ارائهٔ یک سامانهٔ پیشنهادگر حافظه‌پایهٔ ترکیبی با استفاده از هستان‌شناسی و محتوا

پیام بحرانی[۱]، بهروز مینایی بیدگلی[۲]، حمید پروین*[۳و۴]، میترا میرزارضایی[۱] و احمد کشاورز[۵]

[۱]گروه مهندسی کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

[۲]دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران

[۳]گروه مهندسی کامپیوتر، واحد نورآباد ممسنی، دانشگاه آزاد اسلامی، نورآباد ممسنی فارس، ایران

[۴]باشگاه پژوهش‌گران جوان و نخبگان، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران

[۵]گروه مهندسی برق، دانشکده مهندسی سامانه‌های هوشمند و علوم داده، دانشگاه خلیج فارس، بوشهر، ایران

چکیده

سامانه‌های پیشنهادگر در زمینه تجارت الکترونیک شناخته شده هستند. از این‌گونه سامانه‌ها انتظار می‌رود که کالاها و اقلام مهمی (از جمله موسیقی و فیلم) را به مشتریان پیشنهاد دهند. در سامانه‌های پیشنهادگر سنتی از جمله روش‌های پالایش محتوا پایه و پالایش مشارکتی، چالش‌ها و مشکلات مهمی از جمله شروع سرد، مقیاس‌پذیری و پراکندگی داده‌ها وجود دارد. اخیراً به‌کارگیری روش‌های ترکیبی توانسته با بهره‌گیری از مزایای این روش‌ها با هم، برخی از این چالش‌ها را تا حد قابل قبولی حل کنند. در این مقاله سعی می شود روشی برای پیشنهاد ارائه شود که ترکیبی از دو روش پالایش محتوا پایه و پالایش مشارکتی (شامل دو رویکرد حافظه پایه و مدل پایه) باشد. روش پالایش مشارکتی حافظه پایه، دقت بالایی دارد؛ اما مقیاس‌پذیری کمی دارد. در مقابل، رویکرد مدل پایه دارای دقت کمی در ارائه پیشنهاد به کاربران بوده، اما مقیاس‌پذیری بالایی از خود نشان می‌دهد. در این مقاله سامانه پیشنهادگر ترکیبی مبتنی بر هستان‌شناسی ارائه شده که از مزایای هر دو روش بهره برده و براساس رتبه‌بندی‌های واقعی، مورد ارزیابی قرار می‌گیرد. هستان‌شناسی، توصیفی واضح و رسمی برای تعریف یک پایگاه دانش شامل مفاهیم (کلاس‌ها) در حوزه موضوعی، نقش‌ها (رابطها) بین نمونه‌های مفاهیم، محدودیت‌های مربوط به رابطه‌ها، همراه با یک مجموعه از عناصر و اعضا (یا نمونه‌ها) است که یک پایگاه دانش را تعریف می‌کند. هستان‌شناسی در بخش پالایش محتوا پایه مورد استفاده قرار می‌گیرد و ساختار هستان‌شناسی به‌وسیلهٔ روش‌های پالایش مشارکتی بهبود می‌یابد. در روش ارائه‌شده در این پژوهش، عملکرد سامانهٔ پیشنهادی بهتر از عملکرد پالایش محتوا پایه و مشارکتی است. روش پیشنهادی با استفاده از یک مجموعه‌داده واقعی ارزیابی شده است و نتایج آزمایش‌ها نشان می‌دهد روش یادشده کارایی بهتری دارد. همچنین با توجه به راه‌کارهای ارائه‌شده در مقاله حاضر، مشخص شد، روش پیشنهادی دقت و مقیاس پذیری مناسبی نسبت به سامانه‌های پیشنهادگری دارد که تنهاً حافظه پایه (*KNN*) و یا مدل پایه هستند.

واژگان کلیدی: سامانه پیشنهادگر، هستان‌شناسی، پالایش حافظه پایه، پالایش مدل پایه، خوشه‌بندی، *KNN*

A New WordNet Enriched Content-Collaborative Recommender System

Payam Bahrani[1], Behrouz Minaei-Bidgoli[2], Hamid Parvin[3,4*], Mitra Mirzarezaee[1] & Ahmad Keshavarz[5]

[1]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

[2]School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

[3]Department of Computer Engineering, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

[4]Young Researchers and Elite Club, Yasooj Branch, Islamic Azad University, Yasooj, Iran

[5]Department of Electrical Engineering, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr, 75169, Iran

* Corresponding author

*نویسندهٔ عهده‌دار مکاتبات

●تاریخ ارسال مقاله: ۱۳۹۹/۰۸/۰۴ ●تاریخ پذیرش: ۱۳۹۹/۱۲/۱۸ ●تاریخ انتشار: ۱۴۰۰/۱۲/۲۹ ●نوع مطالعه: پژوهشی

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

Abstract

The recommender systems are models that are to predict the potential interests of users among a number of items. These systems are widespread and they have many applications in real-world. These systems are generally based on one of two structural types: collaborative filtering and content filtering. There are some systems which are based on both of them. These systems are named hybrid recommender systems. Recently, many researchers have proved that using content models along with these systems can improve the efficacy of hybrid recommender systems. In this paper, we propose to use a new hybrid recommender system where we use a WordNet to improve its performance. This WordNet is also automatically generated and improved during its generation. Our ontology creates a knowledge base of concepts and their relations. This WordNet is used in the content collaborator section in our hybrid recommender system. We improve our ontological structure via a content filtering technique. Our method also benefits from a clustering task in its collaborative section. Indeed, we use a passive clustering task to improve the time complexity of our hybrid recommender system. Although this is a hybrid method, it consists of two separate sections. These two sections work together during learning.

Our hybrid recommender system incorporates a basic memory-based approach and a basic model-based approach in such a way that it is as accurate as a memory-based approach and as scalable as a model-based approach. Our hybrid recommender system is assessed by a well-known data set. The empirical results indicate that our hybrid recommender system is superior to the state of the art methods. Also, our hybrid recommender system is more accurate and scalable compared to the recommender systems, which are simply memory-based (*KNN*) or basic model-based. The empirical results also confirm that our hybrid recommender system is superior to the state of the art methods in terms of the consumed time.

While this method is more accurate than model-based methods, it is also faster than memory-based methods. However, this method is not much weaker in terms of accuracy than memory-based methods, and not much weaker in terms of speed than model-based methods.

Keywords: Recommender System; Ontology; Memory-based Filtering; Model-based Filtering; Clustering; *KNN*.

۱- مقدمه

پیداکردن اطلاعات در تارنماهای بزرگ یک روند دشوار و وقت‌گیر است. رویکرد استفاده از سامانه‌های پیشنهادگر در خط مقدم بازیابی اطلاعات و سامانه‌های پالایش اطلاعات ظاهر می‌شود. این سامانه‌ها به‌منظور توصیه کالاها/خدمات مورد نیاز کاربران (بدون نیاز به جستجوی صریح) توسعه داده شده‌اند. تاریخچه سامانه‌های پیشنهادگر به سال ۱۹۷۹، در ارتباط با علوم شناختی [1]، برمی‌گردد. این سامانه‌ها در سایر زمینه‌های کاربردی مانند بازیابی اطلاعات [2]، علم مدیریت [3]، نظریه تقریبی [4]، مدل‌سازی انتخاب مصرف‌کننده در کسب و کار و بازاریابی [5] و نظریه‌های پیش‌بینی [6] حائز اهمیت هستند. بر اساس نیاز افراد، سامانه‌های پیشنهادگر به آن‌ها در یافتن اقلام مناسب کمک می‌کنند [7 , 8]. سامانه‌های پیشنهادگر با ارائه پیشنهادهایی (تولید مجموعه‌ای از پیشنهادهای سفارشی) مطابق با علایق کاربران، به آن‌ها کمک می‌کنند تا اطلاعات مورد نیاز خود را پیدا کنند. سامانه‌های پیشنهادگر در تجارت الکترونیک به‌طور گسترده مورد استفاده قرار می‌گیرند. شرکت‌هایی مانند Amazon.com برخی از سرویس‌های قدرتمند و جالب را در این زمینه طراحی کرده‌اند. از آن‌جایی که به‌طورمعمول، عوامل متعددی در انتخاب و خرید محصول حائز اهمیت می‌باشد و این امر تصمیم‌گیری را برای مشتری دشوار می‌سازد، این سامانه‌ها به‌منظور پیشنهاد محصولات به مشتریان توسعه یافته‌اند تا در انتخاب هرچه بهتر کالاها/خدمات به آن‌ها کمک کنند.

سامانه‌های پیشنهادگر به دو دسته کلی محتوا پایه و مشارکتی تقسیم می‌شوند. سامانه‌های محتواپایه کندتر و سامانه‌های مشارکتی محبوب‌تر هستند. سامانه‌های مشارکتی اگرچه عملکرد خوبی در ابتدا (شروع سرد) ندارند اما کارایی خوبی بعد از آن دارند. به‌همین دلیل برخی از پژوهش‌گران به سراغ ترکیب این دو روش (سامانه‌های تلفیقی) رفته‌اند.

سامانه‌های محتوا پایه که از هستان‌شناسی استفاده کرده، بنا به موفقیت‌های زیاد در زمینه بازیابی اطلاعات محبوبیت پیدا کرده‌اند. در همین راستا سامانه‌های محتوا پایه که از هستان‌شناسی موردی[1] (مثلاً هستان‌شناسی فیلم) استفاده می‌کنند، به‌وجود آمدند. به‌کمک این سامانه‌ها تا حدود قابل قبولی بر چالش شروع سرد غلبه شده است، لیکن استفاده از wordnet به معنای تمام، در سامانه‌های پیشنهادگر ترکیبی استفاده نشده است. استفاده از wordnet به‌دلیل پیچیدگی و سربار زمانی

[1] Case Study

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

پیاده‌سازی آن، کمتر مورد توجه پژوهش‌گران در سامانه های پیشنهادگر قرار گرفته است.

در این مقاله قرار است سامانه پیشنهادگر تلفیقی ارائه شود که از ابزار هستان‌شناسی و شباهت‌سنج معنایی به‌کمک wordnet برای حل مشکل شروع سرد استفاده کند. همچنین مقیاس‌پذیری کار، مورد تحلیل قرار خواهد گرفت. در ضمن از آن‌جایی‌که سامانه‌های پیشنهادگر (محتوا پایه و مشارکتی) در مواجهه با مشکل پیشنهاد اقلام جدید و به‌طور بی‌ربط از سوابق گذشته کاربر، نمی‌توانند به‌صورت مطلوبی عمل نمایند، در این مقاله به‌سمت ایجاد شباهت‌های مشارکتی- معنایی اقلام رفته و سعی در بهبود مشکل مذکور می‌شود. همچنین رویکردی برای بهبود پروفایل کاربران به‌کمک سازوکار فوق استفاده خواهد شد. در این مقاله این رویکرد کمک می‌کند تا با حل چالش شروع سرد، اقلام جدیدی که از نظر محتوایی به اقلام دیگر نزدیک است، به کاربران پیشنهاد شود.

همان‌طور که درقبل نیز بیان شد در این مقاله، از راه‌کار پالایش ترکیبی و محتوا پایه استفاده می‌شود. در پالایش محتوا پایه و خوشه‌بندی (پیشرفته) به‌کار رفته در روش یادشده از روش‌های شباهت معنایی و هستان‌شناسی استفاده می‌شود. همچنین به‌منظور به‌دست‌آوردن توصیه‌هایی با کیفیت بالاتر، اندازه‌گیری شباهت معنایی (به‌منظور خوشه‌بندی) بر روی فرا داده مبتنی بر هستان‌شناسی انجام خواهد شد. علاوه‌بر ترکیب شباهت معنایی و پالایش محتوا پایه، راه‌کار دیگر این مقاله، استفاده از الگوریتم‌های داده‌کاوی است. به‌منظور به‌دست آوردن مدلی برای فرآیند پالایش مشارکتی جهت ارائه پیشنهاد به کاربر فعال از الگوریتم‌های داده‌کاوی استفاده خواهد شد. درواقع، دقت پیشنهادهای مبتنی بر خوشه‌بندی را می‌توان با استفاده از سازوکارهایی هم‌چون هستان‌شناسی بهبود داد. هدف نهایی این پژوهش ارائه یک سامانه پیشنهادگر ترکیبی است که در آن هم مقیاس‌پذیری روش مدل پایه و هم دقت روش حافظه پایه لحاظ شده باشد. این بدان معنی است که می‌توان انتظار داشت، با استفاده از روش ترکیبی فوق، پیشنهادهای ارائه‌شده به کاربر فعال با سلایق و نیازهای او منطبق باشد. همچنین می‌توان به‌منظور افزایش دقت سامانه پیشنهادی در ارائه پیشنهاد به کاربران، ویژگی‌های تصویری اقلام را نیز مد نظر قرار داد.

در حال حاضر، سامانه پیشنهادگری که در پردازش ارائه پیشنهادها، تمامی موارد بالا را مدنظر قرار دهد، وجود ندارد. از این‌رو، این پژوهش سعی در توسعه سامانه پیشنهادگر ترکیبی جدیدی مبتنی بر پالایش محتوا پایه، پالایش مشارکتی و هستان‌شناسی دارد.

در بخش پالایش مشارکتی، از روش خوشه‌بندی، پروفایل کاربر بر اساس هستان‌شناسی، هستان‌شناسی اقلام، شباهت معنایی بین دو هستان‌شناسی و الگوریتم رده‌بندی پیشنهادشده، برای غلبه بر مشکلات پراکندگی و شروع سرد استفاده می‌شود. این ادغام، به‌ترتیب، مقیاس‌پذیری و دقت سامانه را بهبود می‌بخشد. از سوی دیگر، هستان‌شناسی بر مبنای اقلام و شباهت معنایی در پالایش محتوا پایه اعمال می‌شوند. برای بهبود دقت در اندازه‌گیری شباهت معنایی، یک روش ابتکاری در این بخش برای سنجش درجه $IS-A$ بین دو گره هستان‌شناسی اقلام که منجر به ساخت یک لیست پیشنهادی دقیق‌تر برای کاربر فعال می‌شود، مورد استفاده قرار می‌گیرد.

به‌طور خلاصه، مقاله حاضر به‌منظور رسیدن به اهداف زیر انجام می‌شود:

۱. افزایش دقت شباهت معنایی به‌وسیله حذف یال‌های همسان در گراف هستان‌شناسی
۲. بهبود مسأله مقیاس‌پذیری با استفاده از خوشه‌بندی بهبودیافته
۳. ترکیب خوشه‌بندی و هستان‌شناسی بهبودیافته به‌منظور افزایش دقت و کارایی بخش پالایش مشارکتی
۴. دست‌یابی به عملکرد مناسب روش مبتنی بر مدل و دقت بالای روش حافظه پایه با ترکیب روش‌های مدل پایه و حافظه پایه با استفاده از هستان‌شناسی در بخش پالایش مشارکتی

به‌طور کلی جنبه‌های نوآوری در روش پیشنهادی این پژوهش عبارتند از:

- بهبود سامانه پیشنهادگر با استفاده از هستان‌شناسی و محتوا در مواجهه با پدیده شروع سرد
- بهبود سامانه پیشنهادگر با استفاده از هستان‌شناسی و محتوا در مواجهه با مشکل عدم توصیه اقلام غیرمنتظره
- استفاده از خوشه‌بندی دوطرفه، و محتوا پایه- مشارکتی برای ساخت بهتر پروفایل کاربران و حل مشکل شروع سرد و عدم توصیه اقلام غیرمنتظره.

بخش‌بندی مقاله حاضر بدین صورت است که: در بخش ۱ این مقاله، مقدمه و توضیحات مختصری راجب به اهداف و نوآوری مقاله آورده شده است. در ادامه در بخش ۲، ادبیات پژوهش و کارهای مرتبط ارائه شده است. در بخش ۳ روش پیشنهادی ارائه شده است. بخش ۴ شامل

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

آزمایش‌های مربوط به روش پیشنهادی و نتایج تجربی است. درنهایت در بخش ۵، نتیجه‌گیری کلی و کارهای آینده مورد بحث واقع شده است.

۲- ادبیات پژوهش و کارهای مرتبط

۲-۱- ادبیات پژوهش

در این بخش سعی می‌شود تا در ابتدا به مباحثی پرداخته شود که راجع به روش پیشنهادی هستند و سپس در ادامه برخی از کارهای مرتبط با روش پیشنهادی این مقاله آورده شود. در مباحث ابتدایی این بخش، توضیحات مفصل و مفیدی در مورد روش‌هایی که قرار هستند با هم ترکیب شوند، ارائه می‌شود. همچنین به مزایا و معایب این روش‌ها پرداخته شده و سعی خواهد شد تا دلیل منطقی برای ترکیب این روش‌ها با هم آورده شود.

با وجود تمام تحولات در سامانه‌های پیشنهادگر و استفاده از نمونه‌های کاربردی موفق در دنیای تجارت، می‌توان گفت هم‌چنان سامانه‌های پیشنهادگر (که بر اساس پالایش محتوا پایه عمل می‌کنند) نیازمند بهبود در بخش‌های مختلف خود هستند تا بتوان بدین طریق میزان اثربخشی و دامنه تحت پوشش آن‌ها را بهبود بخشید.

