

Providing a User-Based Behavior Model to Recommend a Movie Using the Social Network Analysis (Case Study: CinemaMarket)¹

MohammadHossein Zohdi

M.A., Department of Management, Economic and Progress Engineering, Iran University of Science and Technology, Tehran, Iran. Mdhzohdi@gmail.com

Mehrdad Maghsoudi

PhD., Student, IT Management, Management and Accounting Faculty, University of Shahid Beheshti, Tehran, Iran (**Corresponding Author**). m_maghsoudi@sbu.ac.ir

HamidReza Nooralizadeh

Assistant Professor, Business Management and Engineering Group, Iran University of Science and Technology, Tehran, Iran. nooralizadeh@iust.ac.ir

Abstract

Objectives: Due to the increasing share of consumption and watching videos - especially movies and series - in the basket of Iranian households, several systems have been set up to facilitate people's access to these videos. One of the most important types of these systems is the video-on-demand system which has taken unprecedented growth in attracting audiences in recent years. Just as the multiplicity of content in these systems causes users to be diverse and satisfied, this multiplicity can be more confusing for them to find interesting content. Therefore, the need for recommendation systems to further predict user interests and provide consistent content is felt more and more day by day. The purpose of this study is to provide an efficient method of recommending videos based on user viewing data in the video-on-demand system.

Methods: In this research, a new bidding algorithm based on users' tastes and video watching data in the video-on-demand system is presented. This algorithm is based on the concepts and indicators of social network analysis. How this algorithm works is that first the similarity of the videos is calculated based on the percentage of movies viewed by the user and based on that the similarity matrix of the movies is formed. In the next step, based on the similarity matrix of the films, the communication graph of the films is formed, and in the next step, while discovering the communities in the graph, the centrality indicators of each film are calculated.

1. **The present study is based on:** Master Thesis entitled: **Providing a User-Based Behavior Model to Recommend a Movie Using the Social Network Analysis**, the Student is Mohammad Hossein Zohdi, The supervisor is Hamid Reza Nooralizadeh, Department of Management, Economic and Progress Engineering, Iran University of Science and Technology, presented on 2019.

Cite: Zohdi, MohammadHossein; Maghsoudi, Mehrdad & Nooralizadeh, HamidReza (2022). Providing a User-Based Behavior Model to Recommend a Movie Using the Social Network Analysis (Case Study: CinemaMarket. *Sciences and Techniques of Information Management*, 8(1): 451-484. DOI: 10.22091/stim.2021.6720.1557

Received: 2021-03-15 ; **Revision:** 2021-06-03 ; **Accepted:** 2021-10-04

© the authors / **Publisher:** University of Qom



Then, based on the viewing data of each user and the history of his favorite and non-favorite movies, the members of the community with the most favorite movies are selected as candidates for the user and after calculating their distance from the user's favorite and non-favorite movies, and if the final index is positive, Bids are offered to the user.

Results: The performance of the proposed algorithm of this research is evaluated on the data of 50, 100, and 200 users of the CinemaMarket site, the results of which show the better performance of the proposed algorithm of this research in comparison with the algorithms of Naive Bayes, k-nearest neighbors and ID3.

Conclusions: Recommendation systems offer content by having a lot of information from users and their behavioral records. The proposed algorithm of this research can perform the task of suggesting the desired content of users with the least possible information, ie information related to users' observations with desirable and acceptable performance.

Keywords: VOD, Recommendation system, Social media analysis, Data mining, CinemaMarket.

ارائه مدلی مبتنی بر رفتار کاربران جهت پیشنهاددهی فیلم با بهره‌گیری از تحلیل شبکه‌های اجتماعی (مورد مطالعه: سینما مارکت)^۱

محمدحسین زهدی

کارشناسی ارشد، دانشکده مهندسی پیشرفت، دانشگاه علم و صنعت، تهران، ایران. Mdhzohdi@gmail.com

مهرداد مقصودی

دانشجوی دکتری، رشته مدیریت فناوری اطلاعات، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران.
m_maghsoudi@sbu.ac.ir. (نویسنده مسئول)

حمیدرضا نورعلیزاده

استادیار، گروه مدیریت و مهندسی کسب و کار، دانشکده مهندسی پیشرفت، دانشگاه علم و صنعت، تهران، ایران.
nouralizadeh@iust.ac.ir

چکیده

هدف: با توجه به افزایش سهم مصرف و تماشای ویدئو در سید مصرفی خانواده‌های ایرانی، سامانه‌های متعددی جهت تسهیل دسترسی مردم به این ویدئوها راه‌اندازی و توسعه داده شده است. یکی از مهم‌ترین انواع این سامانه‌ها، سامانه‌های ویدئوی درخواستی می‌باشند که در سال‌های اخیر، روند رشد بی‌سابقه‌ای را در زمینه جذب مخاطب طی کرده‌اند. به همان اندازه که تعدد محتوا در این سامانه‌ها موجب تنوع و رضایت کاربران می‌شود، این تعدد می‌تواند موجب سردرگمی بیشتر برای یافتن محتوای مورد علاقه کاربران شود. هدف این پژوهش ارائه روش کارآمد پیشنهاددهی فیلم بر مبنای داده‌های مشاهده کاربران در سامانه ویدئوی درخواستی است.

روش: در این پژوهش یک الگوریتم جدید پیشنهاددهی مبتنی بر سلايق کاربر و داده‌های تماشای فیلم کاربران در سامانه ویدئوهای درخواستی ارائه می‌گردد. این الگوریتم مبتنی بر مفاهیم و شاخص‌های تحلیل شبکه اجتماعی بوده و براساس داده‌های تماشای فیلم کاربران، به شناسایی و پیشنهاددهی فیلم‌های مورد علاقه کاربران می‌پردازد.

۱. پژوهش حاضر برگرفته از: پایان‌نامه کارشناسی ارشد با عنوان: ارائه مدلی مبتنی بر رفتار کاربران جهت پیشنهاددهی فیلم با بهره‌گیری از تحلیل شبکه‌های اجتماعی (مورد مطالعه: سینما مارکت)، دانشجو: محمدحسین زهدی، استاد راهنما: حمیدرضا نورعلیزاده، گروه مدیریت و مهندسی کسب و کار، دانشکده مهندسی پیشرفت، دانشگاه علم و صنعت، ارائه شده در سال ۱۳۹۷ می‌باشد.

استناد به این مقاله: زهدی، محمدحسین؛ مقصودی، مهرداد؛ نورعلیزاده، حمیدرضا (۱۴۰۱). ارائه مدلی مبتنی بر رفتار کاربران جهت پیشنهاددهی فیلم با بهره‌گیری از تحلیل شبکه‌های اجتماعی (مورد مطالعه: سینما مارکت). علوم و فنون مدیریت/اطلاعات، ۸(۱): ۴۵۱-۴۸۴.

DOI: 10.22091/stim.2021.6720.1557

تاریخ دریافت: ۱۳۹۹/۱۲/۲۵ ؛ تاریخ اصلاح: ۱۴۰۰/۰۳/۱۳ ؛ تاریخ پذیرش: ۱۴۰۰/۰۷/۱۲

ناشر: دانشگاه قم

نتایج: عملکرد الگوریتم پیشنهادی این پژوهش بر روی داده‌های ۵۰، ۱۰۰ و ۲۰۰ کاربر سایت سینما مارکت ارزیابی شده که نتایج آن، نشان‌دهنده عملکرد بهتر الگوریتم پیشنهادی این پژوهش در مقایسه با الگوریتم‌های Naive bayes، k-nearest neighbors و ID3 است.

نتیجه‌گیری: سیستم‌های پیشنهاددهنده با در اختیار داشتن اطلاعات فراوان و متعددی از کاربران و سوابق رفتاری آنان به پیشنهاد محتوا می‌پردازند. الگوریتم پیشنهادی این پژوهش قادر است با حداقل اطلاعات ممکن یعنی اطلاعات مربوط به مشاهدات کاربران با عملکرد مطلوب و قابل قبولی وظیفه پیشنهاد محتوای مطلوب کاربران را انجام دهد.

کلیدواژه‌ها: سامانه ویدئوی درخواستی، سیستم‌های پیشنهاددهنده، داده‌کاوی، تحلیل شبکه‌های اجتماعی، سینما مارکت، فیلم.

۱. مقدمه

با توجه به افزایش سهم مصرف و تماشای ویدئو- به خصوص فیلم و سریال - در سبد مصرفی خانواده‌های ایرانی، کسب و کارها و صنایع متعددی جهت تسهیل دسترسی مردم به ویدئوهای مورد نظرشان راه‌اندازی و توسعه داده شده است. یکی از این صنایع، صنعت نمایش آنلاین فیلم است که رشد سریع آن بین سال‌های ۲۰۱۳ تا ۲۰۱۸، سهم بازاریش را از درآمد کل بازار جهانی، تقریباً دو برابر کرده است (اریکسون^۱، ۲۰۱۷). یکی از مهم‌ترین بخش‌های این صنعت را سامانه‌های ویدئوی درخواستی^۲ تشکیل می‌دهند که به واسطه همین رشد سریع، روز به روز شاهد توسعه و افزایش محتواها و به تبع آن کاربران آن‌ها هستیم.

به همان اندازه که تعدد محتوا در این سامانه‌ها موجب تنوع و رضایت کاربران می‌شود، این تعدد می‌تواند موجب سردرگمی بیشتر کاربران، برای یافتن محتواهای مورد نظرشان بشود. طبق گزارش شرکت اریکسون، متوسط زمانی که کاربران برای جستجوی محتواها در سامانه‌های ویدئوی درخواستی صرف می‌کنند، ۲۴ دقیقه در روز است (اریکسون، ۲۰۱۷). به همین جهت، نیاز به سیستم‌هایی برای پیش‌بینی نیازمندی و علائق کاربران و ارائه محتواها بر مبنای این پیش‌بینی، روز به روز بیشتر می‌شود و توسعه و به‌کارگیری آن‌ها روند غیر قابل باوری را در پیش گرفته است تا حدی که یکی از روندهای مهم برای سال ۲۰۱۹، شخصی‌سازی تمامی نیازمندی‌ها پیش‌بینی شده است (اریکسون، ۲۰۱۷).

سیستم‌های پیشنهاددهنده مبتنی بر محتوا، مبتنی بر دانش، مشارکتی، ترکیبی و... بخشی از انواع الگوریتم‌ها برای پیش‌بینی علائق کاربر و ارائه پیشنهاداتی متناسب با آن می‌باشد که مدل‌ها و روش‌های مختلفی برای هر یک تاکنون ارائه شده است، ولی با پیشرفت تکنولوژی، رشد روزافزون استفاده کاربران از ویدئو و نیازمندی آن‌ها به سیستم‌هایی برای صرفه‌جویی در وقت مورد نیاز برای جستجو میان این حجم از محتوا، نیاز به الگوریتم‌هایی دقیق‌تر و سریع‌تر بسیار بیشتر از گذشته احساس می‌شود.

تحقیقات انجام شده بر روی سیستم‌های پیشنهاددهنده، عموماً از جمع‌آوری اطلاعات صریح به دلیل دسترسی آسان‌تر به آن‌ها استفاده کرده‌اند (سان و کیم^۳، ۲۰۱۷؛ جالولی و همکاران^۴،

1. Ericsson Consumer and Industry Insight report
2. Video on demand (VOD)
3. Son & Kim
4. Jallouli & et al.

(۲۰۱۷)، اما به دلیل در دسترس نبودن اطلاعات ضمنی و الگوریتم‌های مبتنی بر آن‌ها در حوزه‌های مختلف و به خصوص سامانه‌های ویدئوی درخواستی، پژوهش‌های مبتنی بر این اطلاعات کم‌تر می‌باشد. همچنین ارزیابی مدل‌های ارائه شده در ادبیات، عموماً با اتکاء بر امتیازهای کاربران به فیلم‌ها در سایت‌های جمعی چون مووی‌لنز^۱ و آی‌ام‌دی‌بی^۲ صورت گرفته است (لین و همکاران^۳، ۲۰۱۴؛ کاتایارا و ورما^۴، ۲۰۱۷؛ سان و کیم^۵، ۲۰۱۷؛ پرا و همکاران^۶، ۲۰۱۳) و جای ارزیابی‌هایی مبتنی بر اطلاعات ضمنی همچون میزان تماشای فیلم‌ها توسط کاربران در پژوهش‌های انجام شده بسیار خالی است. از طرف دیگر، بهره‌گیری از تحلیل شبکه اجتماعی روشی است که می‌تواند راهکار و منطق نوینی را برای طراحی سیستم پیشنهاددهنده برای تأمین کنندگان محتوا و به ویژه محتوای ویدئویی ارائه کند.

