها نشان می‌دهند که مدل ما از سایر روش‌های توصیه مقایسه‌شده کارآمدتر است. برای کار آینده، سایر داده‌های جانبی مانند روابط اجتماعی و اطلاعات برچسب می‌توانند برای افزایش توانایی روش پیشنهادی در ایجاد توصیه‌های دقیق‌تر استفاده شود؛ همچنین، مدل پیشنهادی گسترش پروفایل امتیازدهی را می‌توان در انواع دیگر سامانه‌های توصیه‌کننده مانند روش‌های توصیه موسیقی استفاده‌کرد.

۶- منابع 6-Refrence

[1] H. A. Rahmani, M. Aliannejadi, S. Ahmadian, M. Baratchi, M. Afsharchi, and F. Crestani, "LGLMF: local geographical based logistic matrix factorization model for POI recommendation, " in AIRS 2019: Information Retrieval Technology, 2019, pp. 66-78.

[2] P. Moradi, F. Rezaimehr, S. Ahmadian, and M. Jalili, "A trust-aware recommender algorithm based on users overlapping community structure, " in 2016 sixteenth international conference on advances in ICT for emerging regions (ICTer), 2016, pp. 162-167.

[3] H. Xia, X. Wei, W. An, Z. J. Zhang, and Z. Sun, "Design of electronic-commerce recommendation systems based on outlier mining, " Electronic Markets, vol. 31, pp. 295–311, 2021.

[4] G. Wei, Q. Wu, and M. Zhou, "A hybrid probabilistic multiobjective evolutionary algorithm for commercial recommendation systems, " IEEE Transactions on Computational Social Systems, vol. 8, pp. 589-598, 2021.

[5] D. Wang, Y. Yih, and M. Ventresca, "Improving neighbor-based collaborative filtering by using a hybrid similarity measurement, " Expert Systems with Applications, vol. 160, p. 113651, 2020.

[6] T. Qu, W. Wan, and S. Wang, "Visual content-enhanced sequential recommendation with feature-level attention, " Neurocomputing, vol. 443, pp. 262-271, 2021.

[7] H. Li and D. Han, "A time-aware hybrid recommendation scheme combining content-based and collaborative filtering, " Frontiers of Computer Science, vol. 15, p. 154613 2021.

[8] P. Moradi, S. Ahmadian, and F. Akhlaghian, "An effective trust-based recommendation method using a novel graph clustering algorithm, " Physica A: Statistical mechanics and its applications, vol. 436, pp. 462-481, 2015.

[9] F. Rezaeimehr, P. Moradi, S. Ahmadian, N. N. Qader, and M. Jalili, "TCARS: Time-and community-aware recommendation system, " Future Generation Computer Systems, vol. 78, pp. 419-429, 2018.

[DOI: 10.61186/jsdp.20.4.89]