اگرچه روش پالایش محتوا پایه به نوعی، دو مسأله کلیدی در سامانه‌های پیشنهادگر (پراکندگی و اولین رتبه) را حل می‌کند، با این‌حال این روش دارای معایبی است که مانع از به‌کارگیری آن در بسیاری موارد مانند سامانه‌های پیشنهادگر فیلم و موسیقی می‌شود [9]. روش‌های محتوا پایه، به‌طور طبیعی دارای محدودیت‌هایی هستند که از جمله این محدودیت‌ها، تعداد و نوع ویژگی‌های اقلامی هستند که به کاربران پیشنهاد می‌شوند [9]. این ویژگی‌ها می‌توانند به‌صورت خودکار یا دستی استخراج شده باشند. علاوه‌براین، مشکل عدم پیشنهاد اقلام جدید و غیرقابل‌انتظار [1][10] یکی دیگر از چالش‌های مهم روش پالایش محتوا پایه است. این مسأله را می‌توان این‌گونه توضیح داد که سامانه پیشنهادگر، تنها اقلامی که دارای رتبه بالایی هستند و با پروفایل کاربر مطابقت داشته باشند، را پیشنهاد می‌دهد. از این‌رو، پیشنهادها به‌ندرت دارای خصیصه تازگی[2] هستند. علاوه‌براین، در روش پالایش محتوا پایه زمانی‌که رتبه‌بندی‌ها کافی نیست، سامانه قادر به ارائه پیشنهادهای قابل اعتماد نیست [10,9]. سامانه‌های پالایش مشارکتی را می‌توان به دو دسته مدل پایه و حافظه پایه تقسیم‌بندی کرد [11]. روش‌های حافظه پایه [11] اقدام به پیش‌بینی و ارائه پیشنهادهایی بر اساس کل رتبه‌های داده‌شده به اقلام توسط کاربران می‌کند. از این‌رو، همه رتبه‌ها باید در حافظه نگهداری شوند. این روش یک رویکرد معمول برای افزایش دقت در فرایند ارائه پیشنهادها است. اما این روش برای وب‌گاه‌های بزرگ که شامل تعداد زیادی کاربر و کالا/خدمات هستند [12] قابل اجرا نیست. با توجه به پژوهشی که نویسندگان در [13] انجام دادند، روش‌های مدل پایه، در فرآیند یادگیری مدل، از رتبه‌بندی‌هایی که توسط گروهی از کاربران به کالاها داده شده است استفاده و پس از آن، اقدام به ارائه پیش‌بینی (رتبه‌بندی)، می‌کنند. به‌عبارت دیگر خروجی این روش‌ها، تولید الگوی رتبه‌بندی (آفلاین) است. به‌منظور بهبود مقیاس‌پذیری و مشکل کارایی پالایش مشارکتی مدل پایه، می‌توان از روش‌های خوشه‌بندی استفاده کرد [14, 15]، اما این کار ممکن است منجر به برخی مشکلات مانند کاهش دقت [16]، هم‌پوشانی و تعمیم‌پذیری شود. با تمام این راه‌کارها، این روش هم‌چنان برای جستجوکردن همه کاربران/اقلام در یک خوشه، به‌منظور یافتن بهترین همسایه برای کاربر فعال، وقت‌گیر و زمان‌بر است. درنتیجه، مشکلات یادشده در مورد سامانه‌های مدل پایه می‌تواند منجر به کاهش دقت این روش در مقایسه با روش‌های حافظه پایه شود. الگوریتم‌های پالایش مشارکتی مدل پایه، می‌توانند به‌عنوان یک راه‌حل جایگزینِ استفاده از روش‌های حافظه پایه به‌منظور حل مسأله مقیاس‌پذیری روش یادشده مورد استفاده قرار گیرند. در راه‌کار مذکور از یک روش مبتنی بر احتمال به‌منظور ساخت مدل پیش‌بینی رتبه اقلام استفاده می‌شود.

الگوریتم‌های مدل پایه معایب روش‌های حافظه پایه را ندارند و می‌توانند در مقایسه با الگوریتم‌های حافظه پایه، پیش‌بینی رتبه اقلام را در زمان نسبتاً کوتاه‌تری انجام دهند؛ زیرا این الگوریتم‌ها، محاسبات لازم جهت آموزش مدل پیش‌بینی رتبه اقلام را به‌صورت آفلاین انجام می‌دهند. یکی از رویکردهای مدل پایه که به‌منظور بهبود کارایی پالایش مشارکتی حافظه پایه به‌کار می‌رود، روش خوشه‌بندی کاربران است [13, 17-23]. با این‌حال، استفاده از روش خوشه‌بندی ممکن است منجر به کاهش کارایی مزایایی همچون مقیاس‌پذیری، دقت، پوشش‌دهی، پیشنهادهای کمتر شخصی و تعمیم‌پذیری کل سامانه

[1] Serendipity
[2] Novelty

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

پیشنهادگر شود. درنتیجه، این مشکلات می‌تواند منجر به کاهش اثربخشی پیشنهادهای ارائه‌شده به کاربر، در مقایسه با روش‌های حافظه پایه شود.

نویسندگان در [16] پژوهشی بر روی الگوریتم‌های خوشه‌بندی در سامانه‌های پیشنهادگر انجام دادند و نتیجه گرفتند که استفاده از روش‌های خوشه‌بندی می‌تواند منجر به کاهش دقت در نتیجه پیشنهادها، در مقایسه با رویکردهای حافظه پایه شود. بدین ترتیب، روش‌های پالایش مشارکتی که از الگوریتم‌های خوشه‌بندی استفاده می‌کنند، راه‌حلی برای بهبود مقیاس‌پذیری و درنتیجه بهبود کارایی و همچنین حل مسأله پراکندگی محسوب می‌شوند. با این‌حال، استفاده از این روش‌ها به‌منظور افزایش دقت پیشنهادهای ارائه‌شده به کاربر فعال، تا حدودی بی‌فایده است. سامانه‌های پیشنهادگر ترکیبی می‌تواند از ترکیب دو یا چند روش به‌دست آیند و هدف آن‌ها بهبود معایب هر کدام از روش‌ها (به تنهایی) است.

سامانه‌های پیشنهادگر ترکیبی به‌طورمعمول با ترکیب روش‌های پالایش محتوا پایه و پالایش مشارکتی ساخته می‌شوند و سعی می‌کنند از مزایای هر دو روش برای ارائه پیشنهادهایی منطبق بر سلایق کاربران، بهره ببرند [13 , 24-27]. درواقع، کارایی سامانه‌های پیشنهادگر با استفاده از روش‌های ترکیبی بهبود خواهند یافت. لازم به ذکر است ویژگی‌های حوزه کاربرد و خصوصیات داده-های مورد استفاده، نقش مهمی در توسعه سامانه‌های پیشنهادگر ترکیبی دارند. به‌طور کلی هفت نوع سامانه پیشنهادگر ترکیبی وجود دارد که عبارتند از: مکانیزم وزندار[1]، مکانیزم راهگزینی[2]، راه‌کار آمیخته[3]، مکانیزم ترکیب خصوصیات[4]، مکانیزم آبشاری[5]، مکانیزم افزایش ویژگی‌ها[6]، مکانیزم فراسطح[7] [28].

از سوی دیگر، پیشنهادات ارائه‌شده توسط روش پالایش محتوا پایه، محدود به ویژگی‌هایی از اقلام می‌شود که به‌صورت صریح توسط کاربران رتبه‌بندی شده‌اند. به‌عنوان مثال، در یک سامانه پیشنهادگر فیلم که بر اساس پالایش محتوا پایه کار می‌کند، پیشنهادهای ارائه‌شده به کاربران تنها می‌تواند بر اساس شاخصه‌های یک فیلم هم‌چون نام بازیگران، خلاصه فیلم، ژانرهای سینمایی و غیره باشد. مشکل دیگر این‌گونه سامانه‌ها این است که در آنها، فقط اقلامی که دارای امتیاز بالایی در مقایسه با ویژگی‌های پروفایل یک کاربر خاص باشند، بازیابی می‌شوند. با این‌حال، محدودیت اصلی روش پالایش محتوا پایه، مشکل شروع سرد به‌دلیل ورود اقلام جدید به سامانه است. کالای جدید بایستی توسط کاربران به دقت رتبه-بندی شود تا یک سامانه پیشنهادگر محتوا پایه بتواند آن را به کاربران دیگر پیشنهاد دهد. یک کالای جدید بدون امتیاز یا با تعداد امتیاز کم، ممکن نیست به‌عنوان یک پیشنهاد دقیق به کاربران دیگر توصیه شود. بنابراین، بیش از حد اختصاصی‌شدن پیشنهادات، عدم تنوع و مسأله شروع سرد به دلیل ورود اقلام جدید به سامانه، در حال حاضر سه مورد مهم از مشکلات این نوع استراتژی است. پالایش مشارکتی می‌تواند برای غلبه بر برخی نقاط ضعف پالایش محتوا پایه استفاده شود. سامانه‌هایی که بر مبنای پالایش مشارکتی کار می‌کنند به‌طورمعمول پیشنهادها را بر اساس شباهت‌های رفتاری کاربران و الگوهای عملکرد آن‌ها، ارائه می‌دهند. درواقع، این سامانه‌ها می‌توانند برخی از نقایص پالایش محتوا پایه مانند محدودیت آنالیز محتوا، پیشنهادهای بیش از حد اختصاصی شده و عدم تنوع، را مرتفع کنند.

روش ترکیبی با هدف استفاده از مزایای روش‌های پالایش محتوا پایه و پالایش مشارکتی و به حداقل‌رساندن معایب آن‌ها مورد استفاده قرار می‌گیرد. اما بایستی این نکته را مد نظر قرار داد که حتی اگر سامانه ترکیبی بتواند مشکلات را به میزان زیادی حل کند ولی همچنان برخی از مشکلات حل‌نشده باقی می‌مانند. به‌همین دلیل به-منظور بهبود دقت سامانه ترکیبی می‌توان از روش‌هایی هم‌چون هستان‌شناسی و شاید در صورت لزوم از محتوای تصویری اقلام استفاده کرد.

در این پژوهش، به‌منظور بهبود دقت سامانه پیشنهادگر ارائه‌شده، از هستان‌شناسی به‌عنوان راه‌کاری مؤثر استفاده خواهد شد. درحقیقت، برای حل مسائلی مانند شروع سرد (در اثر ورود قلم جدید به سامانه)، یکی از ابزارهای اصلی وب معنایی، به نام هستان‌شناسی، می‌تواند مورد استفاده قرار گیرد. در ضمن برای حل مشکلاتی همچون پراکندگی داده‌ها و مقیاس‌پذیری، می‌توان از روش‌های خوشه‌بندی در قسمت پالایش مشارکتی استفاده کرد.

به‌طور خلاصه می‌توان گفت سازوکارهای مربوط به شباهت معنایی می‌توانند جهت افزایش دقت خوشه‌بندی کاربران (در بخش پالایش مشارکتی) و خوشه‌بندی اقلام (در بخش پالایش محتوا پایه) نقش مؤثری داشته باشند.

درخت مربوط به یک WordNet در شکل (۱) نشان داده شده است. تمامی "IS – A"های استفاده شده در این درخت، که ارتباط بین مفاهیم را نشان می‌دهند، مشابه یکدیگر هستند. این بدان معنی است که فرزندان

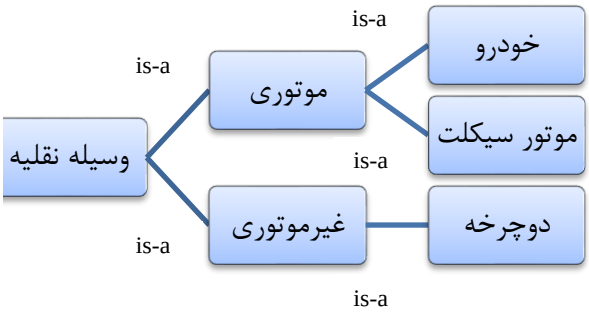
[1] Weighted
[2] Switching
[3] Mixed
[4] Feature Combination
[5] Cascade
[6] Feature Augmentation
[7] Meta-Level

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

یک والد دارای ارزش یکسانی هستند. در روش‌های قبلی هستان‌شناسی، تمام یال‌هایی که روابط بین مفاهیم را نشان می‌دهند، مشابه بوده و وزن همه آن‌ها برابر ۱ (همه یال‌ها یکنواخت) است. این مسأله دقت تشابه بین دو مفهوم را کاهش می‌دهد. درواقع، بهبود ساختار هستان‌شناسی که می‌تواند منجر به افزایش دقت سامانه پیشنهادگر شود، توسط بیشتر نویسندگان نادیده گرفته شده است. این امر به این معنی است که دو مفهوم که در یک سطح از درخت سلسله‌مراتبی واقع شده‌اند به لحاظ معنی بسیار نزدیک به مفاهیم دیگر هستند. درنتیجه این امر می‌تواند بر روی دقت سامانه‌های پیشنهادگر در طی فرایند یافتن اقلام و یا کاربران مشابه، تأثیر به‌سزایی داشته باشد.

به‌عنوان مثال، در WordNet نمایش داده شده در شکل (۱)، خودرو و موتورسیکلت فاصله یکسانی تا دوچرخه دارند، بنابراین می‌توان نتیجه گرفت، خودرو و موتورسیکلت به یک اندازه با دوچرخه دارای شباهت هستند. با این حال، همگی می‌دانیم که شباهت (موتور سیکلت، دوچرخه) بیشتر از شباهت (خودرو، دوچرخه) است. از این‌رو می‌توانیم تعریف کنیم:

فاصله (خودرو، دوچرخه) = فاصله (موتورسیکلت، دوچرخه) ⇐ شباهت (خودرو، دوچرخه) = شباهت (موتورسیکلت، دوچرخه)



(شکل-۱): بخشی از یک WordNet [29]

(Figure-1): Part of a WordNet [29]

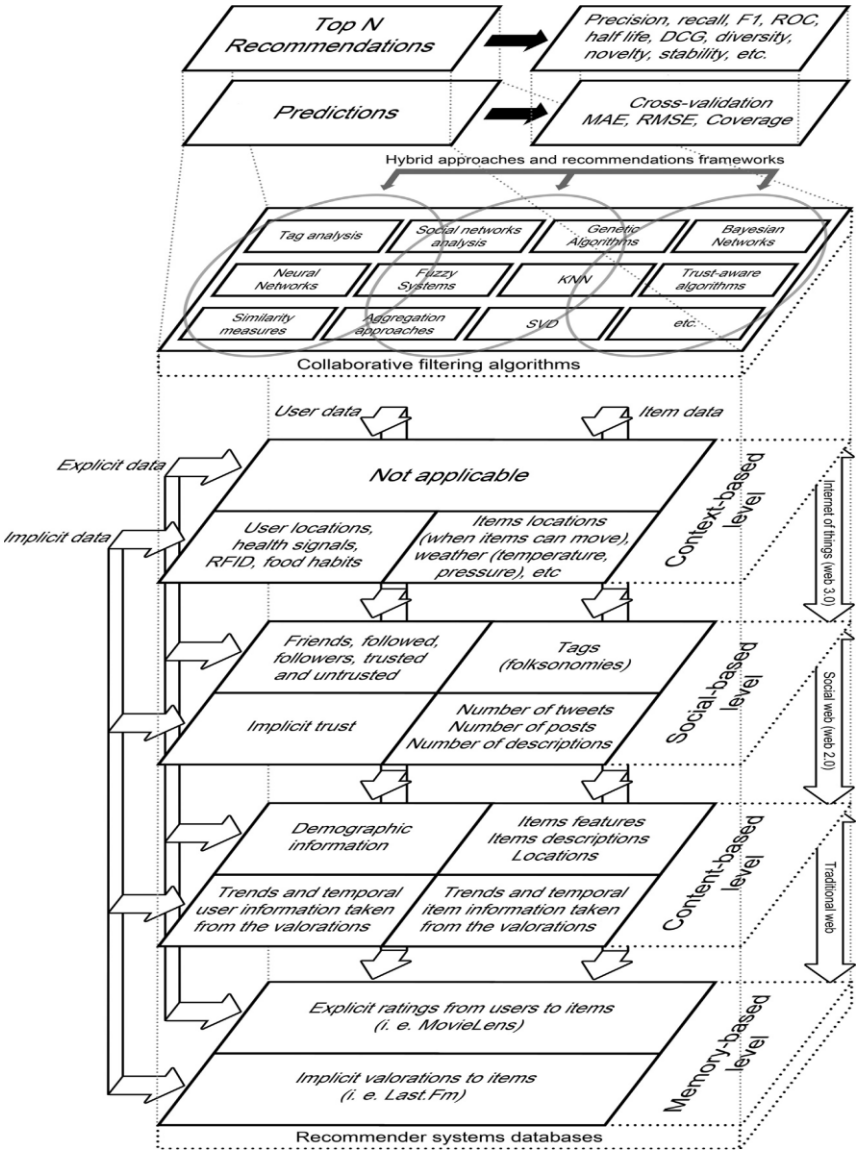
(جدول-۱): انواع سامانه‌های پیشنهادگر و نقاط قوت و ضعف آن‌ها

(Table-1): Types of Recommender Systems and their strengths and weaknesses

رویکردها	روش‌ها	توصیف مختصر روش	نقاط قوت	نقاط ضعف
رویکرد مبتنی بر یادگیری	روش پالایش مشارکتی	– شناسایی کاربرانی که علایق مشابه باکاربر هدف دارند و پیشنهاد اقلام مورد علاقه آنها به کاربر هدف	عدم نیاز به تحلیل محتوا پیشنهاد اقلام مختلف مستقل از دانش کیفیت بالا پیشنهاد اقلام غیر منتظره	کاربر غیرعادی مسأله پراکندگی مسأله شروع سرد مسأله مقیاس‌پذیری مسأله عدم شفافیت ترادف خطای محبوبیت عدم انعطاف‌پذیری
	روش پالایش محتوا پایه	– تحلیل محتوای آیتم‌هایی که کاربر درگذشته انتخاب کرده و پیشنهاد آیتم‌هایی با محتوای مشابه	مستقل از کاربر شفافیت آیتم جدید	زیاد اختصاصی‌کردن مسأله کاربر جدید محدودیت آنالیز محتوا
	روش مبتنی بر داده‌های شخصی	– مبتنی بر اطلاعات جمعیت شناختی کاربران، پیشنهاد ارائه می‌شود	عدم وجود مسأله شروع سرد مسأله از دامنه پیشنهاد اقلام غیرمنتظره پیاده‌سازی سریع و آسان اجرای سریع و آسان	جمع‌آوری اطلاعات مسأله پراکندگی ارائه پیشنهادهای عمومی مسأله کاربر غیرعادی عدم انعطاف‌پذیری
رویکرد دانش‌پایه	روش دانش پایه	– با اکتشاف نیازهای صریح کاربران و دانش عمیق درباره حوزه اقلام به ارائه پیشنهاد می‌پردازند	عدم نیاز به جمع‌آوری اطلاعات کاربر عدم وجود مسأله پراکندگی مناسب برای اکتشافات اتفاقی قابلیت اطمینان بالا عدم وجود مسأله کاربر غیرعادی کارایی بالا با وجود دانش اندک پاسخ سریع به کاربر هنگام تغییر ارجحیت کاربر	نیاز به کسب دانش مشکل با تعداد آیتم‌های بالا وابستگی کیفیت پیشنهادها به کیفیت دانش کسب شده
	روش مبتنی بر سودمندی	– پیشنهادهایی مبتنی بر سودمندی هر آیتم برای کاربر ایجاد می‌کند	عدم وجود مسأله پراکندگی مشارکت فاکتورهای کیفی در ارزش‌گذاری آیتم پاسخ سریع به کاربر هنگام تغییر ارجحیت کاربر	انعطاف‌پذیری کم
رویکرد ترکیبی	روش های ترکیبی	– روش‌های ترکیبی مبتنی بر ترکیب روش‌های ذکرشده در بالاست	استفاده از مزایای یک روش برای غلبه بر معایب روش‌های دیگر بهبود عملکرد سامانه‌های پیشنهادگر ارائه پیشنهادهایی با کیفیت‌تر	