در این راستا، پژوهش حاضر درصدد است تا با بهره‌گیری از مفاهیم تحلیل شبکه‌های اجتماعی و داده‌های مربوط به تماشای فیلم توسط کاربران در سامانه‌های ویدئوی درخواستی، روش بهبود یافته‌ای را برای پیشنهاددهی مبتنی بر سلايق کاربر ارائه دهد.

۲. مبانی نظری

۲-۱. سیستم‌های پیشنهاددهنده^۷

سیستم‌های پیشنهاددهنده، سیستم‌هایی مؤثر در راهنمایی و هدایت کاربران در میان حجم عظیمی از انتخاب‌های ممکن برای رسیدن به گزینه مفید و مورد علاقه هستند. در تعریف کلی، سیستم‌های پیشنهاددهنده را زیرمجموعه‌ای از سیستم‌های حامی تصمیم‌گیری^۸ می‌دانند و آن‌ها را سیستم‌های اطلاعاتی تعریف می‌کنند که توانایی تحلیل رفتارهای گذشته را دارد و توصیه‌هایی برای مسائل جاری ارائه می‌دهد. به طور کلی توصیه‌های سیستم پیشنهاددهنده می‌تواند دو نتیجه داشته باشد: اولاً، به کاربران در تصمیم‌گیری کمک می‌کند و ثانیاً، موجب می‌شود آگاهی کاربر در مورد

1. MovieLens.com

2. IMDB.com

3. Lin & et al.

4. Kataraya & Verma

5. Son & Kim

6. Pera & et al.

7. Recommendation systems

8. Decision Support System (DSS)

زمینه مورد علاقه‌اش بیشتر شود (رابین^۱، ۲۰۰۲؛ استوارت^۲، ۲۰۰۳). در سال‌های اخیر شیوه‌ها، رویکردها، الگوریتم‌ها و مدل‌های مختلفی در زمینه توسعه انواع سیستم‌های پیشنهاددهنده مطرح گردیده است که از ابعاد مختلف و با نگاه‌های متفاوتی به مسأله ارائه پاسخ و پیشنهاد مناسب به کاربران پرداخته‌اند (آهن و شی^۳، ۲۰۰۹؛ جالولی و همکاران^۴، ۲۰۱۷).

سیستم‌های پیشنهاددهنده براساس ترجیحات کاربرانشان برای مجموعه‌ای از آیتم‌ها (از قبیل فیلم، موسیقی، کتاب، برنامه‌های کاربردی، مقصد سفر و...) اطلاعاتی را جمع‌آوری می‌کنند. جمع‌آوری اطلاعات هم به صورت صریح (مثل پرسش‌نامه یا رتبه‌های داده شده توسط کاربر) و هم به صورت ضمنی که اغلب با کاوش رفتار کاربر بدست می‌آید (مثل رصد کردن فیلم‌هایی که تماشا کرده یا برنامه‌هایی که بارگیری کرده است)، انجام می‌شود (ریکی و همکاران^۵، ۲۰۱۰).

سیستم‌های پیشنهاددهنده برای پیشگویی و پیشنهاددهی آیتم‌ها، از منابع اطلاعاتی مختلف استفاده می‌نمایند. این سیستم‌ها سعی می‌کنند بین دقت، جدید بودن، عدم پراکندگی و تناسب در پیشنهادات، توازن ایجاد نمایند. نیاز به سیستم‌های پیشنهاددهنده زمانی برجسته می‌شود که کاربران با اطلاعات و آیتم‌های گسترده‌ای مواجه می‌شوند. در این مواقع، سیستم باید بتواند از میان انبوه اطلاعات، براساس مواردی همچون موقعیت، نیازها و سوابق کاربر در سیستم پیشنهاد را ارائه کند (پارک و همکاران^۶، ۲۰۱۲). در واقع هدف سیستم‌های پیشنهاددهنده معرفی آیتم‌هایی است که نزدیک‌ترین شباهت را با علایق کاربر دارد (مترن و سامرن^۸، ۲۰۰۰؛ آدوماویسیوس و توژیلین^۹، ۲۰۰۵).

۲-۲. الگوریتم‌های ایجاد سیستم توصیه‌گر

• الگوریتم فیلترینگ مشارکتی^{۱۰}: این الگوریتم، پرکاربردترین و محبوب‌ترین الگوریتم

1. Robin

2. Stuart

3. Ahn & Shi

4. Jallouli & et al.

5. Item

6. Ricci & et al.

7. Park & et al.

8. Meteren & Someren

9. Adomavicius & Tuzhilin

10. Collaborative Filtering Algorithm (CFA)

پیشنهاددهی است. این الگوریتم به همان شکلی عمل می‌کند که ما نیز معمولاً در تصمیم‌گیری‌های خود در زندگی عمل می‌کنیم و آن هم استفاده از تجربیات و نظرات دیگران است. مثلاً برای خرید یک کالای خاص، به امتیازهایی که سایر کاربران به آن کالا داده‌اند، توجه می‌کنیم (سو و خوش‌گفتار^۱، ۲۰۰۹). در این روش، معمولاً سیستم، محیطی را فراهم می‌کند تا کاربران بتوانند برای هر آیتام موجود در سیستم، رتبه دلخواه خود را ثبت کنند. در بعضی موارد هم که رتبه‌بندی به صورت صریح ممکن نیست، سیستم از روی بعضی نشانه‌ها رتبه‌ای را برای آیتام‌های خود محاسبه می‌کند، برای مثال تعداد دفعاتی که از یک آیتام بازدید به عمل آمده و یا تعداد دفعات بارگیری یک آیتام یا متعلقات آن و یا تعداد دفعاتی که صفحه‌ای با محتوای شناخته شده برای سیستم، دیده شده و میانگین زمانی که کاربران معمولاً در آن صفحه صرف کرده‌اند (بوبادیللا و همکاران^۲، ۲۰۱۱).

در این روش سیستم کلیه آیتام‌های موجود را براساس امتیازهای داده شده توسط کاربران (صریح و ضمنی)، رتبه‌بندی می‌کند و به هر کاربر با توجه به علایقش آیتام‌هایی با بالاترین رتبه را پیشنهاد می‌دهد. این الگوریتم اینگونه عمل می‌کند که با توجه به امتیازهای داده شده توسط کاربران، کاربرانی را که از نظر امتیازهای داده شده با یکدیگر مشابه هستند، در یک گروه قرار می‌دهد و زمانی که می‌خواهد برای کاربری پیشنهادی مهیاء کند، آیتام‌هایی که توسط کاربران موجود در همسایگی کاربر، رتبه بالاتری را دریافت کرده‌اند، به کاربر پیشنهاد می‌دهد. الگوریتمی که در این روش، به طور معمول مورد استفاده قرار گرفته است، الگوریتم نزدیک‌ترین همسایه^۳ می‌باشد (پارک و همکاران^۴، ۲۰۱۵).

پایه و اساس این الگوریتم، استفاده از داده‌ها مبتنی بر حافظه^۵ است؛ به این معنا که داده‌های موجود در پایگاه داده مستقیماً استفاده می‌شوند. پیاده‌سازی این الگوریتم معمولاً ساده است و نتایج قابل قبولی را در پیش‌گویی‌ها و پیشنهادات تولید می‌کند. الگوریتم‌های مبتنی بر حافظه، شامل دورویکرد مبتنی بر کاربر^۶ و مبتنی بر آیتام^۷ می‌باشند (پارک و همکاران^۸، ۲۰۱۵).

1. Su & Khoshgoftaar

2. Bobadilla & et al.

3. K-Nearest Neighbors (KNN)

4. Park & et al.

5. Memory-based

6. User-based

7. Item-based

8. Park & et al.

در سیستم‌های پیشنهاددهنده بر مبنای فیلترینگ مشارکتی، اساس کار بر تعیین میزان مشابهت بین کاربران و آیت‌ها بوده و بر مبنای میزان مشابهت محاسبه شده، همسایگی‌ها ایجاد می‌شوند. برای محاسبه میزان مشابهت، نیاز به یک معیار مشابهت می‌باشد. روش‌های سنتی، از رتبه‌ها به عنوان معیار استفاده می‌کردند؛ به این شکل که کاربرانی را که به آیت‌های مشابه رتبه‌های مشابه داده بودند، در یک گروه همسایگی قرار می‌دادند. برای تعیین مشابهت بین آیت‌ها نیز اینگونه عمل می‌شد که آیت‌هایی که توسط کاربران یکسانی رتبه دریافت کرده‌اند، دارای مشابهت می‌باشند. از بین معیارهایی که اخیراً ایجاد شده‌اند می‌توان به میانگین تفاضل مربعات، پیوستگی پیرسون و مشابهت‌های کسینوسی اشاره کرد (سو و خوش‌گفتار^۱، ۲۰۰۹).

● **الگوریتم فیلترینگ مبتنی بر محتوا:** در فیلترینگ مشارکتی، پیشنهادات بر مبنای تجربیات کاربران دیگر تولید می‌شوند، در حالی که در فیلترینگ مبتنی بر محتوا، پیشنهادات براساس ویژگی‌های محتوا و سوابق و تجربیات خود کاربر تشکیل می‌شود. در واقع تجربیات و پسندهای کاربر در گذشته، مبنای پیشنهاد برای آینده قرار می‌گیرد. در این روش نیاز به آنالیز و تحلیل محتویات و اطلاعات مرتبط با کاربران و آیت‌ها می‌باشد تا با کمک آن بتوان میزان مشابهت‌های بین کاربران و آیت‌ها را محاسبه نمود. در این سیستم‌ها موضوع مهم، تعیین یک معیار مشابهت مناسب است. در این الگوریتم، آیت‌هایی به کاربر پیشنهاد داده می‌شوند که شباهت بیشتری با آیت‌هایی که در گذشته کاربر مورد نظر، آن‌ها را پسندیده، داشته باشند.

برای اینکه بتوان میزان مشابهت بین آیت‌های مختلف را محاسبه کرد، نیاز داریم که صفات و خصوصیات آیت‌ها را از منابع مختلف استخراج کنیم که این کار نیاز به راهکاری دارد که به صورت خودکار بتواند این صفات را برای مقایسه و تعیین میزان شباهت، از بین داده‌ها و اطلاعات موجود در سیستم واکنشی^۳ کند که این کار در بعضی موارد بسیار دشوار و پیچیده می‌شود (لوپس و همکاران^۴، ۲۰۱۱).

● **الگوریتم فیلترینگ مبتنی بر دانش:** سیستم‌های توصیه‌کننده که از منابع داده پیشرفته استفاده

1. Su & Khoshgoftaar
2. Content-based
3. Fetching
4. Lops & et al.
5. KNowledge-based

می‌کنند- به جز سیستم‌های پیشین-، به عنوان سیستم‌های پیشنهاددهنده بر مبنای دانش شناخته می‌شوند. این سیستم‌ها براساس دانش موجود در رابطه با کاربران و آیتم‌ها به وجود آمده‌اند. این سیستم‌ها از دیدگاه تئوری نسبت به سایر روش‌های ذکر شده، از دقت و کیفیت بیشتری برخوردار هستند، به این دلیل که سلايق و نیازهای کاربر را مورد تفسیر قرار می‌دهند و حاوی دانش هستند (میدلتون و همکاران^۱، ۲۰۰۴).