[10] X. Yuan, L. Han, S. Qian, L. Zhu, J. Zhu, and H. Yan, "Preliminary data-based matrix factorization approach for recommendation, " Information Processing & Management, vol. 58, p. 102384, 2021.
[11] S. Ahmadian, M. Meghdadi, and M. Afsharchi, "Incorporating reliable virtual ratings into social recommendation systems, " Applied Intelligence, vol. 48, pp. 4448-4469, 2018.
[12] S. Ahmadian, M. Afsharchi, and M. Meghdadi, "An effective social recommendation method based on user reputation model and rating profile enhancement, " Journal of Information Science, vol. 45, pp. 607-642, 2019.
[13] F. Tahmasebi, M. Meghdadi, S. Ahmadian, and K. Valiallahi, "A hybrid recommendation system based on profile expansion technique to alleviate cold start problem, " Multimedia Tools and Applications, vol. 80, pp. 2339-2354, 2021.
[14] S. Ahmadian, M. Afsharchi, and M. Meghdadi, "A novel approach based on multi-view reliability measures to alleviate data sparsity in recommender systems, " Multimedia tools and applications, vol. 78, pp. 17763-17798, 2019.
[15] S. Ahmadian, N. Joorabloo, M. Jalili, and M. Ahmadian, "Alleviating data sparsity problem in time-aware recommender systems using a reliable rating profile enrichment approach, " Expert Systems with Applications, p. 115849, 2021.
[16] S. Ahmadian, P. Moradi, and F. Akhlaghian, "An improved model of trust-aware recommender systems using reliability measurements, " in 2014 6th Conference on Information and Knowledge Technology (IKT), 2014, pp. 98-103.
[17] S. Ahmadian, N. Joorabloo, M. Jalili, M. Meghdadi, M. Afsharchi, and Y. Ren, "A temporal clustering approach for social recommender systems, " in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 1139-1144.
[18] B. Chen, Y. Ding, X. Xin, Y. Li, Y. Wang, and D. Wang, "AIRec: Attentive intersection model for tag-aware recommendation, " Neurocomputing, vol. 421, pp. 105-114, 2021.
[19] Z. Y. Khan, Z. Niu, A. S. Nyamawe, and I. Haq, "A deep hybrid model for recommendation by jointly leveraging ratings, reviews and metadata information, " Engineering Applications of Artificial Intelligence, vol. 97, p. 104066, 2021.
[20] A. Breitfuss, K. Errou, A. Kurteva, and A. Fensel, "Representing emotions with knowledge graphs for movie recommendations, " Future Generation Computer Systems, vol. 125, pp. 715-725, 2021.
[21] U. Thakker, R. Patel, and M. Shah, "A comprehensive analysis on movie recommendation system employing collaborative filtering, " Multimedia Tools and Applications, vol. 80, pp. 28647–28672, 2021.
[22] B. Walek and V. Fojtik, "A hybrid recommender system for recommending relevant movies using an expert system, " Expert Systems with Applications, vol. 158, p. 113452, 2020.
[23] H. Tahmasebi, R. Ravanmehr, and R. Mohamadrezaei, "Social movie recommender system based on deep autoencoder network using Twitter data, " Neural Computing and Applications, vol. 33, pp. 1607–1623, 2021.
[24] R. Katarya and O. P. Verma, "An effective collaborative movie recommender system with cuckoo search, " Egyptian Informatics Journal, vol. 18, pp. 105-112, 2017.
[25] K. Indira and M. K. Kavithadevi, "Efficient machine learning model for movie recommender systems using multi-cloud environment, " Mobile Networks and Applications, vol. 24, pp. 1872–1882, 2019.
[26] M. Gan and H. Cui, "Exploring user movie interest space: A deep learning based dynamic recommendation model, " Expert Systems with Applications, vol. 173, p. 114695, 2021.
[27] Y. L. Chen, Y. H. Yeh, and M. R. Ma, "A movie recommendation method based on users' positive and negative profiles, " Information Processing & Management, vol. 58, p. 102531, 2021.
[28] A. Roy and S. A. Ludwig, "Genre based hybrid filtering for movie recommendation engine, " Journal of Intelligent Information Systems, vol. 56, pp. 485–507, 2021.
[29] F. Ortega, R. L. Cabrera, Á. G. Prieto, and J. Bobadilla, "Providing reliability in recommender systems through Bernoulli Matrix Factorization, " Information Sciences, vol. 553, pp. 110-128, 2021.
[30] P. Moradi and S. Ahmadian, "A reliability-based recommendation method to improve trust-aware recommender systems, " Expert Systems with Applications, vol. 42, pp. 7386-7398, 2015.
[31] L. Huang, H. Ma, X. He, and L. Chang, "Multi-affect (ed): improving recommendation with similarity-enhanced user reliability and influence propagation, " Frontiers of Computer Science, vol. 15, p. 155331, 2021.
[32] C. A. Zayani, L. Ghorbel, I. Amous, M. Mezghanni, A. Péninou, and F. Sèdes, "Profile reliability to improve recommendation in social-learning context, " Online Information Review, vol. 44, pp. 433-454, 2020.
[33] Y. Jiang, H. Ma, Y. Liu, and Z. Li, "Exploring user trust and reliability for recommendation: A hypergraph ranking approach, " in International Conference on Neural Information Processing, 2020, pp. 333-344.
[34] S. Alonso, J. Bobadilla, F. Ortega, and R. Moya, "Robust model-based reliability approach to tackle shilling attacks in collaborative filtering recommender systems, " IEEE Access, vol. 7, pp. 41782-41798, 2019.