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]



(شکل-۲): انواع مختلف سامانه‌های پیشنهادگر و رابطه بین آن‌ها [33]

(Figure-2): Various types of RSs and their relations [33]

در جدول (۱) رویکردهای مختلف سامانه‌های پیشنهادگر به همراه توضیحات (توصیف هر روش، نقاط ضعف و نقاط قوت) هر کدام آورده شده است. به‌منظور بهبود دقت شباهت معنایی، وزن یال‌ها در رده‌بندی باید مورد توجه قرار گیرد، برای این‌کار می‌توان از روش‌هایی همچون هستان‌شناسی، داده‌کاوی و الگوریتم‌هایی که بتواند به‌طور خودکار ارتباط بین مفاهیم را کشف کند، استفاده کرد. پس از این‌که وزن یال‌ها مشخص شد می‌توان رابطه شباهت معنایی را با درنظرگرفتن درجه "IS – A" که بین دو مفهوم تعریف شده است، محاسبه کرد. پژوهش‌گران پیشین، در فرآیند خوشه‌بندی (پالایش مشارکتی)، تنها از رتبه‌بندی صریح که توسط کاربران ایجاد شده است، استفاده کرده‌اند. وابستگی پالایش مشارکتی به رتبه‌بندی ایجادشده توسط عامل انسانی (کاربران) علاوه‌بر این‌که می‌تواند راه‌کاری جهت بالابردن کارایی کل سامانه پیشنهادگر باشند اما در عین‌حال می‌تواند یک نقص نیز محسوب شود. این‌گونه سامانه‌ها، به‌طورمعمول دارای دقت کمتری هستند و ممکن است منجر به ایجاد مشکلاتی هم‌چون ارائه پیشنهادهای بیش از حد کلی (کمتر شخصی‌سازی‌شده) به کاربران شوند. بنابراین، سامانه‌های پیشنهادگر می‌توانند از هستان‌شناسی دامنه[1]، برای ارتقاء شخصی‌سازی استفاده کنند. در این روش، علایق کاربران طی فرآیند خوشه‌بندی (پالایش مشارکتی)، به‌وسیله روش‌های هستان‌شناسی و استفاده از روش استنتاج مبتنی بر دامنه، به شیوه‌ای مؤثرتر و دقیق‌تر مدل‌سازی می‌شود. درواقع نوآوری روش پیشنهادی که این روش را از سایر روش‌های رقیب متمایز می‌کند شامل موارد زیر است:

- بهبود سامانه پیشنهادگر با به‌کارگیری هستان‌شناسی و محتوا در مواجهه با مشکل عدم توصیه اقلام غیرمنتظره

[1] Domain Ontologies

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

- به‌کارگیری هستان‌شناسی در سامانه و استفاده از خوشه‌بندی دوطرفه، و محتوا پایه- مشارکتی برای ساخت بهتر پروفایل کاربران و حل مشکل شروع سرد و عدم توصیه اقلام غیرمنتظره.

به‌طور خلاصه می‌توان گفت در این پژوهش از یک روش ابتکاری برای بهبود دقت پیشنهادات ارائه‌شده به کاربر فعال و ارتقاء شخصی‌سازی در بخش پالایش مشارکتیِ سامانه پیشنهادگر ترکیبی، استفاده خواهد شد. به‌منظور دست‌یابی به اهداف مذکور از روش‌هایی هم‌چون هستان‌شناسی و به‌کارگیری ویژگی‌های محتوا پایه اقلام، جهت ارتقاء شخصی‌سازی پیشنهادات استفاده خواهد شد و از ترکیب روش‌های حافظه پایه و مدل پایه (در بخش پالایش مشارکتی) به منظور بهبود دقت سامانه پیشنهادگر، استفاده می‌شود.

۲-۲- کارهای گذشته

هر لحظه تعداد مقالات، فایل‌های موسیقی، فیلم‌ها، کتاب‌ها و صفحات وب در اینترنت در حال افزایش هستند. بدون شک، پدیده غلبه بر داده‌ها، به‌وضوح در جامعه اطلاعاتی امروز ظهور کرده است [30,29]. در چنین محیطی، مردم واقعا نمی‌دانند با این مقدار عظیم اطلاعات چه کار کنند و اغلب به‌دلیل حجم بالای داده‌ها فرصت‌های موجود را نمی‌دانند و حتی در برخی موارد تصمیم‌گیری در این زمینه کاملا نادیده گرفته می‌شود. درواقع، یک RS یک راه‌کار قوی برای انجام کار اصلاح اطلاعات است [32,31]. این سامانه‌ها به یک روش محبوب برای فشرده‌سازی فضاهای اطلاعاتی بزرگ تبدیل شده‌اند که کاربران را به بهترین موارد مورد نیاز هدایت می‌کنند. هدف اصلی RS را می‌توان کاهش پیچیدگی برای کاربرانی بیان کرد که با داده‌های گسترده از مجموعه داده‌های عظیم متشکل از آیتم‌های متعدد، مانند ویژگی‌ها، رفتار و علایق کاربر احاطه شده‌اند و تلاش می‌کنند بهترین و مناسب‌ترین آیتم‌ها را در میان حجمی انبوه از آیتم‌ها به آن‌ها ارائه دهند [32,31]. یکی دیگر از اهداف اصلی در RSها افزایش دقت توصیه‌ها است. با گذشت زمان، با تغییر اولویت‌های کاربر، آیتم‌های مرتبط نیز باید تغییر کنند. لازم به ذکر است که به‌منظور استنتاج شباهت‌های بین کاربران، علاوه بر بازخورد کاربر در زمان‌های مختلف، اطلاعات شخصی کاربران نیز بسیار تأثیرگذار است. RSهای مختلف از منابع اطلاعاتی مختلف برای پیش‌بینی و پیشنهاد آیتم‌ها استفاده می‌کنند. این سامانه‌ها سعی می کنند بین دقت، تازگی، عدم انتشار و تناسب در توصیه‌ها تعادل ایجاد کنند. نیاز به RSها زمانی برجسته می‌شود که کاربران با مقدار زیادی از اطلاعات و موارد مواجه هستند. در چنین وضعیتی، سامانه باید پیشنهادهای خود را در میان طیف گسترده‌ای از اطلاعات با درنظرگرفتن موارد زیر به کاربران ارائه دهد: شرایط و محیطی که کاربران در آن قرار دارند، نیازهای کاربران، دانش سامانهاتیک کاربران، سوابق فعالیت کاربران [33]. در شکل (۲)، انواع دیگر سامانه‌های پیشنهادگر، رابطه بین آن‌ها و همچنین الگوریتم‌های استفاده‌شده در هر یک مشاهده می‌شود. در اواسط دهه ۱۹۹۰ هنگامی که نخستین بار پژوهش‌ها در زمینه سامانه‌های پیشنهادگرآغاز شد، سمت و سوی این پژوهش‌ها بیشتر متوجه روش‌هایی بر پایه رتبه‌بندی اقلام توسط کاربران تجارت الکترونیک بود. در چنین سامانه-هایی اساس پیشنهاد یک قلم را به یک کاربر فقط رتبه هایی تشکیل می‌داد که کاربران به آن قلم داده بودند. در این سامانه‌ها به هر کاربر c به‌عنوان عضوی از مجموعه همه کاربران C، یک قلم s از مجموعه همه اقلام موجود S، بر اساس رابطه (۱) پیشنهاد می‌شود:

$$\forall c \in C, \quad s'_c = \arg\max_{s \in S} u(c,s) \qquad (۱)$$

در رابطه (۱)، $u(c,s)$ یک تابع سودمندی[1] است که نشان می‌دهد که قلم s چقدر برای کاربر c مقبولیت داشته است. بیشینه مقدار این سودمندی‌ها برای هر کاربر می‌تواند در مورد یک قلم به‌خصوص مانند s'_c اتفاق بیافتد. این همان قلمی است که به کاربر مربوطه پیشنهاد خواهد شد. عمل انتخاب و ارائه یک یا چند قلم خاص به یک کاربر مشخص اصطلاحاً فیلترکردن[2] نام دارد.

انواع مختلفی از منابع اطلاعاتی در RSها وجود دارد. به‌طور خلاصه، ورودی اطلاعات بهRSها به‌صورت زیر است:

بنا به کاربرد سامانه پیشنهادگر، ممکن است انواع مختلفی از منابع اطلاعاتی در سامانه وجود داشته باشد. این اطلاعات می‌توانند امتیازهای کاربران به اقلام[3]، اطلاعات شخصی کاربران، محتوای مربوط به اقلام سامانه، ارتباطات موجود در شبکه‌های اجتماعی و اطلاعات مربوط به موقعیت کاربر[4] باشند. طبیعی است که در فرآیند طراحی یک سامانه پیشنهادگر باید به نوع داده‌های در اختیار توجه بسیار کرد.

[1] Utility Function
[2] Filtering
[3] Ratings
[4] Location-Aware Information

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

به‌طور خلاصه ورودی اطلاعات به سامانه‌های پیشنهادگر از طریق موارد زیر می‌باشد [34]:

رتبه‌ها[1]: نظرات کاربران در مورد اقلام موجود

اطلاعات پروفایل[2]: اطلاعاتی از قبیل جنسیت، سن، محل زندگی و ...

اطلاعات محتوایی[3]: تجزیه وتحلیل اسنادی که در مورد اقلام نوشته شده است

کاربر هدف[4]: کاربری که سامانه به او پیشنهادی می‌دهد.

هدف سامانه‌های پیشنهادگر در واقع رتبه‌بندی اقلام موجود به لحاظ نزدیک‌بودن به علایق کاربران است تا در هنگام ارائه پیشنهاد، اقلامی با رتبه بالاتر را به کاربر پیشنهاد دهند. در این میان بدون شک اساسی‌ترین جزء در این سامانه‌ها (پیشنهادگر)، الگوریتم و راه‌کار پالایش آن است [35]. برای این منظور الگوریتم‌های متعددی پیشنهاد شده‌اند که مهم‌ترین آنها به شرح زیر هستند [38-36]:

الگوریتم پالایش مشارکتی[5]

الگوریتم پالایش محتوا پایه[6]

الگوریتم پالایش مبتنی بر شبکه‌های اجتماعی[7]

الگوریتم پالایش دانش پایه[8]

الگوریتم پالایش آگاه از متن[9]

الگوریتم پالایش آگاه از مکان[10]

الگوریتم پالایش مبتنی بر جمعیت شناختی[11]

الگوریتم پیش‌بینی تصادفی

الگوریتم توالی مکرر[12]

الگوریتم پالایش ترکیبی[13]

نویسندگان [39] برخی از نوآوری‌ها را در هر دو رویکرد $ColF$، یعنی مدل‌های عامل نهفته (به‌طورمستقیم هم کاربران و اقلام را نمایه می‌کند) و مدل‌های همسایگی (شباهت‌های بین محصولات یا کاربران را تجزیه و تحلیل می‌کند)، معرفی می‌کند. آن‌ها با ادغام هر یک از آن‌ها در HRS، مشارکت‌هایی را به هر دو معرفی می‌کند. از مزایای این روش این است که توانسته تا حد مطلوبی مشکل پراکندگی داده‌ها را برطرف کند. در مورد روش [39] باید گفت که علی‌رغم نتایج بسیار خوب، این روش سربار پردازشی بالایی دارد و از آنها بهتر است در کاربردهای آفلاینی استفاده کرد که تغییرات زیادی در اطلاعات آن‌ها وجود ندارند. HRS از بازخورد ضمنی و صریح کاربران استفاده می‌کند. به‌طور معمول، داده‌های بازخورد صریح بسیار کم هستند. درنتیجه، روش‌هایی مانند کارهای قبلی [40] ممکن است با مشکل روبه‌رو شوند. با هدف مقابله با این چالش، [42,41] مدل جدیدی از عوامل نهفته را بر اساس فاکتورسازی[14] ماتریس احتمالی ارائه می‌دهند. علاوه‌بر اطلاعات ضمنی و صریح، برخی سامانه‌ها نیز وجود دارند که از اطلاعات شخصی کاربران استفاده می‌کنند. به‌عنوان مثال، سن، جنسیت و ملیت کاربران می‌تواند منبع مفیدی برای شناخت کاربر و بهبود پیشنهادها به آن‌ها باشد. به این نوع اطلاعات "اطلاعات جمعیتی[15]" گفته می‌شود.

$ColF$ پالایش مشارکتی ($ColF$) [43]، به‌عنوان فیلتر اجتماعی شناخته می‌شود. $ColF$ که با فیلترسازی اجتماعی نیز شناخته می‌شود [43]، پرکاربردترین و محبوب‌ترین الگوریتم پالایش و پایه و مبنای کار در بسیاری از راه‌کارهای دیگر است [44]. در این رویکرد پیشنهادها بر اساس جمع‌آوری و تحلیل اطلاعات در مورد رفتار کاربران سامانه ارائه می‌شوند. روش‌های مبتنی بر این رویکرد، مطلوبیت اقلام از نظر یک کاربر را با توجه به شباهت وی با سایر کاربران محاسبه می‌کنند. به عبارت دیگر، اقلام ارائه‌شده به یک کاربر، آنهایی هستند که از نظر کاربران مشابه با وی مطلوب بوده‌اند. این رویکرد، ریشه در رفتار انسان‌ها دارد. افراد به‌طورمعمول برای انتخاب یک فیلم برای مشاهده، یک محل گردشگری برای بازدید و غیره به پیشنهادهای دیگران متکی هستند [45]. این رویکرد سعی می‌کند مشارکتی را که میان کاربران در دنیای واقعی اتفاق می‌افتد شبیه‌سازی کند؛ بدین معنا که آنها نظرات خود در مورد اقلام را به اشتراک گذاشته و پیشنهاداتی را ارائه می‌دهند [45]. با توجه به نمایش قدرتمند سامانه‌های فیلترسازی مشارکتی در ارائه پیشنهادها باکیفیت و با درنظرگرفتن نیاز حداقلی آنها به اطلاعات دامنه‌ای پیرامون اقلام، این سامانه‌ها به‌نوعی به محبوب‌ترین رویکرد در سامانه‌های پیشنهادگر بدل شده‌اند [51-46]. در این روش به‌طورمعمول سامانه محیطی را فراهم می‌کند تا کاربران بتوانند برای هر قلم

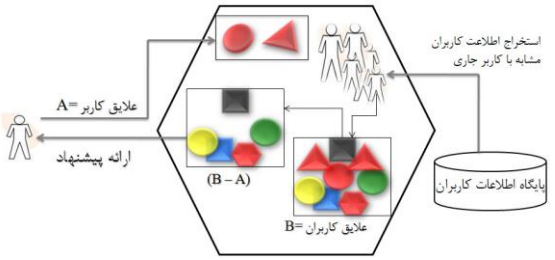
[1] Ranking
[2] Profile Information
[3] Content Information
[4] Current User
[5] Collaborative Filtering
[6] Content-Based Filtering
[7] Social-Based Filtering
[8] Knowledge-Based Filtering
[9] Context-Aware Filtering
[10] Location-Aware Filtering
[11] Demographic Filtering
[12] Frequent Sequence
[13] Hybrid Filtering

[14] matrix factorization
[15] demographic information

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

موجود در سامانه، رتبه دلخواه خود را ثبت کنند. در بعضی موارد هم که رتبه‌دهی به‌صورت صریح ممکن نیست، سامانه از روی بعضی نشانه‌ها رتبه‌ای برای اقلام خود محاسبه می‌کند [45]. برای مثال تعداد دفعاتی که از یک قلم بازدید به عمل آمده و یا تعداد دفعات بارگیری یک قلم یا متعلقات آن و یا تعداد دفعاتی که صفحه‌ای با محتوای شناخته‌شده برای سامانه، دیده شده و میانگین زمانی که کاربران به‌طورمعمول در آن صفحه صرف کرده‌اند. در این روش، سامانه، کلیه اقلام موجود را بر اساس امتیازهای داده‌شده توسط کاربران (صریح و ضمنی)، رتبه‌بندی می‌کند و به هر کاربر با توجه به علایقش اقلامی با بالاترین رتبه را پیشنهاد می‌کند. سامانه این‌گونه عمل می‌کند که با توجه به امتیازهای داده‌شده توسط کاربران، کاربرانی که از نظر امتیازهای داده‌شده با یکدیگر مشابه هستند را در یک گروه قرار می‌دهد و زمانی که می‌خواهد برای کاربری پیشنهادی مهیا کند، اقلامی که توسط کاربران موجود در همسایگی کاربر، رتبه بالایی را دریافت کرده‌اند به کاربر پیشنهاد می‌دهد. سامانه‌های پیشنهادگر مبتنی بر پالایش مشارکتی دو فاز یادگیری و پیش‌بینی دارند. با استفاده از یک مجموعه‌داده که نتیجه بازخورد کاربران با سامانه است (که همان رتبه‌بندی‌ها است)، سامانه یاد می‌گیرد و در مرحله بعد پیش‌بینی می کند. یکی از مهم‌ترین مسائل در این زمینه محاسبه شباهت بین دو کاربر است که با استفاده از رتبه‌بندی کاربران روی اقلام و همچنین تعاملات دیگر کاربران با سامانه محاسبه می‌شود. با توجه به نقش کلیدی معیار شباهت در دقت نتیجه پیشنهادها، در ادامه به بررسی مهم‌ترین روش‌های ارائه‌شده جهت محاسبه میزان شباهت بین کاربران و اقلام پرداخته می‌شود.

در شکل (۳) روند کار روش‌های $ColF$ آورده شده است.



(شکل-۳): روند کار پالایش در الگوریتم مشارکتی
(Figure-3): flowchart for the $ColF$ techniques

این روش‌ها بر روی ماتریس نظرات کاربران به اقلام تمرکز می‌کنند. این ماتریس یک ماتریس دو بعدی است که سطرهای آن متناظر با کاربران و ستون‌های آن متناظر با اقلام هستند. مقدار سلول i و j در این ماتریس به‌عنوان نظر کاربر i به قلم j تعریف می‌شود. که می‌تواند بولین، صحیح و یا اعشاری باشند. اگر مقادیر این ماتریس منفی باشد، نشان‌دهنده میزان عدم علاقهٔ کاربر به قلم مورد نظر خواهد بود. نظر کاربران به اقلام می‌تواند به‌صورت صریح [52] یا ضمنی [53] مشخص شده باشد. اگر نظر کاربران به اقلام به‌صورت صریح مشخص شده باشد، سامانه تنها نیاز است مقادیر را جمع‌آوری کرده و در ماتریس نظرات یادشده ذخیره کند. در غیر این صورت سامانه علاقه کاربران به اقلام را که به‌صورت ضمنی مشخص شده است اندازه‌گیری کرده و در ماتریس ذخیره می‌کند.