• **الگوریتم فیلترینگ مبتنی بر جمعیت‌شناختی^۲:** این سیستم‌ها تنها به ویژگی‌های جمعیت‌شناختی کاربران (نظیر سن، جنسیت، تحصیلات و...) برای یافتن روابط بین آنها و سایرین و درک آنچه دوست دارند، تکیه می‌کنند. مثلاً، وقتی پیشنهادات فیلم به کاربری ارائه می‌شود که علاقه‌مند به مستندهای فرهنگی است، بردی از اطلاعات جمعیت‌شناختی شامل زمان، سن، ملیت و پس‌زمینه فرهنگی باید بررسی شود. پازانی با استفاده از یادگیری ماشینی، یک طبقه‌بندی براساس داده‌های دموگرافیک انجام داده است (پازانی^۳، ۱۹۹۹).

• **الگوریتم فیلترینگ ترکیبی^۴:** این سیستم‌ها بر مبنای ترکیبی از روش‌هایی هستند که پیش‌تر به آنها اشاره شد. برای دستیابی به پیشنهاداتی دقیق‌تر، سیستم‌های پیشنهاددهنده ترکیبی، منابع داده مختلف حاصل از سایر سیستم‌های پیشنهاددهنده را با یکدیگر ترکیب و یکپارچه‌سازی می‌کنند. لینا و همکارانش سیستم پیشنهاددهنده ترکیبی را ارائه می‌دهند که به کاربران این امکان را می‌دهد تا محصولات مورد علاقه خود را محرمانه‌تر از حد تصور انتخاب کنند (لینا و همکاران^۵، ۲۰۱۸).

۲-۳. تحلیل شبکه‌های اجتماعی

شبکه اجتماعی^۶، یک ساختار اجتماعی است که از افراد (یا سازمان‌ها) به عنوان «گره‌ها»^۷ تشکیل شده که براساس یک یا چند نوع وابستگی متقابل مانند دوستی، خویشاوندی، منافع مشترک، مبادله مالی و یا روابطی چون باور، دانش و اعتبار به یکدیگر متصل شده‌اند. تجزیه و

1. Middleton & et al.

2. Demographic

3. Pazzani

4. Hybrid Filtering

5. Lina & et al.

6. Social network

7. Nodes

تحلیل شبکه‌های اجتماعی^۱، روابط اجتماعی را در قالب نظریه شبکه‌ای متشکل از گره‌ها و یال‌های^۲ میان آن‌ها توصیف می‌کند. گره‌ها نقش‌آفرینان منحصربه‌فرد شبکه‌ها و یال‌ها ارتباطات میان این نقش‌آفرینان هستند. ساختارهای مبتنی بر گراف^۳، اغلب بسیار پیچیده می‌باشند. تحقیقات در زمینه‌های مختلف علمی نشان داده است که شبکه‌های اجتماعی در سطوح مختلف- از خانواده‌ها گرفته تا ملت‌ها- فعالیت می‌کنند و نقش بسزایی در حل مسائل مختلف افراد و سازمان‌ها و تعیین میزان دستیابی به اهداف از پیش تعیین شده دارند. در ساده‌ترین شکل، یک شبکه اجتماعی نقشه‌ای از روابط مشخص شده- مانند دوستی- میان گره‌های مورد مطالعه نمایش می‌دهد (جلیل‌وند و همکاران، ۱۳۹۹). گره‌هایی که هر فرد به آن متصل می‌شود، نمایانگر ارتباطات اجتماعی^۴ آن فرد است. همچنین از یک شبکه می‌توان برای اندازه‌گیری سرمایه اجتماعی^۵- ارزشی که فرد از یک شبکه اجتماعی به دست می‌آورد- نیز استفاده کرد. نکته قابل توجه این است که این شبکه‌ها بیانگر روابط بین افراد و تأثیری که بر هم دارند است و نه مشخصات و خصوصیات هر فرد. این مفاهیم اغلب در یک نمودار شبکه اجتماعی نمایش داده می‌شود؛ نموداری که در آن، گره‌ها نقاط و روابط خطوط هستند (واسرمن و فاوست^۶، ۱۹۹۵).

۲-۴. شاخص‌های تحلیل شبکه‌های اجتماعی

ویژگی‌های شبکه‌های اجتماعی به سه دسته ساختی، تعاملی و کارکردی تقسیم می‌شوند که هر کدام شاخص‌هایی دارند. این شاخص‌ها بنابر مسأله و هدف تحقیق انتخاب می‌شوند. منظور از ویژگی‌های ساختی شبکه، ویژگی‌هایی است که بیشتر با ساخت و نه محتوای شبکه ارتباط دارند، مانند اندازه، تراکم و ترکیب. ویژگی‌های تعاملی بیشتر به خصوصیات مربوط به روابط بین اعضا مانند فراوانی تماس‌ها، قوت، چندگانگی، نزدیکی، مدت رابطه و... می‌پردازد. در ویژگی‌های کارکردی به کارکردهایی که شبکه برای اعضا دارد، مانند انواع حمایت‌های اجتماعی شبکه، توجه می‌شود (باستانی و رئیسی، ۱۳۹۱).

<http://stlm.gom.ac.ir>

1. Social network analysis (SNA)
2. Ties
3. Graph-based
4. Social contacts
5. Social capital
6. Wasserman & Faust

هدف پژوهش حاضر تنها بررسی ساخت شبکه است. بنابراین، تنها به بررسی مهم‌ترین شاخص‌های مربوط به ساخت می‌پردازد که با نام شاخص‌های مرکزیت^۱ شناخته می‌شوند.

● **مرکزیت درجه:**^۲ مرکزیت درجه نشان‌دهنده میزان فعالیت یا ارتباطات یک گره با سایر گره‌های موجود در شبکه است. این شاخص نشان‌دهنده تخصص، قدرت و تجربه اعضای شبکه محسوب می‌شود و مشخص می‌کند کدام گره یا گره‌ها در شبکه مهم و اثرگذارتر هستند. عدد مرکزیت درجه یک گره معادل تعداد یال‌های متصل به آن گره است.

● **مرکزیت بینیت:**^۳ این شاخص برای یک گره بیان می‌کند که چه تعداد از گره‌های شبکه برای ارتباط سریع‌تر با هم (با واسطه کم‌تر) به این گره نیاز دارند. برای یک فرد یا گره، این شاخص به صورت تعداد دفعاتی که آن گره در کوتاه‌ترین فاصله^۴ میان هر دو گره دیگر در شبکه قرار می‌گیرد، محاسبه می‌شود. هرچه میزان این شاخص برای یک گره زیادتر باشد، بدان معناست که گره در مکان استراتژیک‌تری قرار گرفته است و اگر این گره از شبکه حذف شود، شبکه به قسمت‌هایی مجزا از هم تفکیک شده و امر انتقال اطلاعات در شبکه دچار اختلال می‌شود. به بیان دیگر، این شاخص نشان‌دهنده درصدی از اطلاعات است که از یک گره می‌گذرد (جلیل‌وند و همکاران، ۱۳۹۹).

● **مرکزیت نزدیکی:**^۵ مرکزیت نزدیکی شاخصی برای مقایسه سرعت دسترسی یک گره به دیگر گره‌های شبکه است. این شاخص از محاسبه معکوس میانگین کوتاه‌ترین مسیر گره با دیگر گره‌ها محاسبه می‌شود. در نتیجه عددی میان صفر و یک خواهد بود که ارقام بالاتر به معنی نزدیکی بیش‌تر و فاصله میانگین کم‌تر است. یک مرکزیت نزدیکی بالا به این معناست که یک گره به بیش‌تر گره‌های دیگر در شبکه متصل بوده یا از آن‌ها فقط یک گام دورتر است (واسرمن و فاوست^۶، ۱۹۹۵).

۲-۵. پیشینه تحقیق

پرا و ان‌جی^۷ (۲۰۱۳) در پژوهش خود مدلی را براساس شباهت و محبوبیت محتوا ارائه

1. Centrality Measures
2. Degree centrality
3. Betweenness centrality
4. Geodesic distance
5. Closeness centrality
6. Wasserman & Faust
7. Pera & Ng

داده‌اند که پیشنهاددهای فیلم را به صورت جذاب (و به میزان مشخصی) به اعضای یک گروه ارائه می‌دهد. به این صورت که اولاً، استراتژی‌هایی را برای بررسی و نمایش ترجیحات اعضای گروه مشخص می‌کند؛ ثانیاً، عوامل همبستگی میان محتواها را برای یافتن فیلم‌های مشابه به کار می‌گیرد؛ ثالثاً، میزان محبوبیت فیلم‌ها را در سایت مورد بررسی در نظر می‌گیرد. برخلاف الگوریتم فیلترینگ مشارکتی که رتبه‌بندی فیلم‌ها را مبنایی برای ارائه پیشنهادات در نظر می‌گیرد، این روش برچسب‌های متناسب با فرد را مبنایی برای توصیه‌ها و منافع مشترک اعضای گروه به کار می‌گیرد. مطالعات تجربی مؤید کارآمدی بیشتر این روش در مقایسه با الگوریتم‌های فیلترینگ مشارکتی است.

لین و همکارانش^۱ (۲۰۱۴) در پژوهش خود یک سیستم پیشنهاددهنده را برای علاقه‌مندان به فیلم ارائه داده‌اند که از الگوریتم‌های فیلترینگ مشارکتی با استفاده از شبکه‌های عصبی برای بهبود تصمیم‌گیری مصرف‌کنندگان بهره برده است. نتایج تجربی نشان می‌دهد که نه تنها این روش توانسته پیش‌بینی رتبه‌بندی فیلم‌ها را بهبود بخشد، بلکه باعث افزایش میزان انتقال داده‌ها و بهبود تجربه کاربری شده است.

بیشتر سیستم‌های پیشنهاددهنده فعلی، مبتنی بر نظریه گراف و روش‌های جبری هستند که مشکل اصلی‌شان این است که نمی‌توانند پیش از تکامل سیستم در زمان و به صورت طولانی مدت، پیشنهاددهای قابل قبولی ارائه دهند. راه‌حل جایگزین، استفاده از شبکه‌های دینامیک است که ژانگ و همکارانش^۲ (۲۰۱۶) در مقاله مدل‌سازی دینامیک و آنالیز شبیه‌سازی سیستم پیشنهاددهی مبتنی بر شبکه فیلم-کاربر با تشکیل شبکه فیلم-کاربر و بررسی فیلم‌های تماشا شده مبتنی بر جستجو و پیشنهاد شده به کاربر، پیشنهاد شده است. به طور خلاصه، براساس میزان تماشای یک فیلم توسط کاربران، جستجوهای آن‌ها، مدل‌های پیشنهاددهی تصادفی و... مدلی پویا و دینامیک برای پیشنهاددهی تولید شده است که بازدهی بسیار خوبی در یک زمان کوتاه دارد. از الگوریتم‌های داده‌کاوی و انواع دسته‌بندی‌ها برای پیش‌بینی و میزان دقت در سیستم‌های توصیه‌کننده فیلم استفاده شده است. یکی از دسته‌بندی‌ها تئوری بیزین^۳ می‌باشد که در پژوهش

1. Lin & et al.

2. Zhang & et al.

3. Bayes Theorem

زهره میثاقیان و همکاران (۱۳۹۵) استفاده شده است و برای نشان دادن کارایی بهتر، آن را با دسته‌بندی‌های دیگر مقایسه کردیم. تئوری بیزین از بین دسته‌بندی‌های درخت تصمیم^۱، نزدیک‌ترین همسایه، شبکه عصبی چندلایه و ماشین بردار پشتیبان، بیشترین دقت را نشان داد. محبوب‌ترین سیستم‌های پیشنهاددهنده، الگوریتم‌های فیلترینگ محتوا محور و مشارکتی هستند. روش مشارکتی از دو مشکل عمده رنج می‌برد: مشکل اول شروع سرد کاربران و مشکل دوم نیز مسأله اعتماد کاربران وفادار است. حیدری و همکاران (۱۳۹۶) در مقاله خود با ترکیب روش محتوا محور و فیلترینگ مشارکتی، یک سیستم توصیه‌کننده طراحی کرده‌اند که علاوه بر حل مشکل شروع سرد، مسأله اعتماد و در نظر گرفتن کاربران وفادار را نیز پوشش می‌دهند. روش پیشنهادی این محققان شامل سه مرحله می‌باشد:

- (۱) خوشه‌بندی اولیه کلیه کاربران و تعیین خوشه مناسب برای کاربر جدید،
 - (۲) تعیین وزن مناسب برای هر کدام از ویژگی‌های کاربران خوشه مورد نظر و تعیین همسایگان نزدیک کاربر جدید،
 - (۳) تشکیل ماتریس مجاورتی امتیازات همسایگان نزدیک به فیلم‌های موجود و محاسبه امتیاز کاربر جدید به هر فیلم. جهت انجام آزمایش‌ها از بانک اطلاعاتی مووی‌لنز^۲ استفاده شده و نتایج بدست آمده نشان‌دهنده بهبود مدل، نسبت به روش‌های پایه مانند نایو بیز^۳ بوده است.
- کاتایارا و ورمای^۴ (۲۰۱۷) در مقاله پژوهشی خود یک سیستم پیشنهاددهنده جدید را با استفاده از روش خوشه‌بندی کا-مینز^۵ و الگوریتم جستجوی کوکو^۶ پیشنهاد داده‌اند که این الگوریتم بر روی مجموعه داده مووی لنز نیز پیاده‌سازی شده است. نتایج تجربی حاصل در این مقاله بیان می‌کند که این روش پیشنهادی می‌تواند عملکرد بالایی را در زمینه قابلیت اطمینان، کارایی و ارائه توصیه‌های فیلم دقیق در مقایسه با روش‌های موجود ارائه دهد.
- فیلترینگ مبتنی بر محتوا، یکی از موفق‌ترین الگوریتم‌های پیشنهاددهنده مبتنی بر همبستگی

-
1. Decision tree
 2. Movielens.org
 3. Naive Bayes
 4. Kataraya & Verma
 5. K-means Clustering
 6. Cuckoo search

بین محتواها است. این رویکرد، از اطلاعات مربوط به آیتم‌ها، به عنوان ویژگی‌هایی برای محاسبه شباهت بین آیتم‌ها استفاده می‌کند. سان و همکاران^۱ (۲۰۱۷) در مقاله خود یک روش جدید مبتنی بر محتوا ارائه کرده‌اند که از شبکه چند عاملی استفاده می‌کند که به طور مؤثر، منعکس‌کننده ویژگی‌های مختلف آیتم‌ها در هنگام محاسبات برای توصیف آن‌ها است. علاوه بر این، روش پیشنهادی، برای تعیین الگوهای ساختاری تعامل میان آیتم‌ها از تکنیک‌های خوشه‌بندی استفاده می‌کند. این رویکرد با رویکردهای موجود با استفاده از داده‌های مووی‌لنز مقایسه شده و نتایج نشان‌دهنده دقت و استحکام بیشتر این رویکرد نسبت به آن‌ها است. نکته قابل توجه در این رویکرد، مبتنی بودن آن بر داده‌های گذشته خود کاربر است.

اینان و همکاران^۲ (۲۰۱۸) در مقاله خود به پیشنهاد سیستم پیشنهاددهنده فیلم مبتنی بر هدف پرداخته‌اند. سیستم‌های پیشنهاددهنده عموماً از کمبود داده رنج می‌برند. به همین علت، سیستم پیشنهاددهنده پیشنهادی، ترکیبی از ویژگی‌های فیلم (کارگردان، بازیگران، ژانر و...) با الگوریتم فیلترینگ مشارکتی است و امتیازهای دقیق‌تری بر مبنای ویژگی‌های فیلم را نسبت به الگوریتم مبتنی بر محتوا ارائه می‌دهد. در این روش، از همبستگی پیرسون برای بخش‌های مربوط به الگوریتم مشارکتی استفاده شده است. این روش بر روی داده‌های وبسایت مووی‌لنز ارزیابی شده و نتایج تجربی نشان داد که بهتر از روش‌های موجود در ادبیات نظری عمل می‌کند. همچنین نتایج نشان می‌دهد که عملکرد کلی سیستم وقتی اطلاعات محتوا با بهره‌گیری از الگوریتم فیلترینگ مشارکتی تقویت می‌شود، افزایش می‌یابد.

محمدرضایی و روانمهر (۱۳۹۹) نیز در پژوهش خود یک روش جدید مبتنی بر ترکیب روابط اعتماد بین کاربران و شباهت علایق آنها برای پیشنهاددهی دوره‌های آموزشی پیشنهاد دادند. در این پژوهش ابتدا کاربران با استفاده از روش فازی c-means خوشه‌بندی شده و پس از آن با استفاده از قوانین انجمنی در هر خوشه N آیتم آموزشی با رتبه بالاتر به عنوان آیتم‌های مورد علاقه کاربر به وی پیشنهاد می‌شود.

این پژوهش در پی آن است که با تکیه بر داده‌های مشاهده ویدئوی کاربران و با استفاده از شاخص‌های تحلیل شبکه اجتماعی، روش جدیدی از نوع فیلترینگ مبتنی بر محتوا را ارائه دهد که

1. Son & et al.
2. Inan & et al.

برای پیشنهاد مناسب محتوا به کاربران نیازمند حداقل اطلاعات مورد نیاز یعنی مدت زمان مشاهده ویدئو توسط کاربر است. در ادامه، نحوه عمل این الگوریتم در گام‌های مختلف تشریح می‌شود.

۳. روش پژوهش

روش پیشنهادی این پژوهش جهت طراحی سیستم پیشنهاددهی فیلم براساس تاریخچه مشاهدات کاربران و با بهره‌گیری از شاخص‌های تحلیل شبکه اجتماعی شامل گام‌های زیر است:

گام اول: تشکیل ماتریس درصد تماشای هر فیلم توسط هر کاربر

در این گام و براساس تاریخچه تماشای فیلم توسط کاربران، ماتریس درصد مشاهده هر فیلم توسط هر کاربر مانند جدول ۱ تشکیل می‌شود.

جدول ۶- ماتریس درصد تماشای هر فیلم توسط هر کاربر

درصد تماشای کاربر	مدت زمان کل فیلم (ثانیه)	مدت زمان تماشای فیلم توسط کاربر (ثانیه)	آیدی کاربر	آیدی فیلم
X1/Y1	Y1	X1	U1	F1
X2/Y2	Y2	X2	U2	F2
X3/Y1	Y1	X3	U3	F1
X4/Y3	Y3	X4	U4	F3

گام دوم: تشکیل ماتریس میزان شباهت دو فیلم از نگاه هر کاربر

در این مرحله و براساس داده‌های گام اول به محاسبه میزان شباهت بین دو فیلم (DS^1) با استفاده از رابطه (۱) می‌پردازیم:

$$DS_{ij} = \frac{2n_{ij}}{n_i + n_j} \quad (1)$$

در این رابطه n_i درصد مشاهده فیلم i توسط کاربر n ، n_j درصد مشاهده فیلم j توسط کاربر n و n_{ij} اشتراک مشاهده دو فیلم i و j توسط کاربر n می‌باشد که براساس منطق فازی و از طریق رابطه (۲) محاسبه می‌گردد:

$$n_i \cap n_j = \min(n_i, n_j) \quad (2)$$

خروجی این گام ماتریسی مانند جدول ۲ می‌باشد.

جدول ۷- ماتریس میزان شباهت هر دو فیلم براساس مشاهده هر کاربر

i	j	U1	U2	U3	...
F1	F2	DS_{F1F2}	DS_{F1F2}	DS_{F1F2}	...
F1	F3	DS_{F1F3}	DS_{F1F3}	DS_{F1F3}	...
F2	F3	DS_{F2F3}	DS_{F2F3}	DS_{F2F3}	...

گام سوم: تشکیل ماتریس مربعی میزان شباهت فیلم‌ها

در این گام و براساس مقادیر DS محاسبه شده در گام گذشته، به محاسبه میانگین شباهت هر دو فیلم (AS) بر مبنای رابطه (۳) می‌پردازیم:

$$AS_{ij} = \frac{\sum_{n=1}^p DS_{ij}(n)}{p} \quad (3)$$

در این رابطه، p تعداد کاربر و $DS_{ij}(n)$ میزان شباهت فیلم i و j بر مبنای مشاهده کاربر n ام است.

در نهایت، ماتریس شباهت بین دو فیلم، براساس این رابطه بدست می‌آید که میزان شباهت بین دو آیتم را از ۰ (به معنای عدم وجود شباهت) تا ۱ (به معنای شباهت کامل دو آیتم) نشان داده و مانند خروجی جدول ۸ است.

جدول ۸- ماتریس مربعی میزان ارتباط (شباهت) فیلم‌ها به یکدیگر

Films	F1	F2	F3	...
F1	1	AS_{F1F2}	AS_{F1F3}	...
F2	AS_{F2F1}	1	AS_{F2F3}	...
F3	AS_{F3F1}	AS_{F3F2}	1	...
....	...	...	...	1

گام چهارم: تشکیل گراف روابط میان فیلم‌ها

در این گام براساس محاسبات گام سوم، به تشکیل گراف ارتباطی فیلم‌ها می‌پردازیم. برای تشکیل این گراف از فیلم‌ها به عنوان گره و از مقادیر AS برای ترسیم یال‌ها و ضخامت آن‌ها استفاده می‌شود.

گام پنجم: محاسبه شاخص‌های مرکزیت هر فیلم

در این گام بر مبنای گراف ترسیمی، سه شاخص اندازه‌گیری مرکزیت درجه^۱، مرکزیت بینیت^۲ و مرکزیت نزدیکی^۳ محاسبه می‌گردند.

مرکزیت درجه برای فیلم i با D_C نمایش داده شده و به صورت زیر محاسبه می‌شود.

$$D_C^w(i) = \frac{\sum_j^N w_{ij}}{n-1} \quad (۴)$$

در این رابطه، w ماتریس همبستگی وزنی^۴ و w_{ij} هم وزن یالی است که فیلم i را به فیلم j متصل می‌کند. در اینجا n هم نمایانگر تعداد گره‌های گراف است. مقدار بالای D_C به معنای اشتراک بیشتر آن فیلم با سایر فیلم‌هاست.

محاسبه D_C محدود به گره‌هایی است که مستقیماً به گره مورد نظر متصل شده‌اند و این معیار، گره‌هایی را که به صورت غیر مستقیم به گره مورد نظر متصل شده‌اند، در نظر نمی‌گیرد. مرکزیت نزدیکی برای گره i با C_C نمایش داده شده و به صورت زیر محاسبه می‌شود.

$$C_C^w(i) = \frac{n-1}{\sum_{j=1}^n d^w(i,j)} \quad (۵)$$

در این رابطه، $d^w(i,j)$ وزن کوتاه‌ترین مسیر بین گره i و j است و n هم نمایانگر تعداد گره‌های گراف (فیلم‌ها) است. مقدار بالای C_C برای یک فیلم، به معنای فاصله زیاد آن فیلم تا مرکز گراف است. مرکزیت بینیت برای گره i با B_C نمایش داده شده و به صورت زیر محاسبه می‌شود.

$$B_C^w(i) = \sum_{j < k} \frac{g_{jk}^w(i)}{g_{jk}^w} \quad (۶)$$

در این رابطه، g_{jk} تعداد تمامی مسیرهای وزن‌دار بین گره j و k است و $g_{jk}^w(i)$ تعداد تمامی مسیرهای وزن‌دار بین گره j و k بوده که از گره i عبور می‌کنند.