[DOI: 10.61186/jsdp.20.4.89]

[35] H. Xie and J. C. S. Lui, "Mathematical modeling and analysis of product rating with partial information, " ACM Transactions on Knowledge Discovery from Data, vol. 9, pp. 1-33, 2015.
[36] C. M. Bishop, Pattern recognition and machine learning: Springer-Verlag New York, 2006.
[37] J. Matoussek and J. Vondrak, The Probabilistic Method: Charles University, 2001.
[38] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization, " in NIPS'07: Proceedings of the 20th International Conference on Neural Information Processing Systems, 2007, pp. 1257–1264.
[39] N. D. Lawrence and R. Urtasun, "Non-linear matrix factorization with Gaussian processes, " in ICML '09 Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, Quebec, Canada, 2009, pp. 601-608.
[40] N. Polatidis and C. K. Georgiadis, "A multi-level collaborative filtering method that improves recommendations, " Expert Systems with Applications, vol. 48, pp. 100–110, 2016.
[41] V. Formoso, D. Fernández, F. Cacheda, and V. Carneiro, "Using profile expansion techniques to alleviate the new user problem, " Information Processing and Management, vol. 49, pp. 659–672, 2013.
[42] L. Huang, W. Tan, and Y. Sun, "Collaborative recommendation algorithm based on probabilistic matrix factorization in probabilistic latent semantic analysis, " Multimedia Tools and Applications, vol. 78, pp. 8711–8722, 2019.
[43] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using Linked Open Data, " Expert Systems with Applications, vol. 149, p. 113248, 2020.
[44] T. Widiyaningtyas, I. Hidayah, and T. B. Adji, "User profile correlation-based similarity (UPCSim) algorithm in movie recommendation system, " Journal of Big Data, vol. 8, pp. 1-21, 2021.
[45] ربی انگورانی، مهرداد، "ارائه راهکاری کارآمد برای گر با حفظ تنوع آرا"،۱۳۹۶، های توصیه‌سامانه پایان‌نامه دانشگاه شیراز، کتابخانه دانشگاه شیراز.
[46] Geetha, G., et al. "A hybrid approach using collaborative filtering and content based filtering for recommender system." Journal of Physics: Conference Series. Vol. 1000. No. 1. IOP Publishing, 2018.
[47] عباسی مقدم، سمانه، "روشی برای بهبود سامانه‌های توصیه‌گر بر اساس شبکه اعتماد "،۱۳۸۷، پایان‌نامه دانشگاه صنعتی شریف، کتابخانه دانشگاه صنعتی شریف.
[48] Alhijawi, Bushra, and Yousef Kilani. "A collaborative filtering recommender system using genetic algorithm." Information Processing & Management 57.6 (2020): 102310.
[49] گوهری، فائزه سادات، "بهبود سامانه‌های توصیه‌گر پالایش جمعی با بهره‌گیری از شبکه اعتماد ضمنی "،۱۳۹۵، پایان‌نامه دانشگاه شهید بهشتی، کتابخانه دانشگاه شهید بهشتی.
[50] Rajendran, Dixon Prem Daniel, and Rangaraja P. Sundarraj. "Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings." International Journal of Information Management Data Insights 1.2 (2021): 100027.
[51] بلوکی، امیدرضا، "شخصی سازی فرآیند آموزشی به‌کمک سامانه‌ای توصیه گر مبتنی بر مدل احساسی کاربر"،۱۳۹۹، پایان نامه دانشگاه صنعتی امیرکبیر، کتابخانه دانشگاه صنعتی امیرکبیر.
[52] مهدوی، مهرگان، "استفاده‌ی موثر از سامانه های توصیه‌گر جهت ارائهٔ توصیه‌های شخصی‌سازی شده "،۱۳۹۹، پایان‌نامه دانشگاه فردوسی مشهد، کتابخانه دانشگاه فردوسی مشهد.

صمد محمدی دکترای تخصصی مهندسی کامپیوتر- سامانه‌های نرم‌افزاری، استاد، گروه مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد تهران مرکزی است. ایشان تدریس دروس هوش‌مصنوعی، مهندسی نرم‌افزار، سامانه‌های خبره، طراحی الگوریتم را برعهده دارند.

نشانی رایانامهٔ ایشان عبارت است از:
s.mohamadi@iauctb.ac.ir

واهه آغازاریان دکترای تخصصی مهندسی کامپیوتر، استاد، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران، تدریس دروس معماری کامپیوتر، مدارهای منطقی را برعهده دارند.

نشانی رایانامهٔ ایشان عبارت است از:
v_aghazarian@iauctb.ac.ir

[DOI: 10.61186/jsdp.20.4.89]

علیرضا هدایتی دکترای تخصصی مهندسی کامپیوتر، استاد، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران، تدریس دروس امنیت شبکه مهندسی اینترنت را برعهده دارند.
نشانی رایانامهٔ ایشان عبارت است از:
hedayati@iauctb.ac.ir

[DOI: 10.61186/jsdp.20.4.89]