به‌عنوان مثال تعداد دفعاتی که یک کاربر به یک موسیقی گوش داده است [54]، تعداد برچسب‌های استفاده شده [55] یا تعداد بازدید یک کاربر از یک صفحه وب می‌تواند به‌عنوان علاقه ضمنی وی در نظر گرفته شود. سامانه‌های پالایش مشارکتی را می‌توان به دو دسته مدل پایه و حافظه پایه تقسیم‌بندی کرد. در روش‌های مدل پایه سامانه تلاش می‌کند تا مدلی را به‌منظور پیشگویی/پیشنهاددهی به کاربر، ایجاد کند. این مدل ممکن است مبتنی بر الگوریتم ژنتیک، شبکه‌های عصبی، مدل‌های فازی، شبکه بیزین، روش‌های خوشه‌بندی یا سایر مدل‌ها باشد [56-58]. روش‌های حافظه پایه، بر ماتریس علایق کاربران به اقلام تمرکز می‌کنند و شامل سه مرحله هستند. در مرحله نخست شباهت بین کاربران محاسبه می‌شود. در مرحلهٔ دوم شبیه‌ترین کاربران به کاربر فعال شناسایی شده و در قالب انجمن همسایگانِ وی در نظر گرفته می‌شوند.

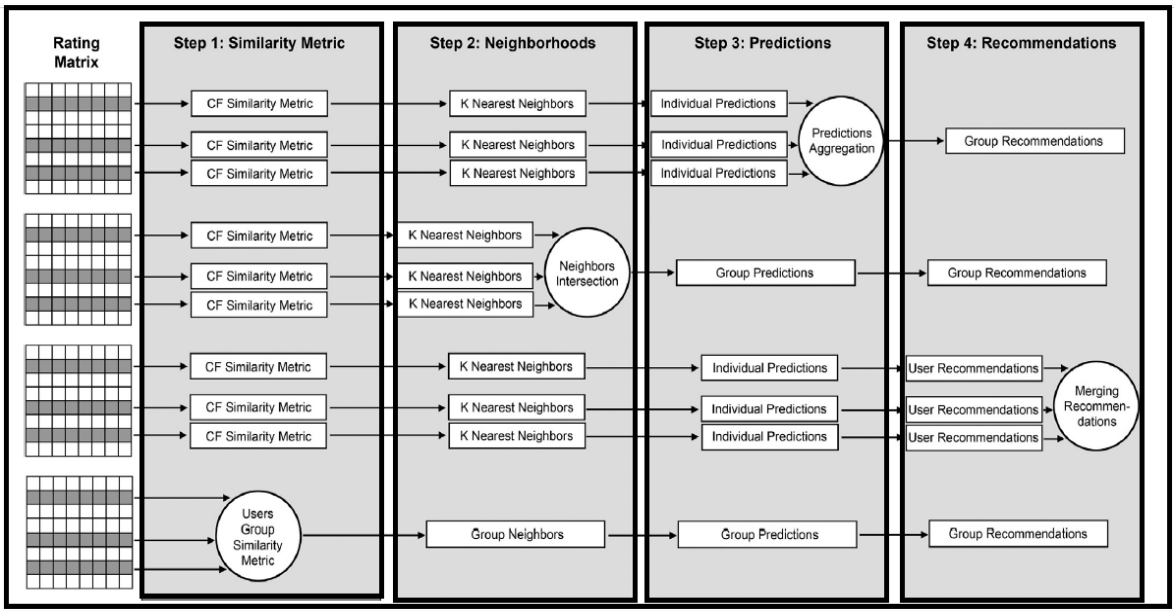
در آخرین مرحله، پیشگویی/پیشنهادهای همسایگان جمع‌بندی شده و نتیجه به کاربر فعال ارسال می‌شود. روش‌های حافظه پایه ممکن است **مبتنی بر کاربر** یا **مبتنی بر قلم** باشند. در روش های مبتنی بر کاربر تمرکز بر یافتن افراد شبیه به کاربر فعال می‌باشد در حالی‌که در روش‌های مبتنی بر قلم تمرکز روی اقلام است. در این حالت سامانه تعدادی از اقلام مشابه به نظرات پیشین کاربر فعال را به‌عنوان جواب ارسال می‌کند.

پیاده‌سازی یک الگوریتم $ColF$ براساس KNN ساده است و معمولا نتایج قابل قبولی در پیش‌بینی‌ها و پیشنهادها ایجاد می‌کند. دو نوع کاربرد برای این الگوریتم وجود دارد: در نوع اول، شباهت کاربر با کاربر در نظر گرفته می‌شود و در نوع دوم، شباهت آیتم با آیتم مورد استفاده قرار می‌گیرد. الگوریتم $ColF$ مبتنی بر KNN کاربر با کاربر، نزدیک‌ترین همسایگان کاربر هدف را براساس معیار شباهت انتخاب می‌کند [46]. سپس، آن‌ها

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

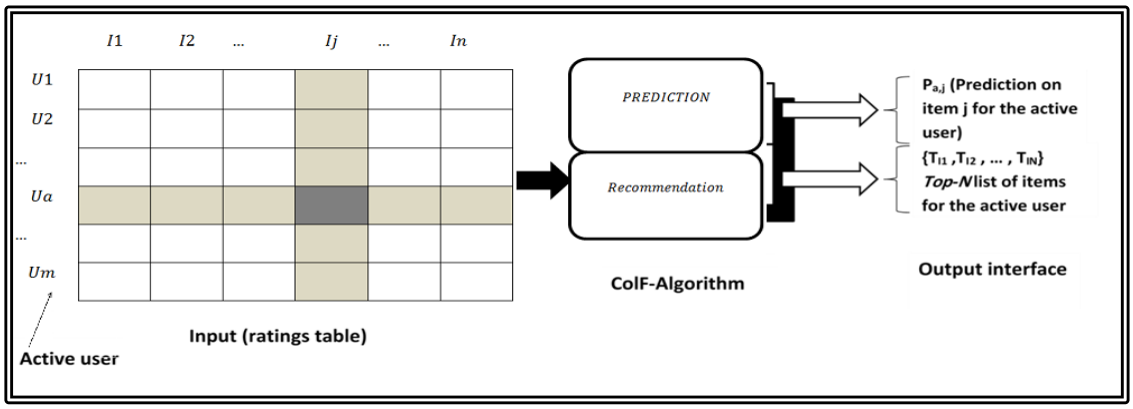
ارزشی را برای هر آیتم i به کاربر a پیش‌بینی می‌کنند. درنهایت، آیتم‌هایی با بالاترین مقادیر پیش‌بینی‌شده به کاربر به او بازگردانده می‌شوند.

شکل (۴) نمونه‌ای از الگوریتم $ColF$ کاربر با کاربر مبتنی بر KNN نشان می‌دهد.



(شکل-۴): نمونه‌ای از RS مبتنی بر الگوریتم KNN با رویکرد شباهت کاربر با کاربر [33]

(Figure-4): An example of a RS based on *KNN* algorithm with the user-by-user similarity approach [33]



(شکل-۵): روندنمای سامانه پیشنهادگر ColF کاربر با کاربر مبتنی بر KNN

(Figure-5): A ColF user-by-user RS flowchart based on *KNN*

یکی از مزایای این الگوریتم سادگی و در عین حال دقت نتایج آن است. البته، آن‌ها دو ضعف اساسی نیز دارند. یکی از مشکلات این الگوریتم ها ضعف آن‌ها در مقابله با کمبود فضای داده است [46]. آن‌ها مشکل اساسی دیگری دارند که مقیاس‌پذیری پایینی دارند [46]. برای مقابله با این مسأله، نسخه دیگری از این الگوریتم‌ها ارایه شده‌است. این الگوریتم‌ها مبتنی بر شباهت آیتم با آیتم KNN هستند. یک الگوریتم $ColF$ مبتنی بر آیتم با آیتم KNN از الگوریتم $ColF$ مبتنی بر کاربر با کاربر، مقیاس‌پذیرتر است. الگوریتم‌های $ColF$ مبتنی بر کاربر با کاربر الگوریتم‌های $ColF$ مبتنی بر KNN در ساختار و رفتار هستند، اما با کمی اختلاف [46]. هدف از الگوریتم‌های $ColF$ مبتنی بر کاربر با کاربر، پیدا کردن کاربرانی است که با کاربر جدید موافق باشند. سپس، آن‌ها آیتم‌های مورد علاقه خود را به کاربر جدید پیشنهاد می کنند. در این فرآیند، فهرستی از کاربران و فهرستی از آیتم‌ها مطابق با رابطه (۲) ایجاد می‌شود.

$$U = u_1, u_2, \ldots, u_m \quad , \quad I = i_1, i_2, \ldots, i_n \qquad (۲)$$

همان‌طور که توسط I_{u_i} نشان داده شده است، هر کاربر u_i لیستی از مواردی را که نظرات خود را ارسال کرده‌اند، دارد. دیدگاه‌های کاربر u_i می‌تواند به صراحت با توجه به نرخ‌هایی که توسط کاربر گرفته می‌شود یا به‌طور ضمنی از تعاملات و خریدهای قبلی کاربر استخراج شود،

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

بیان شود. شکل (۵) چارچوب $ColF$ کاربر با کاربر مبتنی بر KNN را نشان می‌دهد.

$ColF$ و $ConF$ از شباهت مبتنی بر کسینوسی برای بازیابی اطلاعات استفاده می‌کنند. در سامانه پیشنهادگر $ConF$، شباهت بین بردارهای وزن‌دهی TF – IDF تعریف می‌شود، در حالی‌که در سامانه پیشنهادگر $ColF$، شباهت در میان بردارهای درجه‌بندی کاربر تعیین می‌شود. برای توسعه روش‌های $ColF$، روش‌هایی مانند فرکانس کاربر معکوس، پیش‌بینی اکثریت وزنی و رأی‌گیری پیش‌فرض برای بهبود عملکرد سامانه پیشنهادگر ارائه شده‌اند [25]. الگوریتم‌های مبتنی بر مدل مجموعه دیگری از الگوریتم‌های مبتنی بر حافظه هستند. همان‌طور که گفته شد، این الگوریتم‌ها از یک مدل یادگیری برای پیش‌بینی درجه‌بندی استفاده می‌کنند [42]. به‌عنوان مثال، نویسندگان مرجع [25] یک روش احتمالی برای اجرای الگوریتم‌های $ColF$ پیشنهاد دادند. به این‌ترتیب، نرخ‌های ناشناخته به‌صورت معادله زیر محاسبه می‌شوند.

(۳)

$$r_{c,s} = E(r_{c,s}) = \sum_{i=0}^{n} i \times \Pr(r_{c,s} = i \,|\, r_{c,s'} \;,\; s' \in S_c)$$

که در آن فرض می‌کنیم مقادیر درجه‌بندی اعداد صحیح بین صفر و n هستند و می‌گوییم که کاربر c احتمالا همان‌طور که درقبل گفته شد، یک رتبه خاص به آیتم می دهد. رویکردهای $ColF$ به‌دلیل ماهیت خود چالش‌های زیر را دارند:

۱- **شروع سرد**[1]: همان‌طور که گفته شد در روش پالایش مشارکتی، سامانه بر مبنای امتیازات، اقلام را رتبه‌بندی می‌کند و اقلامی با بیشترین امتیاز را به کاربر پیشنهاد می‌دهد. به همین دلیل، در صورتی که سامانه تازه شروع به کار کرده باشد و یا آیتم جدیدی به سامانه اضافه شود، اطلاعات کافی از آیتم‌ها (یا آن آیتم) در دسترس نخواهد بود و درنتیجه نمی‌توان به‌درستی امتیازدهی و رتبه‌بندی[2] را انجام داد. این یکی از مشکلات اساسی و مهم در این‌گونه سامانه‌هاست که به‌عنوان شروع سرد شناخته می‌شود [47]. مسأله شروع سرد ممکن است به یکی از دلایل زیر رخ دهد:

۲- **شروع کار سامانه پیشنهادگر:** راه‌کاری که در چنین حالاتی پیشنهاد می‌شود این است که با استفاده از روش‌های مناسب، کاربران را تشویق به دادن رأی به آیتم‌ها نماییم و زمانی اقدام به پیشنهاد به کاربر کنیم که به‌اندازه کافی اطلاعات جمع‌آوری شده باشد.

۳- **ورود کاربر جدید به سامانه:** مهم‌ترین مشکل برای سامانه‌های پیشنهادگر مبتنی بر پالایش مشارکتی زمانی است که کاربر جدیدی وارد سامانه می‌شود. در این صورت اطلاعات کافی در مورد اقلام وجود دارد اما از آنجا که کاربر جدیدالورود هنوز به آیتمی رأی نداده است نمی‌توان از روش‌های معمول مورد استفاده در پالایش مشارکتی استفاده کرد. برای حل چنین مشکلی در سامانه، عموماً روش پالایش مشارکتی را با دیگر روش‌های رایج در سامانه‌های پیشنهادگر ترکیب کرده و یک سامانه ترکیبی را می‌سازند.

۴- **درج آیتم جدید در سامانه:** عموماً آیتم‌های جدید دارای هیچ امتیازی نیستند. بر همین اساس در لیست پیشنهادات هرگز آورده نمی‌شوند و از دیدگاه کاربران نیز پنهان می‌مانند. این مسأله باعث می‌شود که در آینده نیز به آنها هیچ امتیازی داده نشود.

۵- **پراکندگی داده‌ها**[3]: این سامانه‌ها از مشکل دیگری نیز رنج می‌برند که پراکندگی داده‌ها است. بدین معنی که اطلاعات در سامانه وجود دارد اما پراکنده هستند و نمی‌توان به‌درستی و با قطعیت گفت که چه آیتمی مقبولیت بیشتری دارد [48]. درواقع این مشکل وقتی به‌وجود می‌آید که حجم وسیعی از کاربران و اقلام در سامانه وجود داشته باشند و سطح پوشش نرخ‌گذاری کاربران بین اقلام کم باشد. یعنی ماتریس نرخ (کاربر- آیتم) یک ماتریس کم پشت[4] باشد که درنتیجه آن پیداکردن همسایه‌های مناسب برای کاربران کاری دشوار است. مشکل دیگر وجود کاربران با علایق غیرمعمول است که سامانه را در انتخاب همسایگان مناسب و درنهایت ارائه پیشنهادات معتبر به این کاربران دچار چالش می‌کند. منظور از ماتریس نرخ، ماتریسی است با ابعاد $n \times m$، که n تعداد کاربران و m تعداد آیتم‌هاست و هر عنصر $i \times j$ در این ماتریس نشان‌دهنده نرخی

[1] Cold Start
[2] Ratings
[3] Sparsity
[4] Sparse

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

است که توسط کاربر i برای آیتم j تخمین زده شده است. و به‌طورمعمول کاربران با کمتر از ۱٪ از آیتم‌های موجود در یک وب‌سایت سروکار دارند و فقط به این آیتم‌ها نرخ می‌دهند ونتیجه آن داشتن یک ماتریس بزرگ است که بیشتر عناصر آن تهی است. و این منجر می‌شود که جستجو در این ماتریس مشکل شود. و نتیجتاً میزان درستی و صحت در این سیستم‌ها پایین می آید. این مسأله در رویکرد ترکیبی تا حدودی بر طرف می‌شود. گان[1] و همکارانش برای رفع مشکل پراکندگی داده‌ها سامانهی را طراحی کرده‌اند، که مقادیر خالی را به‌کمک شبکه عصبی پس انتشار پر می‌کند. در این سامانه از الگوریتم شباهت بر اساس آیتم استفاده شده است [49]. برای رفع مشکل خلوت‌بودن ماتریس و افزایش کارایی، یک راه حل دیگر، کاهش بعد است. در مقاله [50] به‌کمک SVD[2] و PSO[3] عملیات کاهش بعد صورت گرفته است. برای به‌دست‌آوردن این‌که بعد چقدر باید کاهش پیدا کند از PSO استفاده شده است. در این روش به‌طور کل مصرف حافظه کاهش پیدا کرده است. همچنین سرعت و کیفیت پیشنهاد رشد داشته است.

۵. **مقیاس‌پذیری[4]**: درالگوریتم‌های پالایش مشارکتی مبتنی بر کاربر زمانی که کاربران صدها یا هزاران تن باشد به‌خوبی پاسخ می‌دهد اما امروزه تجارت الکترونیکی به سرعت در حال گسترش است و تعداد کاربران به بیشتر از میلیون‌ها تن رسیده است و این سیستم‌ها دیگر پاسخ‌گو نیستند زیرا در این سیستم‌ها محاسبات به‌صورت برخط محاسبه می‌شود و اگر حجم اطلاعات زیاد باشد زمان پاسخ‌گویی بسیار طولانی می‌شود که دیگر قابل قبول نیست. برای رفع این مشکل از الگوریتم بر اساس آیتم استفاده می‌کنند. یک الگوریتم افزایشی مقیاس‌پذیر، که مبتنی بر SVD با مقیاس‌پذیری خوب است، توسط نویسندگان معرفی شده است. نام آن افزایشی ApproSVD است [51].

به‌دلیل مشکلات شروع سرد و نیز پراکندگی داده‌ها عموماً سامانه‌های پالایش مشارکتی را به‌صورت ترکیبی با سایر راه‌کارها به‌کار می‌برند تا از مزایای آن‌ها بهره‌مند شده و در عین حال معایب آن را نیز بر طرف کنند. به عنوان مثال یکی از راه‌کارهایی که پیشنهاد شده است پیاده‌سازی تگ‌های مشارکتی[5] در یک سامانه مبتنی بر پالایش مشارکتی است تا بتوان سلایق کاربران را شناخت و اقلام را بر اساس تمایلات کاربران دسته‌بندی کرد [52].