برای ترکیب این سه شاخص، شاخصی به نام مرکزیت میانگین^۵ با $AC(i)$ تعریف شده است که حاصل میانگین سه شاخص قبلی برای گره i می‌باشد.

$$AC(i) = \frac{D_C^w(i) + C_C^w(i) + B_C^w(i)}{3} \quad (۷)$$

1. Degree centrality
2. Betweenness centrality
3. Closeness centrality
4. Weighted adjacency matrix
5. Average centrality

خروجی این گام ماتریسی مانند جدول ۴ می باشد.

جدول ۹- نمونه ای از محاسبه شاخص های مرکزیت برای فیلم ها

Films ID	مرکزیت درجه D_C	مرکزیت نزدیکی C_C	مرکزیت بینیت B_C	مرکزیت میانگین AC
F1	۰,۰۵۳۶	۰,۸۲۲۹	۰,۵۶۲۲	۰,۴۷۹۶
F2	۰,۰۳۴۵	۰,۶۴۷۵	۰,۱۲۰۴	۰,۲۶۷۵
F3	۰,۰۴۹۷	۰,۸۴۹۵	۰,۵۲۰۷	۰,۴۷۳۳
....	۰,۰۴۳۸	۰,۷۷۴۵	۰,۴۲۶۰	۰,۴۱۴۸

گام ششم: خوشه بندی^۱ فیلم ها

خوشه بندی فیلم ها صحت پیش بینی های رتبه بندی کاربران را بهبود می بخشد و سرعت محاسبات آنلاین را به میزان قابل توجهی افزایش می دهد. هدف این مرحله، قرار دادن فیلم های مشابه در همسایگی یکدیگر و در یک خوشه است که برای انتخاب آسان تر کاندیدهای پیشنهاد به کاربر استفاده می شود.

تجزیه و تحلیل ماژولار^۲، یکی از محبوب ترین روش ها برای تشخیص ساختارهای جامعه در یک شبکه است. در این شبکه ها، ارتباطات متراکم میان گره ها در درون یک خوشه شکل می گیرد و ارتباطات ضعیف منجر به خوشه بندی های مختلف می شود. نیومن^۳ بر مبنای این مدل تجزیه و تحلیل، Q را به صورت زیر تعریف کرده است (نیومن، ۲۰۰۶).

$$Q = \sum_r (e_{rr} - a_r^2) \quad (۸)$$

در این رابطه، e_{rr} بیانگر تعداد لینک های متصل کننده (آشکار و پنهان) گره ها به یکدیگر است (که با تعریف یال تفاوت دارد) و a_r تعداد لینک هایی را می شمارد که یک یا دو سمت آن به یک گره متصل شده است. مقدار بالای Q به معنای ساختار مستحکم جوامع موجود در شبکه است.

خوشه بندی به روش ماژولار، یکی از بهترین الگوریتم های خوشه بندی شبکه محور است که نتایج قابل قبولی را در سیستم های پیشنهاددهی ارائه داده است (جلیلوند و همکاران، ۱۳۹۹). این

1. Clustering
2. Modularity analysis
3. Newman

شیوه برای تعیین گروه‌هایی از گره‌ها شامل کاربران و فیلم‌هایی که با یکدیگر ارتباط دارند، استفاده می‌شود. در سیستم‌های پیشنهاددهی، فیلم‌هایی که به طور مستقیم به یکدیگر متصل نشده‌اند، ممکن است در یک گروه قرار گیرند.

• پیشنهاددهی (براساس امتیازهای محاسبه شده هر آیتم در هر خوشه)

آیتم‌های موجود در خوشه ترجیحات هر کاربر، کاندیدیهایی برای پیشنهاد به کاربر هستند که مبنای اولویت بندی آن را اطلاعات مربوط به تجربیات گذشته کاربر از نظر مشاهده تعیین می‌کند. ابتدا میزان مرکزیت آیتم‌ها در شبکه محاسبه می‌شود تا ساختار ارتباط و شباهت میان آیتم‌ها تشکیل شود. سپس برای پیشنهاددهی شخصی‌سازی شده، فاصله مهم‌ترین آیتم‌های هر خوشه را از آیتم‌های مورد علاقه کاربر براساس میزان مرکزیت محاسبه شده و نزدیک‌ترین آیتم‌ها به کاربر پیشنهاد می‌شود.

گام هفتم: تشکیل ماتریس ترجیحات و غیر ترجیحات برای هر کاربر

تاریخچه فعالیت هر کاربر شامل اطلاعاتی از مشاهدات او است که می‌تواند بیانگر ترجیحات^۱ و غیر ترجیحات^۲ کاربر باشد. در این پژوهش مبنای تقسیم ترجیحات و غیر ترجیحات درصد مشاهده آن فیلم توسط کاربر می‌باشد، بدین صورت که هر فیلم با درصد مشاهده بیش از ۵۰ درصد در لیست ترجیحات کاربر و با درصد مشاهده کم‌تر از ۵۰ درصد در لیست غیر ترجیحات قرار می‌گیرند.

جدول ۱۰- ماتریس ترجیحات و غیر ترجیحات کاربران

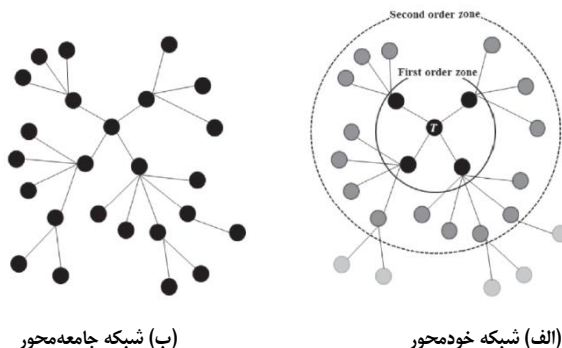
User	لیست ترجیحات (فیلم‌های مورد علاقه)			لیست غیر ترجیحات (فیلم‌های ضد علاقه)		
U1	Film1	Film4	Film43	Film2	Film7	Film10
U2	Film2	Film6		Film5	Film13	Film4
...		...				

برای هر یک از فیلم‌ها، مقدار C_{EF} براساس گره‌های محوری (یعنی ترجیحات و غیر ترجیحات کاربر) محاسبه گردیده که در ماتریسی مانند جدول ۵ نمایش داده می‌شوند.

1. Preference
2. Non-preference

گام هشتم: محاسبه شاخص مرکزیت خودمحور^۱

برای تحلیل یک شبکه، دو نوع رویکرد جامعه‌محور^۲ و خودمحور وجود دارد که در شکل ۱ نمایش داده شده است.



شکل ۱- مثالی از شبکه‌های خودمحور و جامعه‌محور

شاخص‌هایی که در گام پنجم محاسبه شدند، شاخص‌های جامعه‌محور هستند. در واقع، مقدار بالای مرکزیت میانگین در معادله (۷)، به معنای این است که آن فیلم عامه‌پسندتر از سایر آیتم‌هاست. این روش برای ارائه پیشنهاد به کاربرانی که هنوز هیچ‌گونه تاریخچه‌ای از عملکردشان وجود ندارد، استفاده می‌شود. اما برای پیشنهاددهی شخصی‌سازی شده، باید داده‌های مربوط به ترجیحات هر کاربر، در پیشنهادات ارائه شده به او اثرگذار باشد.

وقتی یک فیلم، مورد علاقه کاربر شناسایی می‌شود، آن فیلم به عنوان هدف^۳ (گره محوری^۴) شناخته می‌شود و سیستم پیشنهاددهنده، به دنبال سایر آیتم‌های مشابه آن آیتم می‌گردد. آیتم‌های کاندید از اعضای هم‌خوشه ترجیحات کاربر انتخاب می‌شوند.

مطابق شکل ۲، در این شبکه خودمحور، گره T به عنوان گره محوری (ترجیح کاربر) در نظر گرفته شده است. لایه اول پیرامون این گره مستقیماً به آن متصل شده‌اند؛ لایه دوم، حداقل با یک واسطه به گره T متصل شده‌اند و بدین صورت رابطه تمامی گره‌های شبکه با گره T مشخص می‌شود.

1. Ego-focused centrality
2. Sociocentric
3. Target
4. Ego Node

با این توصیفات، مرکزیت خودمحور با C_{EF} به صورت زیر تعریف می‌شود.

$$C_{EF}(i, T) = \frac{AC(i)}{d(i, T)} \quad (8)$$

در این رابطه، $AC(i)$ از رابطه شماره (۷) و فاصله بین گره i و T براساس تعداد لینک‌های بین آن دو، یعنی فاصله بین دو گره محاسبه خواهد شد.



شکل ۲- یک شبکه خودمحور که گره محوری آن، T است.

به عنوان مثال، در جدول ۱۱ با اینکه عدد مرکزیت میانگین گره i_1 پایین‌تر از گره i_3 است، اما به دلیل فاصله کم‌تر آن با گره محوری i_1 ، مرکزیت خودمحوری آن بیشتر شده است.

جدول ۱۱- نحوه محاسبه مرکزیت خودمحوری گره‌های i_1 تا i_5

مرکزیت خودمحوری	فاصله گره	مرکزیت میانگین AC	مرکزیت بینیت B_C	مرکزیت نزدیکی C_C	مرکزیت درجه D_C	گره
۰,۷۳۹	۱	۰,۷۳۹	۰,۵۶۷	۰,۸۱۷	۰,۸۳۳	i_1
۰,۴۴۵	۲	۰,۸۸۹	۱	۱	۰,۶۶۷	i_2
۰,۲۸۸	۳	۰,۸۶۵	۰,۷۱۹	۰,۸۷۵	۱	i_3
۰,۱۷۵	۳	۰,۵۲۴	۰,۱۶۳	۰,۷۴۲	۰,۶۶۷	i_4
۰,۲۱۶	۱	۰,۲۱۶	۰	۰,۴۸۰	۰,۱۶۷	i_5

گام نهم: تشکیل ماتریس شاخص مرکزیت خودمحور براساس ترجیحات و غیر ترجیحات و

محاسبه پیشنهاددهی خودمحور

در این گام می‌بایست به تشکیل ماتریس شاخص مرکزیت خودمحور براساس لیست ترجیحات و غیر ترجیحات هر کاربر پرداخت. فیلم‌هایی که محاسبه شاخص خودمحوری در مقایسه با آن‌ها صورت می‌گیرد، مهم‌ترین فیلم‌های هر خوشه هستند که تعداد آن‌ها می‌تواند در هر سیستم پیشنهاددهی متفاوت باشد. نتایج حاصل از این گام در جدول ۱۲ نمایش داده شده است.

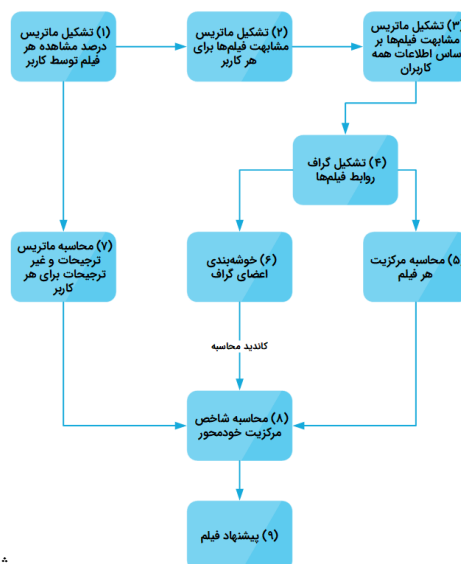
جدول ۱۲- مثالی از محاسبه امتیازهای پیشنهاددهی خودمحمور

User ID						
فیلم کاندید	لیست ترجیحات			لیست غیر ترجیحات		RS_{EF}
	F10	F11	F9	F3	F13	
F1	$C_{EF}(F10, F1)$	$C_{EF}(F11, F1)$	$C_{EF}(F9, F1)$	$C_{EF}(F3, F1)$	$C_{EF}(F13, F1)$	۰
F2	$C_{EF}(F10, F2)$	$C_{EF}(F11, F2)$	$C_{EF}(F9, F2)$	$C_{EF}(F3, F2)$	$C_{EF}(F13, F2)$	۰,۱۳۳۷
F7	$C_{EF}(F10, F7)$	$C_{EF}(F11, F7)$	$C_{EF}(F9, F7)$	$C_{EF}(F3, F7)$	$C_{EF}(F113, F7)$	-۰,۱۴۴۲
....						

در جدول فوق، امتیاز پیشنهاددهی خودمحمور با RS_{EF} نمایش داده می‌شود که نحوه محاسبه آن براساس ترجیحات و غیر ترجیحات کاربر به صورت رابطه (۹) است.

$$RS_{EF}(i) = (\sum_{j=1} C_{EF-P}(i, T_j)) - \sum_{k=1} C_{EF-NP}(i, T_k) \quad (9)$$

در این رابطه C_{EF-P} همان C_{EF} مربوط به ترجیحات کاربر و C_{EF-NP} همان C_{EF} مربوط به غیر ترجیحات کاربر است. همچنین T_j مجموعه ترجیحات کاربر و T_k مجموعه غیر ترجیحات کاربر است. در واقع، مثبت بودن این رابطه به معنای مرتبط بودن فیلم i با سلیق کاربر است. هرچه مقدار این رابطه بیشتر باشد، نشان‌دهنده آن است که آن فیلم به میزان بیشتری مورد پسند کاربر است و هرچه کم‌تر و یا منفی باشد، به معنای فاصله بیشتر آن فیلم با سلیق کاربر خواهد بود. شکل ۳ نمایشی از گام‌های این پژوهش برای پیشنهاددهی آیت‌ها می‌باشد.


<http://stjm.gom.ac.ir>

شکل ۴- گام‌های مدل تحقیق برای پیشنهاد فیلم

۳-۱. داده‌های مورد استفاده

داده‌های مورد بررسی در این پژوهش، شامل اطلاعات مشاهده کاربران سامانه ویدئوی درخواستی سینما مارکت در ۱۰ ماه و بین بازه زمانی ۱۳۹۷/۰۱/۰۱ تا ۱۳۹۷/۰۹/۳۰ بوده و شامل ۸۰ عنوان فیلم و ۳۲۸ کاربر می‌باشد. سایت سینما مارکت برای پاسخ‌گویی به نیاز دانلود قانونی محتوای تصویری و با تاکید بر محتوای مستند در سال ۱۳۹۳ تأسیس گردید. این پلتفرم، محتوای ویدئویی را در موضوعات مختلف و به دو صورت مشاهده آنلاین و دانلود در اختیار مخاطبان قرار می‌دهد.

۴. یافته‌های پژوهش

• ماتریس درصد تماشای هر فیلم توسط هر کاربر

گام اول این پژوهش، تشکیل جدول رابطه هر فیلم با هر کاربر و درصد تماشای آن فیلم توسط آن کاربر است که حاصل تقسیم مدت زمان تماشای فیلم توسط کاربر بر کل مدت زمان فیلم است. بخش‌هایی از این اطلاعات در جدول ۱ آمده است.

جدول ۱- بخشی از داده‌های تماشای فیلم‌ها توسط کاربران

درصد تماشای کاربر	مدت زمان کل فیلم (ثانیه)	مدت زمان تماشای فیلم توسط کاربر (ثانیه)	شماره کاربر	شماره فیلم
۰,۹۸	۲۴۵۸	۲۴۰۰	۵۹۶۳۵	۱۴۰۱
۰,۲۲	۲۴۶۹	۵۴۰	۴۴۵۳۰	۶۳۵۲
۰,۶۸	۲۴۶۹	۱۶۸۰	۷۱۳۲۶	۶۳۵۲
۰,۹۹	۲۵۳۳	۲۵۲۰	۵۸۱۹۷	۲۳۶
۰,۴۰	۲۴۷۲	۱۰۲۰	۵۹۲۴۳	۵۳
۰,۸۲	۲۵۷۲	۲۱۰۰	۸۸۴	۵۳
۰,۹۶	۲۵۷۲	۲۴۶۰	۵۹۱۰۰	۵۳
...	...	...	...	...

• ماتریس میزان شباهت دو فیلم از نگاه هر کاربر

گام دوم شامل تشکیل ماتریس شباهت بین دو فیلم از نظر هر کاربر، براساس رابطه شماره (۱) می‌باشد. بخشی از این ماتریس ۸۰ در ۳۲۴ بُعدی، در جدول ۲ ارائه شده است. با توجه به رابطه شماره (۱)، در حالتی که هم صورت و هم منخرج کسر برابر ۰ بودند، عدد ۱- در ماتریس قرار

گرفته است که در محاسبات بعدی، این اعداد در نظر گرفته نشوند؛ چراکه این حالت به معنای عدم تماشای هیچ یک از دو فیلم توسط کاربر است و لذا نمی‌تواند معیاری برای فهم شباهت یا عدم شباهت دو فیلم به یکدیگر باشد. همچنین عدد ۰ در این ماتریس به معنای عدم تماشای یکی از دو فیلم توسط کاربر و در نتیجه عدم شباهت آن دو فیلم از نظر آن کاربر است.

جدول ۲- بخشی از ماتریس میزان شباهت دو فیلم از نگاه هر کاربر

		کاربر ۷۳۰۹۴	کاربر ۷۰۳۲۴		کاربر ۵۳۸۳	کاربر ۴۴۸۶
فیلم ۶۱	فیلم ۵۱	-۱	-۱	...	-۱	۰
فیلم ۵۴۹۰	فیلم ۶۱	-۱	-۱	...	-۱	۰,۸۲
فیلم ۵۵۹۱	فیلم ۶۱	-۱	-۱	...	-۱	۰,۹۹
...	...	...	...	...	...	...
فیلم ۵۳	فیلم ۲۳۶	-۱	-۱	...	-۱	-۱
فیلم ۵۹۰	فیلم ۵۹۰	-۱	-۱	...	۱	۱
فیلم ۶۵۱۸	فیلم ۵۹۰	-۱	۰	...	۰,۷۱	۰,۵۳

● ماتریس مربعی میزان شباهت فیلم‌ها

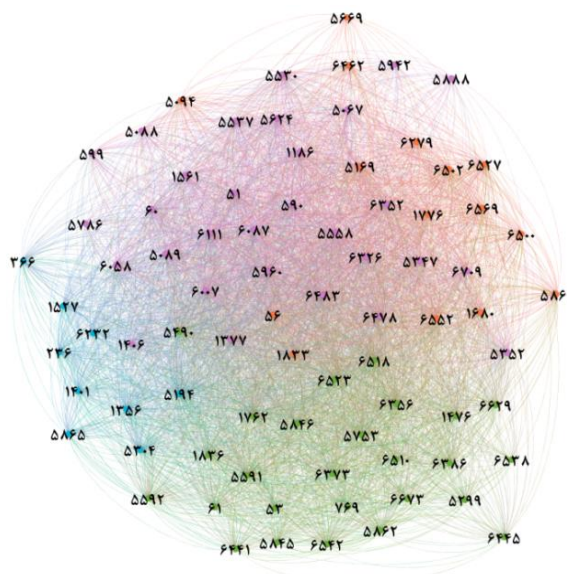
در گام سوم این بخش براساس رابطه شماره (۲)، با جمع کردن تمامی اعداد مثبت هر سطر و تقسیم آن‌ها بر تعدادشان، میزان ارتباط (شباهت) دو فیلم موجود در هر سطر در بازه ۰ تا ۱ تعیین می‌شود. جدول ۳ این رابطه را به صورت یک ماتریس مربعی نمایش می‌دهد. در این جدول بخش کوچکی از این ماتریس ۸۰ در ۸۰ بُعدی نمایش داده شده است.

جدول ۳- ماتریس مربعی میزان ارتباط (شباهت) فیلم‌ها به یکدیگر

	فیلم ۶۷۰۹	فیلم ۶۵۱۸	...	فیلم ۶۱	فیلم ۵۱
فیلم ۶۷۰۹	۰,۰۹۰۱	۰,۰۳۸۱	...	۰,۰۱۹۴	۱
فیلم ۶۵۱۸	۰,۰۲۸۵	۰,۰۴۳۹	...	۱	۰,۰۱۹۴
...	...	...	...	...	...
فیلم ۶۵۱۸	۰,۰۵۳۷	۱	...	۰,۰۴۳۹	۰,۰۳۸۱
فیلم ۶۷۰۹	۱	۰,۰۵۳۷	...	۰,۰۲۸۵	۰,۰۹۰۱

● تشکیل گراف روابط میان فیلم‌ها

در این گام، براساس نتایج جدول ۳ گراف ارتباطی فیلم‌ها ترسیم شده‌اند که نتایج آن در شکل ۱ نمایش داده شده است.



شکل ۱- گراف مبتنی بر مرکزیت و نشان‌دهنده ارتباط (شباهت) میان فیلم‌ها

● محاسبه شاخص‌های مرکزیت هر فیلم

جدول ۴ بخشی از نتایج حاصل از محاسبه شاخص‌های مرکزیت هر فیلم را نشان می‌دهد.

جدول ۴- محاسبه پارامترهای مرکزیت هر فیلم

شماره فیلم	مرکزیت درجه	مرکزیت نزدیکی	مرکزیت بینیت	مرکزیت میانگین
۵۱	۰,۰۵۳۶	۰,۸۲۲۹	۰,۵۶۲۲	۰,۴۷۹۶
۵۳	۰,۰۳۴۵	۰,۶۴۷۵	۰,۱۲۰۴	۰,۲۶۷۵
۵۶	۰,۰۴۹۷	۰,۸۴۹۵	۰,۵۲۰۷	۰,۴۷۳۳
۶۰	۰,۰۴۳۸	۰,۷۷۴۵	۰,۴۲۶۰	۰,۴۱۴۸
۶۱	۰,۰۴۳۰	۰,۷۳۸۳	۰,۲۷۱۱	۰,۳۵۰۸
...	...	...	...	...

● خوشه‌بندی گراف فیلم‌ها

در این گام خوشه‌بندی فیلم‌ها با استفاده از تجزیه و تحلیل ماژولار انجام می‌شود که نتایج آن در جدول ۵ و در رنگ‌های موجود در شکل ۱ ارائه شده است.

آیتم‌های حاضر در خوشه‌ای که بیشترین ترجیحات کاربر را شامل می‌شود، به عنوان کاندید پیشنهاد برگزیده می‌شوند.