یکی دیگر از راه‌های مقابله با مشکلات ذکرشده استفاده از روش‌های خوشه‌بندی است که عموماً برای حل مشکل شروع سرد به‌کار گرفته می‌شود. در این روش می‌توان اقلام یا کاربران و یا هر دوی آن‌ها (دو خوشه[6]) را خوشه‌بندی کرد. این روش‌ها علاوه‌بر، برطرف‌کردن مشکل ذکرشده باعث بهبود کارایی سامانه پیشنهادگر نیز می‌شوند. برای برطرف‌کردن مشکل پراکندگی داده‌ها، نیز عموماً از روش‌های کاهش ابعاد[7] استفاده می‌شود. در کنار این روش، روش‌های LSI[8] و SVD[9] نیز وجود دارند. در مورد روش SVD باید گفت که برخلاف نتایج بسیار خوب، این روش سربار پردازشی بالایی دارد و از آن‌ها بهتر است در کاربردهای آفلاینی استفاده کرد که تغییرات زیادی در اطلاعات آن‌ها وجود ندارند. خوشه‌بندی به‌طور گسترده‌ای در RS برای سرعت‌بخشیدن به آن‌ها استفاده می‌شود [53 , 54]. SVD می‌تواند به‌عنوان کاهش ابعاد در RS نیز استفاده شود [55 , 56]. برای سرعت‌بخشیدن به RS، خوشه‌بندی اولیه انجام شده و سپس، از SVD برای هر خوشه استفاده می‌شود. پژوهش‌گران رویکردی را ارائه داده‌اند که از طبقه‌بندی به‌طور خودکار با استخراج قاعده ارتباط استفاده می‌کند [57].

با توجه به خوشه‌بندی کاربران، بر اساس علایق یکسان ناشی از رتبه‌بندی خوشه‌بندی می‌شوند. با تشکیل خوشه‌ها، جمع‌آوری نظرات در هر خوشه برای انجام وظیفه پیش‌بینی خاص کاربر هدف استفاده می‌شود. از این‌رو، این موضوع به عملکرد بهتر منجر می‌شود، زیرا خوشه‌هایی که باید تجزیه و تحلیل شوند، در مقایسه با مقدار کل کاربران (به دلیل اینکه گروه مورد تجزیه و تحلیل از اندازه بسیار کوچکتری برخوردار است)، تعداد کاربران کمتری را شامل می‌شود [49].

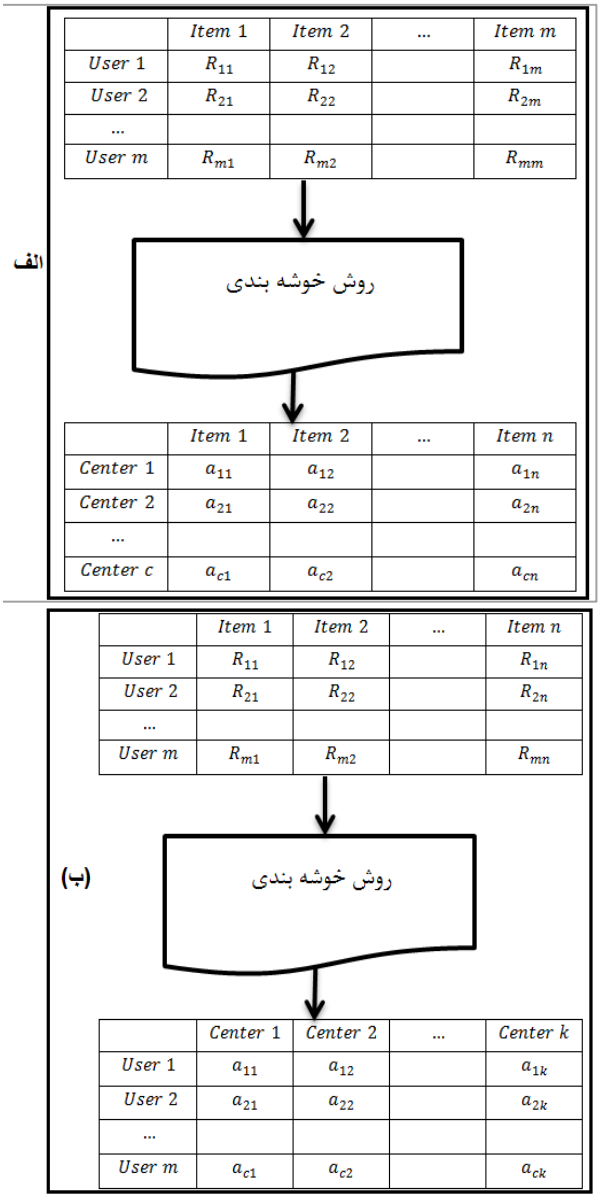
در شکل (۶) قسمت (الف)، m مقدار کل کاربران را نشان می‌دهد، R_{ij} رده‌بندی را که توسط کاربر i برای آیتم j ارائه می‌شود، تعیین می‌کند و n مقدار اقلام کلی را نشان می‌دهد. درواقع، ما می‌خواهیم کاربران را به تعدادی خوشه در شکل (۶) قسمت (الف)، تقسیم کنیم. در شکل (۶) قسمت (ب) هر ستون به‌عنوان یک رکورد و هر سطر

[1] Gone
[2] Singular Value Decomposition
[3] Partical Swarm Opitmization
[4] Scalability
[5] Collaborative Tagging
[6] Bi-Clustering
[7] Dimensionality Reduction
[8] Latent Semantic Index
[9] Singular Value Decomposition

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

به‌عنوان یک ویژگی توسط الگوریتم خوشه‌بندی در نظر گرفته شده است. درواقع، ما می‌خواهیم آیتم‌ها را به تعدادی خوشه در شکل (۶) قسمت (ب)، تقسیم کنیم.

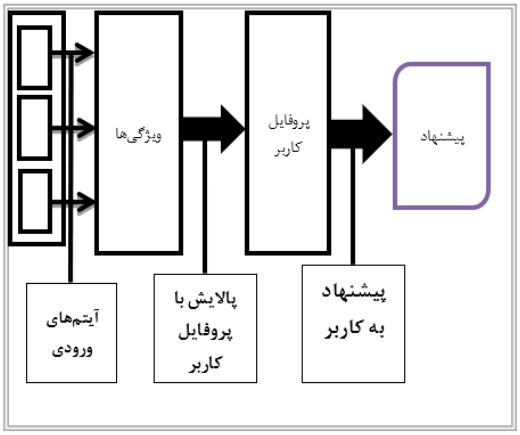
برای خوشه‌بندی آیتم، آیتم‌ها بر اساس رتبه‌بندی‌های یکسان ارائه‌شده توسط کاربران، خوشه‌بندی می‌شوند. پس از تشکیل خوشه‌ها، از تجمیع نظرات مرتبط با سایر آیتم‌ها در هرخوشه برای پیش‌بینی رتبه آیتم هدف استفاده می‌شود. از این‌رو، منجر به بهبود عملکرد می‌شود، زیرا خوشه‌های مورد تجزیه و تحلیل در مقایسه با مقدار کل اقلام از اقلام قابل توجه کم‌تری برخوردار هستند [49].



(شکل-۶): (الف) خوشه‌بندی کاربر و (ب) خوشه‌بندی آیتم در RSهای مبتنی بر ColF [55]

(Figure-6): (a) User Clus tering and (b) Item Clustering in ColF-based RSs [55]

روش پالایش مبتنی بر محتوا ($ConF$) برای پیشنهاددهی، ریشه در بازیابی اطلاعات [52] و فیلترکردن اطلاعات [53] دارد. از آن‌جایی که پیشرفت‌های قابل توجه و چشم‌گیری توسط بازیابی اطلاعات و انجمن‌های فیلترینگ در زمینه سیستم‌های مبتنی بر متن انجام شده است، بسیاری از سامانه‌های پیشنهادگر، روی اقلام مبتنی بر اطلاعات متنی مثل اسناد، آدرس وب سایت‌ها و متن پیام‌های خبری متمرکز شده‌اند. روش‌های بازیابی اطلاعات به شیوه سنتی از پروفایل کاربر به‌منظور اطلاع از اولویت‌های مشتری و نیازهای او استفاده می‌کنند. اطلاعات پروفایل یا به‌صورت صریح از طریق پرسش‌نامه یا به‌صورت ضمنی از رفتار تراکنش‌ها به‌دست می‌آید. به‌صورت فرمال، محتوا از یک پروفایل آیتم که مجموعه‌ای از مشخصه‌های آیتم S است به‌وسیله استخراج مجموعه‌ای از خصوصیات آیتم S برای مشخص‌کردن تناسب آیتم برای اهداف پیشنهاد استخراج می‌شود به‌عنوان نمونه در شکل (۷) می‌توانید فرآیندی که در یک سامانه پالایش محتوا پایه در حال انجام است را مشاهده کنید:



(شکل-۷): فرآیند پالایش محتوا پایه

(Figure-7): Content-Based filtering process

سامانه‌های مبتنی بر محتوا اغلب برای اقلامی مبتنی بر متن طراحی شده‌اند. محتوا دراین سامانه‌ها معمولاً با کلمات کلیدی[1] مشخص می‌شود. اهمیت وسودمندی کلمه k_i در سند d_j با وزن w_{ij} تعریف می‌شود. یکی از بهترین روش‌های اندازه‌گیری وزن کلمات کلیدی در بازیابی اطلاعات استفاده از اندازه‌گیری تعداد تکرار واژه/معکوس تعداد تکرار سند[2] است که به‌صورت زیر تعریف می‌شود: فرض کنیم که N تعداد کل متن‌هایی باشد که می‌توانند به کاربر پیشنهاد داده شوند و همچنین کلمه کلیدی k_i در n_i تا از آن‌ها وجود دارد. به‌علاوه فرض

[1] Key Words

[2] TF-IDF: Term Frequency/Inverse Document Frequency

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

کنیم که f_{ij} تعداد دفعات تکرار کلمه کلیدی k_i در متن d_j است. سپس $TF_{i,j}$، تکرار واژه کلیدی k_i در متن d_j است، که به‌صورت زیر تعریف می‌شود: بیشینه تعداد تکرارهای $f_{z,j}$ برای تمام کلمات کلیدی k_z که در متن d_j ظاهر شده است، و به‌صورت رابطه (۴) محاسبه می‌شود.

$$TF_{i,j} = \frac{f_{i,j}}{\max_z f_{z,j}} \quad (۴)$$

هنگامی‌که یک کلمه کلیدی در متن‌های زیادی تکرار می‌شود دیگر آن کلمه کلیدی برای نمایش تفاوت بین متن‌ها نمی‌تواند زیاد مفید باشد و با کمک آن کلمه کلیدی نمی‌توان تشخیص داد کدام متن مناسب و کدام متن غیرمناسب است. به‌همین دلیل اغلب از اندازه‌گیری معکوس تعداد تکرار سند[1] IDF_i در ترکیب با تعداد تکرار واژه‌های ساده $TF_{i,j}$ استفاده می‌کنند. معکوس تعداد تکرار سند برای کلمه کلیدی k_i معمولاً به‌صورت زیر تعریف می‌شود:

$$IDF_i = \log \frac{N}{n_i} \quad (۵)$$

سپس، وزن *TF-IDF* برای کلمه کلیدی k_i در متن d_j به‌صورت زیر تعریف می‌شود:

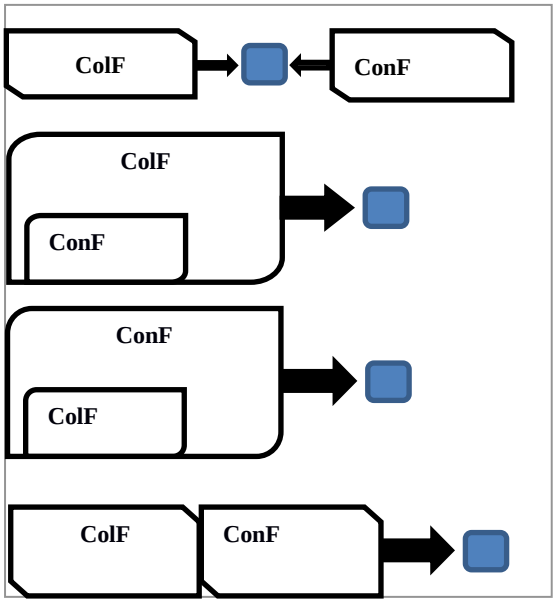
$$w_{i,j} = TF_{i,j} \times IDF_i \quad (۶)$$

و محتوای متن d_j نیز مانند زیر است:

$$Content(d_j) = (w_{1j}, \ldots, w_{kj}) \quad (۷)$$

همان‌گونه‌که درقبل توضیح داده شد، سیستم‌های مبتنی بر محتوا آیتم‌های مشابه با آنچه که کاربر درقبل انتخاب کرده است، به کاربر پیشنهاد می‌دهد [52]. آیتم‌های کاندیدهای گوناگون، با آیتمی که درقبل توسط کاربر نرخ‌گذاری شده است، مقایسه می‌شوند و بهترین آیتم که با آیتم مورد نظر هماهنگ است، پیشنهاد داده می‌شود. و برای فرمال‌ترکردن این منظور، یک پروفایل مربوط به کاربر با عنوان پروفایل محتوا پایه (c)[2] در نظر می‌گیریم. این پروفایل شامل علایق و ترجیحات کاربر c است. پروفایل مذکور با استفاده از تحلیل محتواهایی از آیتم‌هایی که درقبل کاربر مشاهده کرده و یا نرخ‌گذاری کرده است به‌دست می‌آید، که این کار به‌طورمعمول به‌کمک روش‌های تحلیل کلمات کلیدی از بازیابی اطلاعات، انجام می‌شود. به‌عنوان نمونه، پروفایل c شامل یک بردار از وزن‌های $(w_{c1}, w_{c2}, \ldots, w_{ck})$ است، که هر وزن w_{ci} نشان‌دهنده اهمیت کلمه کلیدی k_i برای کاربر c است.

پس از به‌وجودآمدن سامانه‌هایی بر اساس روش‌های بالا و مشخص‌شدن برخی مشکلات آنها، سامانه‌های پیشنهادگربرای فرار از این مشکلات و یا لااقل حذف ناکارآمدی‌های غیرمشترک از طریق هم‌پوشانی، به روش‌های ترکیبی روی آوردند [46]. به‌عنوان مثال روش پالایش مشارکتی از خواص کالاها استفاده نمی‌کند و فقط از تعاملات کاربران بهره می‌برد که با توجه به اینکه یک کاربر تازه وارد تعاملات اندکی با سامانه داشته است، این روش در ارائه پیشنهادات دقیق به او بهینه عمل نمی‌کند. پس می‌توان با ترکیب روش پالایش مشارکتی و محتوا پایه به‌طور دقیق‌تر از سلایق کاربر آگاه شد و پیشنهادات مؤثرتری به او ارائه داد. در شکل (۸) راه‌کارهای موجود جهت ترکیب دو روش پالایش مشارکتی و پالایش محتوا پایه نشان داده شده است:



(شکل-۸): انواع راه‌کارهای ترکیب دو روش پالایش مشارکتی و پالایش محتوا پایه [28]

(Figure-8): Different approaches to combine ColF and ConF models [28]

(الف) استفاده مجزا از هر روش و ترکیب نتایج برای ارائه بهترین پیشنهاد نهایی

(ب) پیاده‌سازی رویکرد پالایش مشارکتی و افزودن برخی ویژگی‌های پالایش محتوا پایه برای دست‌یابی به نتایج باکیفیت‌تر

[1] IDF: Inverse Document Frequency

[2] Content-Based Profile(C)

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

(ج) پیاده‌سازی رویکرد پالایش محتوا پایه و افزودن برخی ویژگی‌های پالایش مشارکتی به‌منظور حصول نتایج باکیفیت‌تر

(د) تجمیع هر دو روش در یک ساختار واحد

پالایش ترکیبی از سازوکارهای مختلفی جهت ترکیب روش‌های مختلف استفاده می‌کند. برخی از این سازوکارها عبارتند از [58]:

مکانیزم وزن‌دار[1]: نتایج (نرخ یا امتیاز) چندین روش پیشنهاددهنده باهم ترکیب می‌شوند تا یک پیشنهاد ساده تولید شود.

مکانیزم راهگزینی[2]: در این روش با توجه به شرایط جاری سیستم یکی از روش‌های پیشنهاددهنده را انتخاب می‌کند.

مکانیزم آمیخته[3]: پیشنهاد از چندین سیستم پیشنهاددهنده متفاوت که در یک زمان نمایش داده شده‌اند، ایجاد می‌شود.

مکانیزم ترکیب خصوصیات[4]: خصوصیات از منابع داده پیشنهاددهنده‌های متفاوت باهم در یک الگوریتم ساده قرار می‌گیرند.

مکانیزم آبشاری[5]: سیستم پیشنهادات دیگر سیستم‌ها را پالایش می‌کند.

مکانیزم افزایش ویژگی‌ها[6]: خروجی یک روش به‌عنوان خصوصیت ورودی سیستم دیگر استفاده می‌شود.

مکانیزم فرا سطح[7]: مدلی که یک سیستم یادگرفته به‌عنوان ورودی دیگران استفاده می‌شود.

هستان‌شناسی در سامانه‌های پیشنهادگر: مدل‌سازی اطلاعات در سطح معنایی یکی از اهداف اصلی استفاده از هستان‌شناسی است [59]. تعریف اولیه هستان‌شناسی در علوم رایانه توسط گربر[8] در سال ۱۹۹۳ میلادی ارائه شد و بعدها توسط استاب[9] و استادر[10] در سال ۲۰۰۹ تصحیح شد. مفهوم هستان‌شناسی در ابتدا توسط گربر [60] به‌عنوان "توصیف صریح از یک مفهوم" ارائه شده است. بورست[11] [61] هستان‌شناسی را به‌عنوان "توصیف رسمی از یک مفهوم مشترک" تعریف می‌کند. علاوه بر این، تانیار[12] و رحاو[13] [62] هستان‌شناسی را "به‌عنوان دانش مفهوم‌سازی دامنه که به‌وسیله کامپیوتر قابل پردازش هستند و قادر است واقعیات، ویژگی‌ها و بدیهیات را مدل‌سازی کند" تعریف می‌کند. بر طبق نظر آنتونیو و ون هارلمن [63]"هستان‌شناسی مشخصاً از یک فهرست لغات و رابطه میان مفاهیم ساخته شده است. هستان‌شناسی شامل ویژگی‌های مفهومی، عبارات، محدودیت‌های مربوط به رابطه‌ها و توصیف روابط منطقی بین اشیاء است. هستان‌شناسی ابزاری است برای ساخت مدل رسمی ساختار یک سامانه، براساس روابط حاصل از مشاهدات در آن.

زمانی که هستان‌شناسی تنها شامل روابط – IS" A" می‌باشد به‌جای عبارت هستان‌شناسی از طبقه‌بندی اصطلاح (سلسله‌مراتب موضوع) استفاده می‌شود و به‌طورمعمول استفاده از کلمه هستان‌شناسی موقعی صحیح است که سامانه مورد بررسی دربرگیرنده انواع روابط متقابل بین مفاهیم (شامل گزاره‌های منطقی که توصیف کننده ارتباط بین مفاهیم هستند) باشد.