جدول ۵- خوشه‌بندی فیلم‌ها براساس روش مازولار

خوشه اول		خوشه دوم		خوشه سوم		خوشه چهارم	
فیلم ۲۳۶	فیلم ۳۶۶	فیلم ۵۱	فیلم ۶۰	فیلم ۵۳	فیلم ۶۱	فیلم ۵۶	فیلم ۱۶۸۰
فیلم ۱۳۵۶	فیلم ۱۴۰۱	فیلم ۵۹۰	فیلم ۵۹۹	فیلم ۷۶۹	فیلم ۱۴۷۶	فیلم ۱۷۷۶	فیلم ۱۸۳۳
فیلم ۱۵۲۷	فیلم ۵۱۹۴	فیلم ۱۱۸۶	فیلم ۱۳۷۷	فیلم ۱۷۶۲	فیلم ۱۸۳۶	فیلم ۵۰۹۴	فیلم ۵۱۶۹
فیلم ۵۳۰۴	فیلم ۵۸۶۵	فیلم ۱۴۰۶	فیلم ۱۵۶۱	فیلم ۵۲۹۹	فیلم ۵۴۹۰	فیلم ۵۶۶۹	فیلم ۵۸۶۱
فیلم ۶۲۳۲		فیلم ۵۰۶۷	فیلم ۵۰۸۸	فیلم ۵۵۹۱	فیلم ۵۵۹۲	فیلم ۶۲۷۹	فیلم ۶۴۶۲
		فیلم ۵۰۸۹	فیلم ۵۳۴۷	فیلم ۵۷۵۳	فیلم ۵۸۴۵	فیلم ۶۵۰۰	فیلم ۶۵۰۲
		فیلم ۵۳۵۲	فیلم ۵۵۳۰	فیلم ۵۸۴۶	فیلم ۵۸۶۲	فیلم ۶۵۲۷	فیلم ۶۵۵۲
		فیلم ۵۵۲۷	فیلم ۵۵۵۸	فیلم ۶۳۵۶	فیلم ۶۳۷۳	فیلم ۶۵۶۹	
		فیلم ۵۶۳۴	فیلم ۵۷۸۶	فیلم ۶۳۸۶	فیلم ۶۴۴۱		
		فیلم ۵۸۸۸	فیلم ۵۹۴۲	فیلم ۶۴۴۵	فیلم ۶۵۱۰		
		فیلم ۵۹۶۰	فیلم ۶۰۰۷	فیلم ۶۵۱۸	فیلم ۶۵۲۳		
		فیلم ۶۰۵۸	فیلم ۶۰۸۷	فیلم ۶۵۳۸	فیلم ۶۵۴۲		
		فیلم ۶۱۱۱	فیلم ۶۳۲۶	فیلم ۶۶۲۹	فیلم ۶۶۷۳		
		فیلم ۶۳۵۲	فیلم ۶۴۷۸				
		فیلم ۶۴۸۳	فیلم ۶۷۰۹				

• تشکیل ماتریس ترجیحات و غیر ترجیحات کاربران

در این گام لیست ترجیحات و غیر ترجیحات کاربران بر مبنای میزان تماشای هر فیلم توسط هر کاربر محاسبه شده که جدول ۶ بخشی از نتایج را نمایش می‌دهد.

جدول ۶- لیست ترجیحات و غیر ترجیحات برخی کاربران

لیست غیر ترجیحات (فیلم‌های ضد علاقه)	لیست ترجیحات (فیلم‌های مورد علاقه)				کاربران
فیلم ۶۷۰۹	فیلم ۵۰۸۸	فیلم ۲۳۶	فیلم ۵۵۹۱	فیلم ۶۵۲۳	کاربر ۴۴۸۶
	فیلم ۶۵۵۲	فیلم ۶۵۳۸	فیلم ۱۳۷۷	فیلم ۵۹۰	کاربر ۵۲۸۳
...	...	...	...	...	...
فیلم ۶۵۶۹	فیلم ۶۵۰۰	فیلم ۸۳۶	فیلم ۶۳۵۶	فیلم ۵۳۰۴	کاربر
فیلم ۵۵۹۲	فیلم ۵۰۸۹	فیلم ۲۳۶	فیلم ۶۳۵۲	فیلم ۵۵۲۷	کاربر

• محاسبه شاخص مرکزیت خودمحمور لیست ترجیحات و غیر ترجیحات یک کاربر

در این گام ابتدا براساس رابطه شماره (۸) این پژوهش، در جدول ۷ تا جدول ۱۰ مقدار C_{EF} برای ترجیحات و غیر ترجیحات یکی از کاربران (کاربر ۵۳۸۳) به عنوان نمونه محاسبه شده است. طبق مدل، فیلم‌هایی که خودمحموری در مقایسه با آن‌ها صورت می‌گیرد، مهم‌ترین فیلم‌های هر خوشه هستند که تعداد آن‌ها می‌توانند در هر سیستم پیشنهاددهی متفاوت باشد.

جدول ۷- محاسبه CEF با خودمحوری فیلم ۵۹۰ برای کاربر ۵۳۸۳

مرکزیت خودمحور	فاصله لینک ^۱ تا فیلم ۵۹۰	مرکزیت میانگین	شماره فیلم
۰,۴۷۹۶	۱	۰,۴۷۹۶	۵۱
۰,۱۳۳۷	۲	۰,۲۶۷۵	۵۳
...	...	...	...
۰,۱۴۴۲	۲	۰,۲۸۸۵	۶۶۷۳
۰,۲۷۴۳	۱	۰,۲۷۴۳	۶۷۰۹

جدول ۸- محاسبه CEF با خودمحوری فیلم ۱۳۷۷ برای کاربر ۵۳۸۳

مرکزیت خودمحور	فاصله لینک تا فیلم ۱۳۷۷	مرکزیت میانگین	شماره فیلم
۰,۴۷۹۶	۱	۰,۴۷۹۶	۵۱
۰,۲۶۷۵	۱	۰,۲۶۷۵	۵۳
...	...	...	...
۰,۲۸۸۵	۱	۰,۲۸۸۵	۶۶۷۳
۰,۲۷۴۳	۱	۰,۲۷۴۳	۶۷۰۹

جدول ۹- محاسبه CEF با خودمحوری فیلم ۶۵۳۸ برای کاربر ۵۳۸۳

مرکزیت خودمحور	فاصله لینک تا فیلم ۶۵۳۸	مرکزیت میانگین	شماره فیلم
۰,۲۳۹۸	۲	۰,۴۷۹۶	۵۱
۰,۲۶۷۵	۱	۰,۲۶۷۵	۵۳
...	...	...	...
۰,۱۴۴۲	۲	۰,۲۸۸۵	۶۶۷۳
۰,۱۳۷۱	۲	۰,۲۷۴۳	۶۷۰۹

جدول ۱۰- محاسبه CEF با خودمحوری فیلم ۶۵۵۲ برای کاربر ۵۳۸۳

مرکزیت خودمحور	فاصله لینک تا فیلم ۶۵۵۲	مرکزیت میانگین	شماره فیلم
۰,۴۷۹۶	۱	۰,۴۷۹۶	۵۱
۰,۱۳۳۷	۲	۰,۲۶۷۵	۵۳
...	...	...	...
۰,۲۸۸۵	۱	۰,۲۸۸۵	۶۶۷۳
۰,۲۷۴۳	۱	۰,۲۷۴۳	۶۷۰۹

● محاسبه شاخص نهایی پیشنهاددهی یا عدم پیشنهاددهی فیلم‌ها به کاربر

در جدول ۱۱، مقدار RS_{EF} طبق رابطه شماره (۹)، برای یکی از همان کاربران (کاربر ۵۳۸۳)، به صورت نمونه محاسبه شده است.

جدول ۱۱- محاسبه RSEF برای یک کاربر نمونه براساس لیست ترجیحات و غیر ترجیحات

RS_{EF}	لیست غیر ترجیحات		لیست ترجیحات	
	فیلم ۶۵۵۲	فیلم ۶۵۳۸	فیلم ۱۳۷۷	فیلم ۵۹۰
۰	۰,۴۷۹۶	۰,۴۷۹۶	۰,۴۷۹۶	۰,۴۷۹۶
۰,۱۳۳۷	۰,۱۳۳۷	۰,۱۳۳۷	۰,۲۶۷۵	۰,۱۳۳۷
...	...	...	...	...
-۰,۱۴۴۲	۰,۲۸۸۵	۰,۲۸۸۵	۰,۲۸۸۵	۰,۱۴۴۲
۰	۰,۲۷۴۳	۰,۲۷۴۳	۰,۲۷۴۳	۰,۲۷۴۳

طبق محاسبات صورت گرفته برای این نمونه، فیلم شماره ۵۳، در پیشنهاددهی به این کاربر نسبت به ۳ مورد دیگر ارجح است و فیلم شماره ۶۶۷۳ نباید به این کاربر پیشنهاد داده شود.

۴-۱. ارزیابی مدل

ارزیابی و اعتبارسنجی مدل پیشنهادی این پژوهش مهم‌ترین بخش این تحقیق می‌باشد، زیرا عدم کارایی مدل ارائه شده، به راحتی می‌تواند فرضیات این پژوهش را زیر سؤال ببرد. به منظور ارزیابی و اعتبارسنجی مدل این پژوهش و مقایسه آن با دیگر الگوریتم‌ها، ابتدا سه نمونه ۵۰، ۱۰۰ و ۲۰۰ نفری از ۳۲۸ کاربر مورد بررسی در این پژوهش به صورت کاملاً تصادفی انتخاب گردیده و داده‌های انتخاب شده به دو دسته ۷۰ درصد داده آموزشی و ۳۰ درصد آزمایش تقسیم شدند.

در این مرحله و برای ارزیابی دقت مدل، ۲ فیلم با بالاترین میزان تماشا و ۲ فیلم با پایین‌ترین میزان تماشا به عنوان لیست ترجیحات و غیر ترجیحات کاربر در نظر گرفته شده و ارزیابی براساس آنها صورت پذیرفته است. آزمون اعتبارسنجی مدل طبق مفروضات پژوهش به این صورت است که اگر میزان RS_{EF} طبق تعریف برای فیلم‌های مورد علاقه کاربر (ترجیحی) مثبت و برای فیلم‌های ضد علاقه کاربر (غیرترجیح) منفی بود، مدل درست عمل کرده و در غیر این صورت پیش‌بینی مدل اشتباه بوده است.

با توجه به طولانی بودن روند محاسبات و جداول مربوطه، محاسبات مربوط به یک کاربر (کاربر ۶۲۲۷۷) در جدول ۱۲ و جدول ۱۳ گزارش شده و در نهایت، مقایسه میان مدل پژوهش و

سه الگوریتم مرسوم پیش‌بینی با ناظر یعنی درخت تصمیم، نزدیک‌ترین همسایه و بیزین در سه حالت ۵۰، ۱۰۰ و ۲۰۰ کاربر در شکل ۲ ارائه شده است.

جدول ۱۲- محاسبه RSEF برای کاربر آزمایشی ۶۲۲۷۷ در حالت ۳۵ داده آموزشی

	لیست ترجیحات		لیست غیر ترجیحات		RS_{EF}
	فیلم ۶۰۰۷	فیلم ۵۰۸۹	فیلم ۱۵۶۱	فیلم ۵۰۹۴	
فیلم ۶۰۰۷	۰,۳۰۸۱	۰,۳۰۸۱	۰,۳۰۸۱	۰,۱۵۴۱	۰,۱۵۴۱
فیلم ۵۰۸۹	۰,۱۹۳۳	۰,۱۹۳۳	۰,۱۹۳۳	۰,۰۹۶۷	۰,۰۹۶۷
فیلم ۱۵۶۱	۰,۱۷۸۸	۰,۱۷۸۸	۰,۱۷۸۸	۰,۰۸۹۴	۰,۰۸۹۴
فیلم ۵۰۹۴	۰,۰۰۰۰	۰,۰۰۰۰	۰,۰۰۰۰	۰,۰۰۰۰	۰,۰۰۰۰

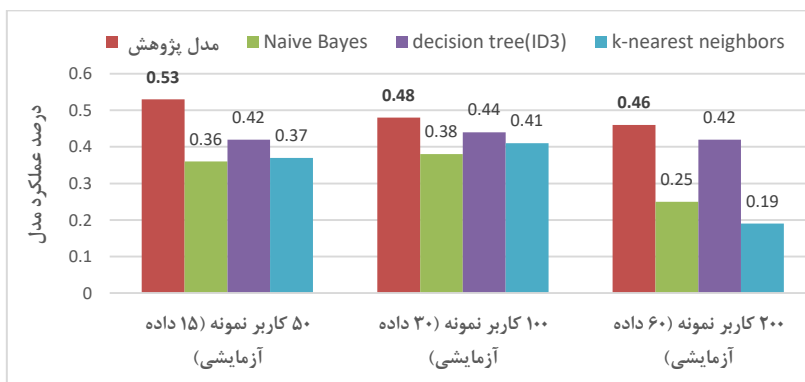
در حالت ۳۵ داده آموزشی، با توجه به مثبت بودن مقدار RS_{EF} برای دو فیلم اول، مدل توانسته لیست ترجیحات کاربر را به درستی تشخیص دهد، اما فیلم شماره ۱۵۶۱ که جزو غیر ترجیحات کاربر است، به عنوان لیست ترجیحات او شناخته شده است.