الگوریتم‌های مبتنی بر پالایش مشارکتی و هستان‌شناسی که در سامانه‌های پیشنهاد شده استفاده می‌شوند، دو شاخص اساسی برای طبقه‌بندی مقاله‌های پژوهشی انجام‌شده در این حوزه است. مفاهیم هستان‌شناسی به‌عنوان ابزاری جهت تسهیل شناسایی اطلاعات موجود در پروفایل کاربران است، اکثر سامانه‌های پیشنهادگر موفق سعی می‌کنند به‌واسطه دانش هستان‌شناسی، دقت پیشنهادات ارائه‌شده به کاربران را افزایش دهند. بر اساس پژوهشی که توسط میدلتون و همکاران [64] صورت گرفته است، در کنار استفاده از الگوریتم‌های یادگیری ماشین می‌توان سامانه‌های پیشنهادگر را با استفاده از مفاهیم هستان‌شناسی بهبود داد.

امروزه، در سامانه‌های پیشنهادگر علاوه‌بر این‌که از روش‌هایی هم‌چون روش‌های تشخیص آماری الگو، یادگیری ماشین و روش‌های ابتکاری استفاده می‌شود، سعی می‌شود از مفاهیم هستان‌شناسی نیز جهت بهبود نتایج استفاده شود [65]. در اکثر سامانه‌های پیشنهادگر، به‌طورمعمول از روش‌هایی استفاده می‌شود که دارای مزایایی همچون دقت، ارتباط بین اقلام و کالاها، بازخورد و یا قابلیت رصدکردن رفتار کاربر در طول زمان باشند. در

[1] Weighted
[2] Switching
[3] Mixed
[4] Feature Combination
[5] Cascade
[6] Feature Augmentation
[7] Meta-Level
[8] Gruber
[9] Staab
[10] Studer
[11] Borst
[12] Taniar
[13] Rahayu

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

این سامانه‌ها، اقلام جدید می‌توانند با استفاده از جستجوی قلم به قلم، مشابه کاری که در روش پالایش محتوا پایه انجام می‌شود، به کاربران پیشنهاد داده شوند. همچنین ارزیابی یک قلم به‌خصوص برای این‌که مشخص شود آیا برای یک کاربر خاص مناسب هست یا خیر، از طریق بررسی پروفایل جمعی از کاربران، مشابه کاری که در روش پالایش مشارکتی انجام می‌شود، قابل انجام است. همچنین می‌توان از مزایای استفاده از ارتباطات معنایی بین اقلام موجود در پایگاه داده، در روش‌هایی همچون روش پالایش ترکیبی و روش‌های ابتکاری بهره جست [66].

اگر داده‌های آموزشی در دسترس باشد، استفاده از روش پالایش محتوا پایه می‌تواند مؤثر باشد. در مقابل، اگر سامانه دارای تعداد قابل ملاحظه‌ای از کاربران باشد، روش پالایش مشارکتی در مقایسه با روش پالایش محتوا پایه بهتر عمل خواهد کرد. با این وجود، هیچ قاعده و قانونی وجود ندارد که بتوان گفت دقیقاً چه نوع از این روش‌ها را می‌توان مورد استفاده قرار داد [67]. به‌منظور بهبود و توسعه سامانه‌های پیشنهادگر محتوا پایه، می‌توان از روش‌های هستان‌شناسی هم‌چون OntoSeek استفاده کرد [68].

از OntoSeek می‌توان به‌منظور فرموله‌کردن درخواست‌ها استفاده کرد. همچنین سامانه‌های مبتنی بر هستان‌شناسی می توانند برای ایجاد (خودکار) پایگاه‌های دانش، از صفحات وب (به‌عنوان مثال Web-KB) استفاده کنند [69].

در Web-KB نمونه‌هایی از صفحات وب وجود دارند که به‌صورت دستی برچسب‌گذاری شده‌اند و این سامانه قادر است به‌وسیله روش‌های یادگیری ماشین به طور خودکار صفحات جدید وب را رده‌بندی کند. این سامانه‌ها اطلاعات پویا و در حال تغییر مانند علایق کاربران را ذخیره نمی‌کنند. گراف هستان‌شناسی اقلام، می‌تواند ارتباط معنایی اقلام مختلف را نمایان سازد و این امر می‌تواند موجب افزایش اثربخشی پیشنهادهای ارائه شده به کاربران شود.

روش‌های پروفایل‌سازی مبتنی بر هستان‌شناسی که برای مثال در سامانه‌های Foxtrot و Quickstep استفاده می‌شود [70] با هدف پرکردن شکاف معنایی بین ویژگی‌های سطح پایین استخراج‌شده از اسناد و مشاهدات مفهومی مورد علاقه کاربران، انجام می‌شود [71]. درواقع، مفاهیم هستان‌شناسی می‌تواند برای بهبود عملکردِ پروفایل کاربر در سامانه‌های پیشنهادگر توسعه یافته، استفاده شوند.

بسط و گسترش مجموعه لغات با استفاده از هستان‌شناسی، یکی از روش‌هایی است که می‌توان از آن برای پرکردن شکاف معنایی بین مفاهیم استفاده شده در پروفایل کاربران و حاشیه‌نویسی‌های مربوط به تصاویر اقلام، از آن استفاده کرد. به‌طورمعمول از هستان‌شناسی‌های دامنه‌ای برای اتصال مفاهیم، بین پروفایل کاربران و اقلام (از طریق ساختار سلسله‌مراتبی) استفاده می‌شود [72]. بر مبنای پژوهشی که توسط گاچ و همکاران [73]، انجام شده است، می‌توان از طریق ارزیابی رفتار یک کاربر خاص، به‌وسیله اندازه‌گیری پارامترهایی همچون محتوا و زمان صرف‌شده در هر صفحه وب، پروفایل کاربر مورد نظر را بهبود بخشید.

در سامانه‌های پیشنهادگر، اطلاعات معنایی مربوط به یک قلم به‌خصوص، شامل مواردی هم‌چون: ویژگی‌ها، روابط میان اقلام و همچنین رابطه بین اطلاعات غیر نمادین[1] و اقلام، می‌شود. در سال‌های اخیر، روش‌های هستان‌شناسی به‌طور موفقیت‌آمیزی در سامانه‌های پیشنهادگر برای غلبه بر نقص‌های این سامانه‌ها مورد استفاده قرار گرفته‌اند [74]. پورسل و همکاران سعی کردند در روش پیشنهادی خود با استفاده از هستان‌شناسی فازی، دقت سامانه‌های پیشنهادگر را بهبود بخشند [75]. بسیاری از پژوهش‌گران در حوزه سامانه‌های پیشنهادگر به‌منظور اندازه‌گیری میزان علاقه کاربران به ویژگی‌های محتوا پایه‌ی اقلام، از روش‌های هستان‌شناسی استفاده کرده‌اند [76]. علاوه‌براین، همان‌طور که می‌دانیم مشکل شروع سرد که در نتیجه ورود قلم (یا اقلام) جدید و یا کاربر جدید به سامانه اتفاق می‌افتد یکی از چالش‌های اساسی در سامانه پیشنهادگر است، به همین خاطر بسیاری از پژوهش‌گران پیشنهاد می‌کنند برای حل این معضل، روش‌های هستان‌شناسی را با روش‌هایی همچون پالایش مشارکتی و محتوا پایه ترکیب کنیم [77]. هم چنین، برخی از پژوهش‌گران پیشنهاد می‌کنند، به‌منظور بهبود سامانه‌های پیشنهادگر مبتنی بر مفاهیم، می‌توان از ترکیب روش پالایش مشارکتی (قلم محور) با روش‌های شباهت معنایی (همچون هستان‌شناسی) استفاده کرد [78]. در همین راستا نویسندگان [79,76]، یک نمونه اولیه از سامانه پیشنهادگر جهت ارائه خدمات گردش‌گری

[1] Meta-Information

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

الکترونیکی را با استفاده از مفاهیم هستان‌شناسی توسعه داده‌اند. همچنین ونگ و کنگ، سامانه پیشنهادگری مبتنی بر پالایش مشارکتی و هستان‌شناسی را با استفاده از داده‌های موجود در پروفایل کاربران و شباهت‌های معنایی، ارائه داده‌اند [80]. به هر حال می‌توان گفت، استفاده از مفاهیم معنایی مربوط به اقلام و کاربران در یک سامانه پیشنهادگر، می‌تواند در جهت ارائه پیشنهادات مؤثر به کاربران به‌منظور رفع نیازهای آنها، کمک شایانی کند [81]. مطالعه حاضر با هدف استفاده از روش ترکیبی دانش پایه انجام شده است.

راه‌کارهای مطرح‌شده در بالا می‌تواند برای تولید پیشنهادات در سامانه‌ها و حوزه‌هایی که برمبنای روابط معنایی و دانش کار می‌کنند، مورد استفاده قرار گیرد (به ویژه در برنامه‌های کاربردی وب ۲٫۰) [81].

هستان‌شناسی، نوعی مفهوم‌سازی دامنه‌ای است که قابل خواندن برای ماشین نیز می باشد. هستان‌شناسی به‌طورمعمول به‌صورت ساختاری است که روابط میان مفاهیم، عناصر و ویژگی‌ها را نشان می‌دهد. پیداکردن شباهت معنایی بین مفاهیم و در کل هستان‌شناسی به‌عنوان یک ساختار پیچیده (به‌عنوان مثال Flickr) اهمیت دارد [81].

درنهایت می‌توان گفت بهبود ساختار هستان‌شناسی که می‌تواند به افزایش دقت منجر شود، توسط نویسندگان نادیده گرفته شده است. به‌عنوان مثال، تمام روابط "$IS-A$"ها که برای اندازه‌گیری شباهت معنایی در یک درخت سلسله‌مراتبی مورد استفاده قرار می‌گیرد، مشابه یکدیگر در نظر گرفته شده است و این درحالی است که این موضوع، دقت اندازه‌گیری شباهت بین دو مفهوم را کاهش می‌دهد. در نتیجه، دقت سامانه‌های پیشنهادگر برای یافتن اقلام یا کاربران مشابه، تحت تأثیر قرار می‌گیرند.

۳- روش پیشنهادی

سامانه پیشنهادگر ارائه‌شده در این پژوهش مبتنی بر روش ترکیبی است. در روش ترکیبی پیشنهادشده به‌منظور به‌دست‌آوردن نتیجه مطلوب، از دو روش پالایش محتوا پایه و پالایش مشارکتی که خود ترکیبی از روش‌های حافظه پایه و مدل پایه می‌باشد، استفاده شده است.

در بخش پالایش مشارکتی چندین فن از قبیل خوشه‌بندی، استفاده از اطلاعات پروفایل کاربر با کمک فن‌های هستان‌شناسی، هستان‌شناسی اقلام، به‌کاربردن فن شباهت معنایی در هستان‌شناسی برای غلبه بر مشکلاتی همچون پراکندگی داده‌ها، مسأله شروع سرد و بهبود مقیاس‌پذیری و افزایش دقت پیشنهادهای ارائه‌شده، استفاده می‌شود. همچنین در بخش پالایش محتوا پایهٔ روش پیشنهادی، از فن هستان‌شناسی مبتنی بر اقلام و شباهت معنایی استفاده می‌شود. لازم به ذکر است، در هنگام استفاده از فن شباهت معنایی، به‌منظور افزایش دقت در اندازه‌گیری میزان شباهت‌های یال‌های $IS-A$ بین دو قلم شاخص، از یک روش ابتکاری به‌منظور ارائه پیشنهادهای دقیق‌تر به کاربر فعال، استفاده می‌شود.

اهداف اصلی این پژوهش عبارت‌اند از: افزایش دقت و بهبود عملکرد سامانه‌های پیشنهادگر با استفاده از روش ترکیبی (پالایش مشارکتی و محتوا پایه) و هستان‌شناسی بهبودیافته. به‌طور خلاصه، مطالعه حاضر به‌منظور رسیدن به اهداف زیر انجام می‌شود:

۱. افزایش دقت شباهت معنایی به‌وسیله حذف یال‌های همسان در گراف هستان‌شناسی
۲. بهبود مسأله مقیاس‌پذیری با استفاده از خوشه‌بندی بهبودیافته
۳. ترکیب خوشه‌بندی و هستان‌شناسی بهبودیافته به‌منظور افزایش دقت و کارایی بخش پالایش مشارکتی
۴. دست‌یابی به عملکرد مناسب روش مبتنی بر مدل و دقت بالای روش حافظه پایه با ترکیب روش‌های مدل پایه و حافظه پایه با استفاده از هستان‌شناسی در بخش پالایش مشارکتی

در شکل (۹)، ساختار کلی سامانه پیشنهادگر ترکیبی نشان داده شده است. همان‌طور که در این شکل دیده می‌شود، روش ترکیبی پیشنهاد شده شامل دو بخش اصلی هست: پالایش محتوا پایه و پالایش مشارکتی که ترکیبی از روش‌های حافظه پایه و مدل پایه است.

Input $t, \{C_1^t, C_2^t, \dots, C_{N_1}^t\}, R, L_{1:N_2}^t$.

C_i^t *indicates the ith most similar cluster to the targeted user (i.e. tth user)*

L_i^t *indicates the index of the ith most similar item to the targeted user (i.e. tth user)*

$R_{t,i}$ *indicates the rate that the targeted user has given to the ith item*

1. $I^t = \{j: 1 \dots n | R_{t,j} \neq NaN\}$

2. $S = \{\ \}$

3. for any $I_i^t \in I^t$

3.1. $L_{I_i^t}^q$ *=The index of the* qth *most similar item to the* I_i^tth *item*

3.2. $\left\{C_1^{L_{I_i^t}^q}, C_2^{L_{I_i^t}^q}, \dots, C_l^{L_{I_i^t}^q}\right\}$ *= Retrieve Clusters Using "User Clustering" Feature of the* I_i^tth *Item Ontology*

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

3.3. for any G in $\left\{C_1^{L_{I_i^t}^q}, C_2^{L_{I_i^t}^q}, \dots, C_l^{L_{I_i^t}^q}\right\}$

3.3.1. $if\left((G \in C_1^t)|(G \in C_2^t)|\dots|(G \in C_{N_1}^t)\right)$

3.3.1.1. $S = S \cup G$

3.3.1.2. $break$

4. Use Semantic Similarity to Find KNN Users of the Targeted User in Cluster S and denote it by KNN^t

5. Use UPs of all the users in KNN^t *to predict any missing* $R_{t,i}$

(شکل-۹): روش پیشنهادی

(Figure-9): The proposed algorithm

در روش پیشنهادی، بخش پالایش مشارکتی از یک‌سو، از پایگاه داده هستان‌شناسی فیلم و اطلاعات ضمنی کاربران برای ساخت هستان‌شناسی پروفایل کاربران استفاده می‌کند و از سوی دیگر، از اطلاعات رتبه‌بندی صریح کاربران به‌عنوان دانش مفید در مرحله خوشه‌بندی استفاده می‌شود و هستان‌شناسی را تکمیل می‌کند.

در بخش پالایش محتوا پایه از دانش موجود در مخزن هستان‌شناسی فیلم‌ها جهت مشخص‌کردن میزان درجه IS − Aهای استفاده‌شده در هستان‌شناسی استفاده می‌شود که درنتیجه منجر به وزن‌دارشدن مفاهیم موجود در درخت سلسله‌مراتبی هستان‌شناسی می‌شود و به‌این‌ترتیب می‌توان بهترین اقلام را برای پیشنهاد به کاربر هدف، مشخص کرد. همان‌طور که ملاحظه می‌شود، شکل (۹) به‌وضوح نشان می‌دهد که اندازه‌گیری درجه IS − A ها، قبل از اندازه‌گیری میزان شباهت معنایی، یکی از مراحل اصلی در روش ترکیبی پیشنهادی است. خوشه‌بندی، هستان‌شناسی پروفایل کاربر و الگوریتم پیشنهادی حافظه پایه (الگوریتم بهبود یافته KNN) در بخش پالایش مشارکتی استفاده می‌شوند.

در ادامه، فرایندی که در هر یک از مراحل اجرای سامانه پیشنهادی رخ می‌دهد و اینکه چه مؤلفه‌هایی در هر فرایند درگیر هستند، توضیح داده می‌شود.

برای تسهیل مطالعه الگوریتم پیشنهادی، فهرستی شامل کلمات اختصاری استفاده شده در این متن، در جدول (۲) آورده شده است.

(جدول-۲): فهرست کلمات اختصاری استفاده شده

(Table-2): List of Used Acronyms

ACRONYMS	DESCRIPTIONS
RS	Recommender Systems
CONF	Content based Filtering
COLF	Collaborative Filtering
HRS	Hybrid Recommender System
IAD	"IsA" Degree
TF-IDF	Term Frequency*Inverse Document Frequency
UP	User Profile
IP	Item Profile
IDF	Inverse Document Frequency
TF	Term Frequency
UCBP	User Content-Based Profile
ICBP	Item Content-Based Profile
CT	Concept Tree
CHC	Children Concept
PAC	Parent Concept

در شکل (۹)، C_i^t شبیه‌ترین خوشه به کاربر مورد نظر را نشان می‌دهد (یعنی tامین کاربر) I_i^t شاخص iامین مشابه‌ترین آیتم به کاربر مورد نظر را نشان می‌دهد و $R_{t,i}$ نرخ کاربر هدف درنظر گرفته شده به iامین آیتم است. این یک طرح کلی از HRS ما را به تصویر می‌کشد.

همان‌طور که گفته شد، روش خوشه‌بندی کاربر یک روش مبتنی بر مدل برای افزایش عملکرد ColF مبتنی بر KNN است. در این حالت، کاربرانی که یک قلم را خریداری یا علاقه‌مند شده‌اند معرفی می‌شوند. سپس، k نزدیکترین کاربر همسایه مورد نظر بر اساس ویژگی "خوشه‌بندی کاربر" از طریق الگوریتم پیشنهادی شناسایی می‌شود تا k کاربرانی که نزدیکترین کاربر هستند را شناسایی کند. این روش در بخش‌های بعدی به تفصیل شرح داده شده است.