جدول ۱۳- محاسبه RSEF برای کاربر آزمایشی ۶۲۲۷۷ در حالت ۷۰ داده آموزشی

	لیست ترجیحات		لیست غیر ترجیحات		RS_{EF}
	فیلم ۶۰۰۷	فیلم ۵۰۸۹	فیلم ۱۵۶۱	فیلم ۵۰۹۴	
فیلم ۶۰۰۷	۰,۱۹۵۰	۰,۱۹۵۰	۰,۱۹۵۰	۰,۰۹۸۰	۰,۰۹۸۰
فیلم ۵۰۸۹	۰,۱۶۶۶	۰,۱۶۶۶	۰,۱۶۶۶	۰,۱۶۶۶	۰,۰۰۰۰
فیلم ۱۵۶۱	۰,۲۰۱۶	۰,۲۰۱۶	۰,۲۰۱۶	۰,۱۰۰۸	۰,۱۰۰۸
فیلم ۵۰۹۴	۰,۱۰۶۱	۰,۲۱۲۱	۰,۱۰۶۱	۰,۲۱۲۱	۰,۰۰۰۰

در حالت ۷۰ داده آموزشی، مدل توانسته فیلم اول را به درستی به عنوان لیست ترجیح کاربر تشخیص دهد، اما فیلم شماره ۱۵۶۱ که جزو غیر ترجیحات کاربر است، به اشتباه به عنوان لیست ترجیحات او شناخته شده است.

با جمع‌بندی تمامی کاربران آزمایشی در دو حالت ۱۵، ۳۰ و ۶۰ کاربر آزمایشی در مقابل ۳۵، ۷۰ و ۱۴۰ کاربر آموزشی، در صورت همسان بودن علامت RS_{EF} با لیست امتیازدهی، عدد یک، در صورت صفر بودن آن، عدد صفر و در صورت غیر همسان بودن این دو، عدد منفی یک، جهت مقایسه دقت مدل پژوهش با روش‌های پیش‌بینی مرسوم در نظر گرفته شده و نتایج آن مطابق شکل ۲ است.



شکل ۲- مقایسه عملکرد مدل پژوهش و دیگر الگوریتم‌ها

۵. نتیجه‌گیری

در این مقاله یک روش جدید پیشنهاددهی فیلم با استفاده از اطلاعات مشاهده کاربران و بهره‌گیری از مفاهیم تحلیل شبکه‌های اجتماعی ارائه گردید. در روش پیشنهادی این پژوهش براساس داده‌های درصد تماشای فیلم کاربران، شباهت فیلم‌ها با یکدیگر محاسبه شده و براساس آن گراف ارتباطی فیلم‌ها تشکیل می‌شود. در گراف ارتباطی فیلم‌ها ابتدا شاخص‌های مرکزیت هر یک از فیلم‌ها محاسبه شده و با استفاده از روش ماژولار خوشه‌های موجود در گراف استخراج می‌گردند. در مرحله بعدی و بر مبنای درصد مشاهده فیلم‌ها، لیست ترجیحات و غیرترجیحات کاربران مشخص شده و فیلم‌هایی که در خوشه با بیشترین تعداد ترجیحات کاربر حضور دارند، به عنوان کاندید پیشنهاد به کاربر انتخاب شده و در نهایت با محاسبه شاخص نهایی پیشنهاددهی که براساس مرکزیت خودمحور و فاصله از ترجیحات و غیرترجیحات هر کاربر است، لیست پیشنهاددهی به هر کاربر مشخص می‌گردد.

مدل ارائه شده در این تحقیق بر روی داده‌های ۵۰، ۱۰۰ و ۲۰۰ کاربر سایت سینما مارکت آموزش داده شده و تست شدند که براساس نتایج ارزیابی آن‌ها، الگوریتم پیشنهادی این پژوهش در مقایسه با الگوریتم‌های پرکاربرد نایویز، درخت تصمیم و نزدیک‌ترین همسایه، دارای برتری می‌باشد. الگوریتم پیشنهادی این پژوهش بر مبنای اطلاعات مشاهده فیلم توسط کاربران بوده و پژوهشگران علاقه‌مند به سیستم‌های پیشنهاددهنده در پژوهش‌های آتی می‌توانند با مبنای قرار دادن کاربردهای تحلیل شبکه‌های اجتماعی و با در اختیار داشتن اطلاعات بیشتری از جمله اطلاعات فیلم‌ها به توسعه سیستم‌های پیشنهاددهنده دقیق‌تری بپردازند.

منابع

- باستانی، س.، رئیسی، م. (۱۳۹۱). روش تحلیل شبکه: استفاده از رویکرد شبکه‌های کل در مطالعه اجتماعات متن باز. *مطالعات اجتماعی ایران*، ۵(۲): ۳۱-۵۷.
- جلیلوند خسروی، م.، مقصودی، م.، صلواتیان، س. (۱۴۰۰). شناسایی و دسته‌بندی کاربران پلتفرم‌های تماشای برخط فیلم به کمک تکنیک تحلیل شبکه اجتماعی؛ مطالعه موردی سینما مارکت. *تحقیقات بازاریابی نوین*، ۱۱(۴): ۲۰-۱.
- DOI: 10.22108/nmrj.2021.126442.2324
- محمدرضایی، ر.، روانمهر، ر. (۱۳۹۹). سیستم پیشنهاددهنده مبتنی بر اعتماد در محیط یادگیری الکترونیکی با استفاده از خوشه‌بندی فازی. *فناوری آموزش*، ۱۵(۳): ۴۳۹-۴۶۴. DOI: 10.22061/tej.2021.6807.2454
- میثاقیان، ز.، بزرگری‌نژاد، ع. (۱۳۹۵). *میزان دقت و پیش‌بینی در سیستم‌های توصیه‌کننده فیلم مبتنی بر داده‌کاوی با استفاده از تئوری بیزین*. در: همایش ملی دانش و فناوری مهندسی برق، کامپیوتر و مکانیک ایران. تهران: مؤسسه برگزارکننده همایش‌های توسعه محور دانش و فناوری سام ایرانیان.

References

- Adomavicius, G. & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering*, 17(6): 734-749. DOI: 10.1109/tkde.2005.99
- Ahn, S. & Shi, C. (2004). **Exploring Movie Recommendation System Using Cultural Metadata**. *Transactions on Edutainment*, II: 119-134. DOI: 10.1109/cw.2008.13
- Bastani, S. & Raissi, M. (2012). Social Network Analysis as a Method: Using Whole Network Approach for Studying FOSS Communities. *Journal of Iranian Social Studies*, 5(2): 31-57. [in persian]
- Bobadilla, J., Hernando, A., Ortega, F. & Bernal, J. (2011). A framework for collaborative filtering recommender systems. *Expert Systems with Applications*, 38(12): 14609-14623. DOI: 10.1016/j.eswa.2011.05.021
- Consumerlab, E. (2017). **TV and media 2017**. A consumer-driven future of media.
- Heidari, A. & Dehghani, H. (2016). The Qualitative Assessment of Married Women Teachers Family- work Conflict and Its Management in Delvar City. *Journal of Applied Sociology*, 27(3): 15-40. DOI: 10.22108/JAS.2016.20503
- Inan, E., Tekbacak, F. & Ozturk, C. (2018). Moreopt: A goal programming based movie recommender system. *Journal of computational science*, 28: 43-50. DOI: 10.1016/j.jocs.2018.08.004
- Jalilvand khosravi, M., Maghsoudi, M. & Salavatian, S. (2021). Identifying and Clustering Users of VOD Platforms Using SNA Technique: A case study of Cinemamarket. *New Marketing Research Journal*, 11(4). DOI: 10.22108/nmrj.2021.126442.2324 [in persian]
- Jallouli, M., Lajmi, S. & Amous, I. (2017). Designing Recommender System: Conceptual Framework and Practical Implementation. *Procedia Computer Science*, 112: 1701-1710. DOI: 10.1016/j.procs.2017.08.195
- Kataraya, R. & Verma, O.P. (2017). An effective collaborative movie recommender system with cuckoo search. *Egyptian Informatics Journal*, 18(2): 105-112. DOI: 10.1016/j.eij.2016.10.002

- Lin, J., Hsu, C.L. & Li, Y. (2014). Improving the effectiveness of experiential decisions by recommendation systems. *Expert Systems with Applications*, 41: 4904-4914.
DOI: 10.1016/j.eswa.2014.01.035
- Lina, N., Hongdi, L., Mengmeng, Z. & Jinquan, Z. (2018). Hybrid Filtrations Recommendation System based on Privacy Preserving in Edge Computing. *Procedia Computer Science*, 129: 407-409.
DOI: 10.1016/j.procs.2018.03.016
- Lops, P., Gemmis, M. & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. *Recommender systems handbook*: 73-105. **DOI:** 10.1007/978-0-387-85820-3_3
- Meteren, R. & Someren, M. (2000). **Using Content-Based Filtering for Recommendation**. In: *Proceedings of the Machine Learning in the New Information Age*.
- Middleton, S.E., Shadbolt, N.R. & Roure, D.C. (2004). Ontological user profiling in recommender systems. *ACM Transactions on Information Systems*, 22(1): 54-88.
DOI: 10.1145/963770.963773
- Misaghian, Z. & Barzeghari Nezhad, A. (2016). **Accuracy and prediction in data mining systems based on data mining using Bayesian theory**. In: *National Conference on Science and Technology of Electrical Engineering, Computer and Mechanics of Iran*. Tehran: Sam Iranian Institute for Organizing Knowledge and Technology Development Conferences. [in persian]
- Mohamadrezaei, R. & Ravanmehr, R. (2021). A trust-based recommender system for e-Learning environment using fuzzy clustering. *Technology of Education Journal (TEJ)*, 15(3).
DOI: 10.22061/TEJ.2021.6807.2454[in persian]
- Newman, M.E. (2006). Modularity and community structure in networks. *National Academy of Sciences*, 103(23): 8577-8582. **DOI:** 10.1073/pnas.0601602103
- Park, D.H., Kim, H.K., Choi, I.Y. & Kim, J.K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11): 10059-10072.
DOI: 10.1016/j.eswa.2012.02.038
- Park, Y., Park, S., Jung, W. & Lee, S. (2015). Reversed CF: A fast collaborative filtering algorithm using a k-nearest neighbor graph. *Expert Systems with Applications*, 42(8): 4022-4028.
DOI: 10.1016/j.eswa.2015.01.001
- Pazzani, M.J. (1999). A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence review*, 13(5-6): 393- 408.
- Pera, M.S. & Ng, Y.K. (2013). A group recommender for movies based on content similarity and popularity. *Information Processing & Management*, 49(3): 673-687.
DOI: 10.1016/j.ipm.2012.07.007
- Ricci, F., Rokach, L. & Shapira, B. (2010). Introduction to recommender systems handbook. *Recommender Systems Handbook*: 1-35. **DOI:** 10.1007/978-0-387-85820-3_1
- Robin, B. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4): 331-370. **DOI:** 10.1109/dictap.2012.6215409
- Son, J. & Kim, S.B. (2017). Content-based filtering for recommendation systems using multiattribute networks. *Expert Systems with Applications*, 89: 404-412.
DOI: 10.1016/j.eswa.2017.08.008
- Stuart, E.M. (2003). **Capturing KNowledge of User Preferences with Recommender Systems, A Thesis Submitted for the Degree of Doctor of Philosophy**. University of Southampton.

- Su, X. & Khoshgoftaar, T.M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*: 1-19. **DOI**: 10.1155/2009/421425
- Wasserman, S. & Faust, K. (1995). *Social network analysis: Methods and applications (Structural Analysis in the Social Sciences)*. Cambridge University Press.
DOI: <https://doi.org/10.1017/CBO9780511815478>
- Zhang, S., Jin, Z. & Zhang, J. (2016). The dynamical modeling and simulation analysis of the recommendation on the user–movie network. *Physica A: Statistical Mechanics and its Applications*, 463: 310-319. **DOI**: 10.1016/j.physa.2016.07.049