۱-۳- پالایش محتوا پایه و پالایش مشارکتی

در بخش پالایش محتوا پایه، یکنواختی تمام روابط IS − A موجود در هستان‌شناسی به کمک اندازه‌گیری درجه IS − A تمام یال‌ها حذف می‌شود. پس از آن، شباهت معنایی بین دو مفهوم (بر اساس وزن‌های داده‌شده) به‌منظور تعیین اقلام مشابه با پروفایل کاربر هدف، مورد ارزیابی قرار می‌گیرد. از سوی دیگر، خوشه‌بندی و هستان‌شناسی بهبودیافته، الگوریتم پیشنهادی جهت یافتن کاربران مشابه (KNN بهبودیافته) و یافتن شباهت معنایی بین دو هستان‌شناسی، فرایندهای اصلی در بخش پالایش مشارکتی روش پیشنهادی می‌باشند. به‌منظور خوشه‌بندی کاربران از اطلاعات مربوط به رتبه‌بندی (صریح) کاربران و ویژگی‌های محتوای فیلم‌ها، استفاده می‌شود. در روش خوشه‌بندی پیشنهادشده، هم‌پوشانی در خوشه‌بندی و دیگر اشکالات مطرح در خوشه‌بندی سنتی، حذف می‌شوند. در مرحله بعد، هستان‌شناسی اقلام با استفاده از مرحله خوشه‌بندی، بهبود می‌یابد. در این مرحله یک ویژگی به نام "User-Clustering" به هستان‌شناسی اقلام اضافه خواهد شد. این ویژگی از اقلام، شامل کاربرانی می‌باشد که قلم موردنظر را خریداری کرده یا به آن علاقه‌مند هستند. در ادامه،

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

کاربرانی که بیشترین شباهت را با کاربر هدف دارند (k نزدیک‌ترین کاربران همسایه به کاربر هدف)، بر اساس ویژگی مذکور، شناسایی می‌شوند. در مرحله بعد، تعداد N قلم برتر با توجه به نیازها و علایق کاربران مشابه با کاربر هدف، مشخص می‌شوند. برخلاف الگوریتم‌های سنتی، برای تشخیص k نزدیک‌ترین کاربران همسایه به کاربر هدف، علاوه بر این‌که لازم نیست همه خوشه‌ها را برای یافتن خوشه کاربران مشابه مورد جستجو قرار دهیم، بلکه نیازی نیست تا تمام کاربران موجود در خوشه کاربران مشابه را بررسی کنیم. برای این منظور فقط آن خوشه‌هایی مورد بررسی قرار می‌گیرند که در مجموعه خوشه‌هایی با شباهت بیشتر نسبت به کاربر هدف قرار دارند. همچنین به‌منظور پیدا کردن شبیه‌ترین کاربران با کاربر هدف درون خوشه موردنظر، فقط کاربرانی مورد بررسی قرار می‌گیرند که بر اساس ویژگی "User-Clustering" درون یک خوشه مشترک با کاربر هدف باشند.

I. $Such *PaCs* as *ChC*({, *ChC*}*(, |)(or|and) *ChC*$|$)

Such *PaCs* as ChC_1, ChC_2, ChC_3,… ChC_{n-1} and ChC_n.

II. $*ChC*{, *ChC*}*(, |)(or|and) other *PaCs*$

ChC_1, ChC_2, ChC_3,… ChC_{n-1}, ChC_n, or other *PaCs*.

III. $*PaCs*(, |)especially *ChC*({, *ChC*}*(, |)(or|and) *ChC*$|$)

PaCs, especially ChC_1, ChC_2,… ChC_{n-1} and ChC_n.

IV. $*PaCs*(, |)including *ChC*({, *ChC*}*(, |)(or|and) *ChC*$|$)

PaCs including ChC_1, ChC_2,… ChC_{n-1}, or ChC_n.

(شکل-۱۰): استخراج ChC های برای PaC . Hyponyms(“ChC_i","PaC") برای همه $1 \le i \le n$ که حداقل یکی از قوانین مذکور رعایت شده است

(Figure-10): Extracting ChC(s) for a PaC. We can conclude Hyponyms(“ChC_i","PaC"), for all $1 \le i \le n$ given at least one of rules I-IV is hold

۱-۱-۳- مخزن هستان‌شناسی فیلم و کاربران

آدرس اینترنتی (URL) فیلم‌های مختلف در وب‌گاه IMDb (پایگاه داده فیلم بر روی شبکه اینترنت) یک شاخص منحصربه‌فرد است که می‌تواند نمایان‌گر یک‌قلم (فیلم) به‌خصوص در سامانه باشد. با استفاده از یک نوع سرویس Web Crawler و استفاده از URLهای منحصربه‌فرد هر فیلم، می‌توان ویژگی‌های محتوا پایهٔ فیلم‌ها را از پایگاه داده IMDb استخراج کرد. این ویژگی‌ها به‌منظور تولید فراداده مبتنی بر هستان‌شناسی[1] در پایگاه داده ذخیره می‌شوند. درواقع، سرویس Web Crawler صفحات وب IMDb را بر اساس ویژگی‌های مهم هر فیلم که از پیش تعیین شده‌اند، مورد تجزیه و تحلیل قرار می‌دهد. در این پژوهش، ده ویژگی مهم از اقلام (فیلم‌ها) مورداستفاده قرار می‌گیرد که عبارت‌اند از: ژانر، بازیگران، کشور سازنده، زمان انتشار، زمان اکران، رتبه IMDb، رنگ، کارگردان، نویسنده و زبان فیلم. پس از آن، هستان‌شناسی کاربران بر اساس هستان‌شناسی اقلام و رتبه‌بندی ضمنی کاربران و همچنین بازخورد کاربران (رتبه‌بندی صریح) که از طریق Web Proxy جمع‌آوری شده است، ساخته می‌شود (شکل ۱۱).

۲-۱-۳- اندازه‌گیری درجه IS – A (وزن یال‌ها)

در این مرحله، ابتدا هستان‌شناسی مربوط به اقلام بایستی تولید شود. برای طراحی هستان‌شناسی، از مفهوم درخت[2] استفاده می‌شود که در آن رابطه بین اقلام با رابطه "IS-A" مشخص می‌گردد. هر گره درخت یک قلم را نشان می‌دهد و هر یال رابطه والد_فرزندی بین دو گره را نشان می‌دهد. CT یکی از ساده‌ترین مدل‌های هستان‌شناسی است که در پژوهش حاضر مورد استفاده قرار می‌گیرد. برای طبقه‌بندی اقلام در این پژوهش از روش کدینگ UNSPSC استفاده می‌شود. هر قلم دارای ویژگی‌های خاص خود مانند کد محصول در UNSPC و کد WordNet می‌باشد. علاوه‌بر ویژگی‌های ذکرشده، در این تحقیق از ویژگی منحصربه‌فرد دیگری به نام "درجه IS-A" برای اقلام استفاده می‌گردد. این ویژگی نشان می‌دهد

[1] Ontology-Based Metadata

[2] CT: Concept Tree

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

که چقدر یک فرزند توسط والدش پشتیبانی می‌شود. الگوریتم زیر به‌منظور کشف خودکار رابطه بین دو مفهوم مورد استفاده قرار می‌گیرد. پیشنهاد می‌شود که وزن یال‌ها به‌عنوان درجه "IS − A" در نظر گرفته شود.

معادله (۱) که در پیوست آورده شده است، برای اندازه‌گیری "درجه IS − A" بین دو مفهوم پیشنهاد می‌شود؛ به‌طوری‌که $ChCSet(\boldsymbol{C}, i)$، مجموعه تمام iامین های سطح ChC از مفهوم Cرا نشان می‌دهد (برای رسمی کردن، $ChCSet(\boldsymbol{C}, 1)$، مجموعه تمام ChCهای مفهوم C است؛ و $ChCSet(\boldsymbol{C}, i) = \cup_{C' \in ChCSet(C, i-1)} ChCSet(\boldsymbol{C}', 1)$ و $SChCSet(\boldsymbol{C_2}|RD_d^{C_2})$ مجموعه تمام ChC های مفهوم C است در dامین سند مربوط به مفهوم C موجود در الگوریتم جستجو، $RD_d^{C_2}$ سند dام مربوط به مفهوم C موجود در الگوریتم جستجو است، β یک عدد صحیح برابر با ۳ در این مقاله است، θ پارامتر می‌باشد، و $\pi(.)$ بر اساس معادله (۸) تعریف می‌شود.

$$\pi(Condition) = \begin{cases} 1 & if\ Condition\ is\ true \\ 0 & otherwise \end{cases} \quad (۸)$$

الگوریتم پیشنهادی برای اندازه‌گیری درجه IS − A شامل مراحل زیر است:

- در مرحله اول، تمام فرزندان یک مفهوم والد با استفاده از استاندارد UNSPSC در درخت سلسله‌مراتبی مفاهیم مشخص می‌شوند.

سپس مجموعه‌ای از اسناد مرتبط با سر واژه (مفهوم والد) را می‌یابیم. برای این منظور، می‌توان از روش‌هایی که برای کشف خودکار ارتباط بین مفاهیم، کاربرد دارد استفاده کنیم [72]. "اسناد مرتبط" به معنی اسناد مربوط به مفهوم والد است. برای یافتن اسناد مربوط به مفهوم والد، عبارات زیر باید در Google جستجو شوند [72 , 73]. در شکل (۱۰)، عبارت "PaC" به معنای "مفهوم والد" و عبارت "ChC" به معنی "مفهوم فرزند" است. همچنین عبارت ("ChC","PaC") نشان‌دهنده "ChC IS − A PaC" است. هر عبارت (فرزند) زیرمجموعه یک عبارت دیگر (والد) باشد، لذا لازم است بدون در نظر گرفتن یک عبارت (فرزند) خاص، به جستجوی عبارات مدنظر بپردازیم. به‌عنوان‌مثال، عبارات زیر باید برای پیداکردن اسناد مرتبط و تعیین درجه "IS − A" برای فرزندان عبارت "وسایل نقلیه موتوری" جستجو شود (شکل ۱۰).

- برای بقیه عبارات، می‌توان تعداد معینی از صفحات یا اسناد یافت شده برای هر عبارت را به‌منظور یافتن مفاهیم موردنظر، بررسی نماییم. برای مثال، می‌توان اولین ۱۰۰۰ صفحه یافته شده برای هر عبارت را انتخاب کنیم.
- در این مرحله، زیرمجموعه (فرزند) یک عبارت خاص با استفاده از روابط مرحله ۲ جستجو می‌شود. برای هر عبارت (فرزند) که در اسناد یافته شده در مرحله ۳ به‌دست می‌آید (و همچنین در روابط مرحله ۲ صادق است)، یک امتیاز مثبت برای ارزش آن عبارت نسبت به والدش، اضافه می‌شود.

مثال۱: فرزندان عبارت مفهومی "خودروهای مسافربری" عبارت‌اند از مینی‌بوس، اتوبوس، ماشین، واگن استیشن، مینی ون یا ون، لیموزین، کامیون‌های سبک و اتومبیل‌های ورزشی. تعداد آن‌ها در فرایند جستجو توسط الگوریتم مذکور به‌ترتیب ۳۴۴، ۲۴۷، ۸۰۱، ۳۰، ۰، ۱۵، ۲۰۵ و ۵ مشخص شد.

- در مرحله بعدی فرض می‌شود که "ChC" یک فرزند برای عبارت "PaC" است و سپس قصد داریم میزان "ChC IS − A PaC" را اندازه‌گیری کنیم. برای این کار، بایستی امتیازات به‌دست‌آمده برای ChC را به مجموع امتیازات تمام فرزندان PaC تقسیم کنیم.

لازم به ذکر است برخی از عبارات فرزند، ممکن است در روابط مرحله ۲ قرار نداشته باشند. بنابراین، وزن پیش‌فرض یال‌ها را می‌توان برابر ۱ فرض کرد و سپس امتیاز به‌دست‌آمده در مرحله ۵ را با آن جمع کرد تا وزن یال موردنظر به‌دست آید.

مثال ۲: فرزندان عبارت مفهومی "خودروهای مسافربری" عبارت‌اند از مینی‌بوس، اتوبوس، ماشین، واگن استیشن، مینی‌ون یا ون، لیموزین، کامیون‌های سبک و اتومبیل‌های ورزشی که مقادیر IS − A آن‌ها نسبت به والدشان به‌ترتیب ۱,۲۰۹۴، ۱,۱۵۰۳، ۱,۴۸۷۵، ۱,۰۱۸۳، ۱,۰۰۹۱، ۱,۱۲۴۸ و ۱,۰۰۰۶ می‌باشد که با استفاده از الگوریتم مذکور اندازه‌گیری شده است. در بعضی از اسناد ممکن است فرزندِ "عبارتِ فرزند" به‌جای "عبارتِ فرزند" در مرحله ۲ تعیین شود. در این صورت یک امتیاز (با ضریب 1/k (مثلاً k = 2)) برای "عبارت فرزند اصلی" در نظر گرفته می‌شود. به‌عنوان‌مثال، چنان‌که در درخت سلسله‌مراتبی داشته باشیم: ChC1 IS − A PaC و ChC2 IS − A ChC1، و در هنگام جستجوی اسناد با عبارت: "PaC شامل ChC2 است" روبه‌رو شویم، این بدان معنی است که رابطه "ChC1 IS − A PaC" صحیح است و یک امتیاز با ضریب $1/k$ برای آن در نظر گرفته می‌شود.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

۳-۱-۳- اندازه‌گیری شباهت معنایی

طبقه‌بندی مفهومی موجودیت‌ها بر اساس میزان شباهت معنایی میان آن‌ها، یکی از مراحل مهم در فرایند ساخت مدل هستان‌شناسی می‌باشد، این کار با توجه به جایگاه هر موجودیت خاص در درخت سلسله‌مراتبی قابل انجام است. اساساً، در فرایند طبقه‌بندی مفهومی می‌توان گفت مفاهیمی که در درخت سلسله‌مراتبی به یکدیگر نزدیک‌ترند، نسبت به هم شبیه‌تر هستند. بااین‌حال، برای محاسبه دقیق میزان شباهت بین مفاهیم، لازم است وزن یال‌های بین دو موجودیت مختلف، محاسبه شود. زیرا شباهت معنایی در طبقه‌بندی مفهومی، بر اساس وزن یال‌ها اندازه‌گیری می‌شود.

برای اندازه‌گیری شباهت معنایی بین دو عبارت در درخت سلسله مراتبی وزن‌دار (یال‌های بین موجودیت‌ها وزن‌دار شده‌اند)، نخستین گام، محاسبه فاصله بین دو عبارت است که این کار با کمک وزن‌های به‌دست‌آمده در مراحل قبلی قابل انجام است. فاصله بین دو عبارت با استفاده از رابطه $Dis(C_1, C_2) = [IaD(C_1, C_2)]^{-1}$ محاسبه می‌شود. که در آن $Dis(C_1, C_2)$ نشان‌دهنده فاصله بین دو عبارت C_1 و C_2 است.

میزان شباهت معنایی بین دو عبارت C_1 و C_2 را می‌توان به‌وسیله رابطه (۹) محاسبه کرد:

$$Sim(C_1, C_2) = \frac{2 \times Dis(\dot{R}, \dot{S}_{C_1,C_2})}{Dis(C_1, \dot{R}) + Dis(C_2, \dot{R})} \times V_{C_1,C_2} \qquad (۹)$$

در رابطه بالا، $\dot{R}$ ریشه درخت را نشان می‌دهد. $\dot{S}_{C_1,C_2}$ نشان‌دهنده فاصله بین عبارت A و نزدیک‌ترین والد مشترک دو عبارت C_1 و C_2 در درخت سلسله‌مراتبی است (مجموع طول یال‌ها) و درنهایت V_{C_1,C_2} مطابق رابطه (۱۰) محاسبه می‌شود:

$$V_{C_1,C_2} = \begin{cases} \dfrac{1}{Dis(C_1, \dot{S}_{C_1,C_2}) - Dis(C_2, \dot{S}_{C_1,C_2})} & C_1 \text{ is an ancestor of } C_2 \\ \dfrac{1}{Dis(C_2, \dot{S}_{C_1,C_2}) - Dis(C_1, \dot{S}_{C_1,C_2})} & C_2 \text{ is an ancestor of } C_1 \\ \dfrac{1}{Dis(C_1, \dot{S}_{C_1,C_2}) + Dis(C_2, \dot{S}_{C_1,C_2})} & otherwise \end{cases} \qquad (۱۰)$$

این معادله نشان می‌دهد که فاصله بین مفاهیم خواهر و برادر بیشتر از فاصله بین ChC و PaC آن است.

با توجه به رابطه مذکور می‌توان به دو نکته اساسی پی برد:

- تفاوت معنایی بین سطوح بالاتر در درخت سلسله مراتبی بیشتر از تفاوت معنایی بین سطوح پایین‌تر است.
- فاصله بین عباراتی که والد مشترک دارند بیشتر از فاصله بین فرزندان و والد است.

در ادامه، با استفاده از روش پالایش محتوا پایه، فهرستی از برترین اقلام بر اساس سوابق گذشته کاربر هدف آماده می‌شود. روش پالایش استفاده‌شده در اینجا بر اساس شباهت معنایی می‌باشد که در مراحل قبلی توضیح داده شد.

در بخش پالایش مشارکتی، به‌منظور خوشه‌بندی اقلام از هر دو ویژگی محتوا پایه (اطلاعات ضمنی) و رتبه‌بندی کاربران (اطلاعات صریح) استفاده می‌شود، زیرا تنها توجه به یکی از این نوع ویژگی‌ها، منجر به مشکلاتی همچون: کاهش دقت، تعمیم‌پذیری بیش‌ازحد[1] و همپوشانی خوشه‌ها خواهد شد. در این پژوهش، جهت خوشه‌بندی می‌توان از الگوریتم‌های متفاوتی استفاده کرد؛ در این میان الگوریتم k-means یکی از ساده‌ترین و قدیمی‌ترین الگوریتم‌های یادگیری است که در زمره الگوریتم‌های بدون نظارت است [73]. به این الگوریتم، الگوریتم Lloyd نیز گفته می‌شود [74].

این روش خوشه‌بندی مدل پایه مرکز خوشه است. در روش خوشه‌بندی k-means هدف این است که مجموعه اقلام به k خوشه مختلف تقسیم شوند. مراکز خوشه‌ها را می‌توان از پیش تعریف کرد. برای اینکه عمل خوشه‌بندی به‌درستی انجام شود، بهتر است مراکز خوشه‌های اولیه از یکدیگر فاصله داشته باشند.

مراحل اجرای الگوریتم k-means به‌صورت زیر است:

گام اول: تعیین k نقطه (امتیاز) در کل فضای مجموعه داده‌های مربوط به رتبه‌بندی کاربران. این k نقطه، نشان‌دهنده مراکز اولیه خوشه‌ها است.

گام دوم: محاسبه فاصله بین داده‌ها (رتبه‌بندی کاربران) و مراکز خوشه‌ها

گام سوم: انتقال هر یک از داده‌ها (رتبه‌بندی‌های کاربران) به نزدیک‌ترین خوشه

گام چهارم: پس از اختصاص تمام داده‌ها (رتبه‌بندی‌ها)، به خوشه‌های مربوطه، مراکز خوشه‌ها مجدداً تعریف می‌شوند.

گام پنجم: مراحل ۲، ۳ و ۴ تکرار می‌شوند و این کار تا زمانی که مراکز خوشه‌ها دیگر تغییری نکنند، ادامه می‌یابد.

بر اساس ویژگی‌های محتوایی اقلام، ابتدا خوشه‌ها را با استفاده از رویکرد خوشه‌بندی انتخابی، تولید می‌کنیم. برای ساخت نمونه اولیه سامانه پیشنهادگر فیلم

[1] Overgeneralization

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

در این پژوهش، ابتدا ویژگی ژانر انتخاب می‌شود. برای این منظور وزن هر ژانر برای هر کاربر باید محاسبه شود. برای رسیدن به این هدف، ابتدا می‌بایست هر قلم را به‌صورت برداری نمایش دهیم.

	Genre 1	Genre 2	Genre 3	Genre 4	…	Genre k
$ITEM_i$	0/1	0/1	0/1	0/1	…	0/1

If $ITEM_i[j]=1 \Rightarrow ITEM_i$ includes Genre j
Else $ITEM_i$ doesn't include Genre j

پس از آن، وزن هر ژانر باید برای هر کاربر خاص مانند کاربر k-ام اندازه‌گیری شود. اگر کاربر k-ام قلم $ITEM_i$ را با وزن W رتبه‌بندی کرده باشد، بنابراین وزن هر ژانر در $ITEM_i$ برای کاربر k از طریق بردار زیر محاسبه می‌شود:

$W^*_{\tilde{i}}$	W*(0/1)	W*(0/1)	W*(0/1)	…	W*(0/1)

جهت تعیین وزن کلی یک ژانر به‌خصوص برای همه اقلامی که کاربر k-ام آن‌ها را رتبه‌بندی کرده است، می‌توان از رابطه (۱۱) استفاده کرد:

(۱۱)

$$\text{wieght of user } k \text{ for genre } x = \frac{\text{sum weigh for genre } x \text{ among all items that user } k \text{ has rated}}{\text{sum weigh for all genres among all items that user } k \text{ has rated}}$$

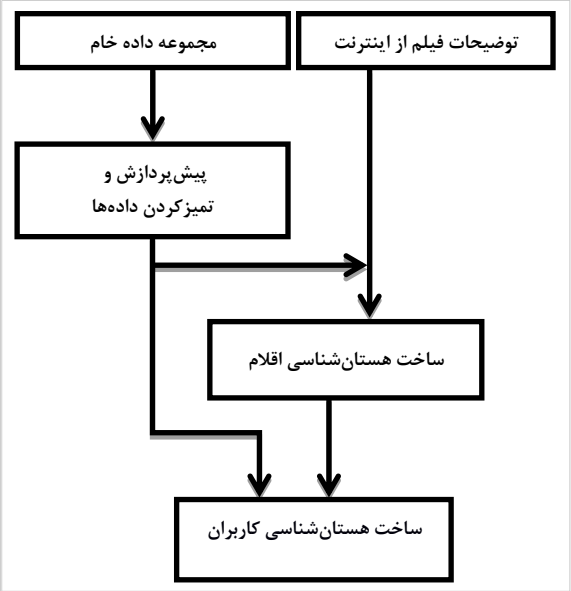
در مرحله بعد، ماتریس کاربر_ژانر به‌دست می‌آید و نشان می‌دهد که وزن هر ژانر برای هر کاربر به‌صورت زیر است.

$$M = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,k} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,k} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ w_{n,1} & w_{n,2} & \cdots & w_{n,k} \end{bmatrix} \qquad (۱۲)$$

در رابطه بالا، k و n به‌ترتیب تعداد ژانرها و تعداد کاربران را نشان می‌دهند و w_{ij} نشان‌دهنده وزن ژانر j-ام برای کاربر i-ام است. در مرحله بعد، با استفاده از الگوریتم k-means (بر اساس ماتریس کاربر_ژانر)، خوشه‌بندی کاربران انجام می‌شود.

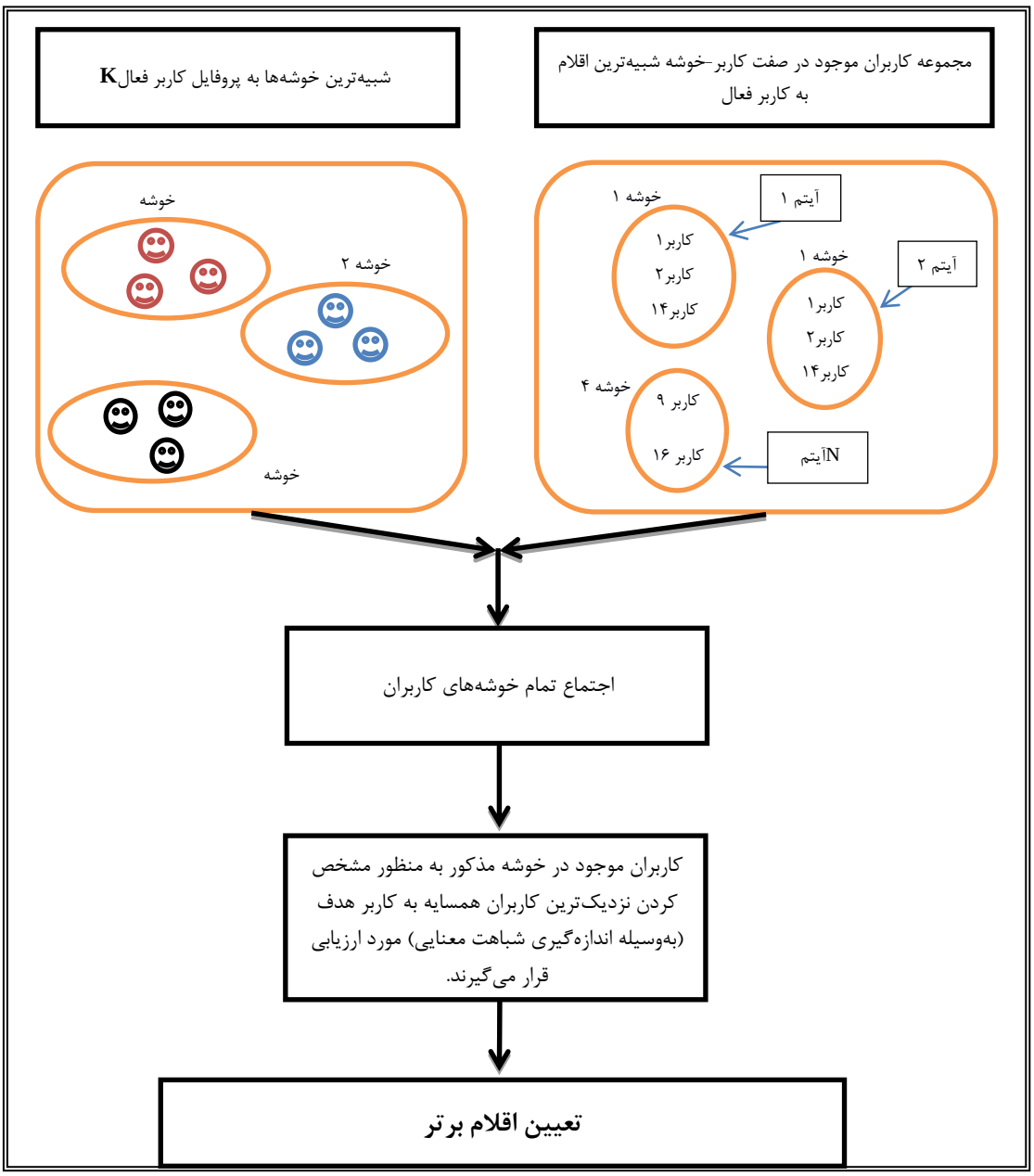
۴-۱-۳- تکمیل هستان‌شناسی

یکی از مهم‌ترین کارهایی که در این پژوهش انجام‌شده است، تکمیل هستان‌شناسی مبتنی بر خوشه‌بندی کاربران به‌منظور دست‌یابی به پالایش مشارکتی کارآمد و دقیق مبتنی بر خوشه‌بندی است. شکل (۱۱) ساختار هستان‌شناسی را نشان می‌دهد. به‌عنوان مثال، تمام کاربرانی که یک قلم به‌خصوص را خریداری کرده‌اند، در یک خوشه مشترک قرار می‌گیرند. به‌عنوان‌مثال، فرض کنید ITEM X توسط کاربران U1، U2، U3، U4، U5، U6، U7 و U8، خریداری شده است. با توجه به ویژگی User_Clustering، کاربران U1، U2 و U3 در خوشه C1، و کاربران U4، U5 و U6 در خوشه C2 و همچنین کاربران U7 و U8 در خوشه C3 خوشه‌بندی می‌شوند.



(شکل-۱۱): ساختار هستان‌شناسی
(Figure-11): Constructing ontology

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]



(شکل-۱۲): پیداکردن شبیه‌ترین کاربران همسایه با کاربر هدف
(Figure-12): Finding the most similar neighbor users to the target user

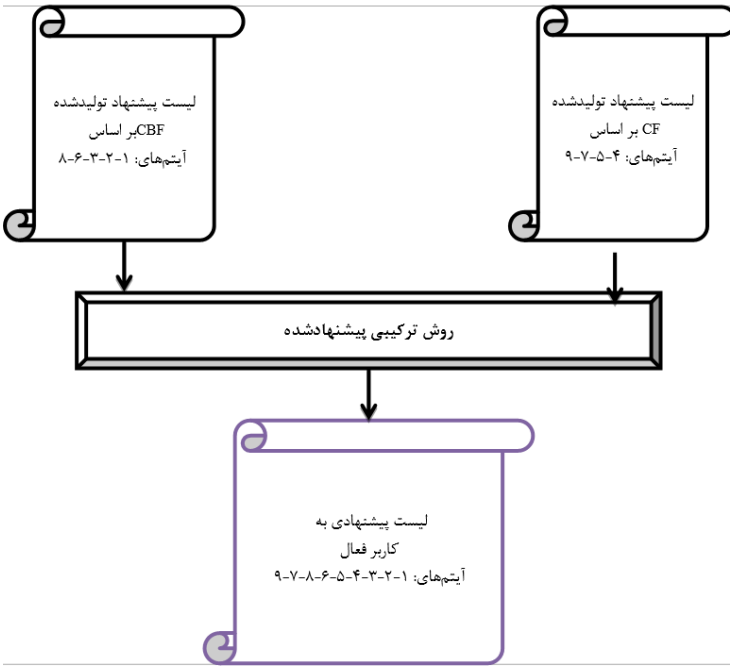
۵-۱-۳- الگوریتم پیشنهادی جهت پیداکردن کاربران مشابه

در این پژوهش به‌منظور تهیه فهرستی از اقلام قابل پیشنهاد بر اساس پالایش مشارکتی، می‌بایست شبیه‌ترین کاربران همسایه با کاربر هدف را تعیین کنیم. برای این منظور، می‌توان از الگوریتم‌های متفاوتی استفاده کرد. الگوریتم *KNN* یکی از روش‌هایی است که می‌توان از آن برای رسیدن به این هدف استفاده کنیم. البته به‌منظور بهبود و افزایش کارایی الگوریتم مذکور، در این پژوهش راه کاری ارائه می‌شود که جهت یافتن کاربران همسایه با کاربر هدف، نیازی به جستجو در میان همه کاربران نباشد. برای انجام این کار، تنها کاربرانی مورد بررسی قرار می‌گیرند که در میان کاربران خوشه‌بندی شده بر اساس ویژگی «User_Clustering» (برای اقلامی که کاربر هدف خریداری کرده است) وجود دارند. علاوه بر این، لازم نیست همه خوشه‌های کاربران مورد بررسی قرار گیرند. برای پیداکردن کاربران همسایه با کاربر هدف، تنها خوشه‌هایی مورد بررسی قرار می‌گیرند که درون خوشه‌هایی با بیشترین شباهت با کاربر هدف وجود داشته باشند. با توجه به راه‌کار پیشنهادی بالا، انتظار می‌رود مقیاس‌پذیری و دقت در الگوریتم مذکور بهبود یابد و همین امر موجب بهبود عملکردِ کل سامانه پیشنهادگر شود. الگوریتم پیشنهادی جهت یافتن کاربران مشابه در شکل (۱۲) آورده شده است.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

در الگوریتم فوق، جهت به‌دست‌آوردن k نزدیک‌ترین کاربران همسایه به کاربر هدف، پروفایل کاربر هدف با پروفایل کاربران دیگر به‌وسیله شباهت معنایی بین هستان‌شناسی‌ها، مقایسه می‌شود [75]. در فن شباهت معنایی استفاده‌شده در این پژوهش، برای اندازه‌گیری شباهت بین دو هستان‌شناسی، هم شباهت‌های واژگان و هم عبارات مفهومی در نظر گرفته می‌شود. مقایسه‌های مفهومی، شامل مقایسه بین دو طبقه‌بندی و مقایسه روابط بین مفاهیم متناظر دو طبقه‌بندی است. پس از تولید مجموعه k نزدیک‌ترین کاربران همسایه، تمام اقلامی که توسط کاربران همسایه خریداری شده‌اند اما توسط کاربر هدف خریداری نشده است، به کاربر هدف توصیه می‌شود.

در ادامه، فهرست نهایی پیشنهادها را می‌توان با توجه به ترکیبی از فهرست پیشنهادهای تولیدشده در فاز پالایش محتوا پایه و فاز پالایش مشارکتی بر اساس رویکرد وزنی، به کاربر فعال ارائه کرد [76] (شکل ۱۳).



(شکل-۱۳): روش ترکیبی مبتنی بر CBF و CF
(Figure-13): Combined method based on CBF and CF

۴- مطالعات تجربی

در این فصل، قصد داریم الگوریتم پیشنهادی را با کمک یک پایگاه داده واقعی، مورد ارزیابی قرار دهیم. بر این اساس، می‌توان نتایج حاصل از روش پیشنهادی را با نتایج روش‌های دیگر در این حوزه مقایسه کرد.

سامانه پیشنهادگر (الگوریتم) پیشنهادی در این مقاله و دیگر الگوریتم‌های مورد مقایسه، با استفاده از نرم‌افزار Mathworks Matlab R2019a تحت پردازنده Intel Core i9-9900K Tray و حافظه TeamGroup XCALIBUR 16GB 8GB*2 4000Mhz و سامانه‌عامل Windows Server 2019 Version 1809 Build 17763.805 Retail مورد بررسی قرار می‌گیرند؛ هدف از این ارزیابی، پاسخ دادن به سؤالات زیر است:

۱.آیا ترکیب هستان‌شناسی، حافظه و محتوا در سامانه پیشنهادگر می‌تواند منجر به بهبود سامانه پیشنهادگر شود؟

۲.آیا ترکیب هستان‌شناسی، حافظه و محتوا در سامانه پیشنهادگر می‌تواند منجر به یک سامانه پیشنهادگر مقیاس‌پذیر شود؟

۳.آیا ترکیب هستان‌شناسی، حافظه و محتوا در سامانه پیشنهادگر می‌تواند بر مشکلات شروع سرد در زمینه سامانه پیشنهادگر فائق آید؟

۴-۱- پایگاه داده

برای انجام ارزیابی و آزمایش الگوریتم‌ها، ابتدا مجموعه‌داده از مجموعه‌داده Movielens جمع‌آوری شد. پس از آن داده‌ها پالایش شده‌اند تا در فرآیند ارزیابی مورد استفاده قرار گیرند. هر کاربر در این مجموعه‌داده حداقل ۲۰ فیلم را رتبه‌بندی کرده است. شرح مجموعه‌داده به شرح زیر ارائه شده است: تعداد فیلم‌ها و کاربران مجموعه‌داده‌های Movielensبه‌ترتیب ۳،۷۰۶ و ۶،۰۴۰ نفر است.کاربران

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

امتیازات صحیح از ۱ تا ۵ با میانگین امتیاز ۳،۵۸ و انحراف استاندارد ۱،۱۷ ارائه کردند.

هر کاربر در مجموعه داده‌های Movielens حداقل ۲۰ رتبه‌بندی ارسال کرده است. همچنین هر کاربر حداکثر ۲۳۱۴ امتیاز ارائه داده است. همچنین به طور متوسط، هر کاربر ۱۶۶ رتبه‌بندی با انحراف معیار ۱۹۳ ارائه داده است.

هر مورد از مجموعه‌داده‌های Movielens حداقل ۱ بار و حداکثر ۳۴۲۸ بار رتبه‌بندی می‌شود. به‌طور متوسط، هر مورد ۲۷۰ بار با انحراف معیار ۳۸۴ رتبه‌بندی می‌شود. این مجموعه‌داده از ۱،۰۰۰،۲۰۹ رتبه‌بندی ناشناس بر اساس تعداد فیلم‌ها و کاربران تشکیل شده است. پراکنـــدگی مجموعـــه‌داده × $\left[1-\frac{1000209}{3706\times6040}\right]$ (95.53% 100) است.

برای این پژوهش، محتوای آیتم‌ها، به‌عنوان مثال فیلم‌ها، از پایگاه داده وب سایت IMDb جمع‌آوری شده است. برای این منظور، از یک نوع سرویس Web Crawler، به‌اسم WebSPHINX، استفاده می‌شود. (http://cs.cmu.edu/~rcm/websphinx) توسط راب میلر از دانشگاه کارنگی ملون ایجاد شده است. از این داده‌ها برای ایجاد و تکمیل هستان‌شناسی استفاده می‌شود.

در ارزیابی هر RS، مجموعه‌داده باید به دو قسمت تقسیم شود: یک مجموعه آموزش و یک مجموعه آزمایش. در حین ساخت RS، فقط باید از مجموعه آموزش استفاده شود. پس از ساخت RS، می‌توان از مجموعه آزمایش برای ارزیابی میزان عملکرد RS استفاده کرد. زیر مجموعه‌ای از اندازه $\frac{4}{5}$ داده‌ها به‌طور تصادفی برای مجموعه آموزش انتخاب می‌شود و مجموعه آزمایش از $\frac{1}{5}$ باقیمانده آن‌ها تشکیل می‌شود.

۴-۲- ارزیابی کیفیت خوشه‌بندی

برای ارزیابی کیفیت خوشه‌بندی، می‌توان از روش Silhouette استفاده کرد. در این راستا، پس از خوشه‌بندی مجموعه‌داده‌ها به k خوشه جداگانه، هر یک از خوشه‌ها با استفاده از روش مذکور [77 , 78] ارزیابی می‌شود. از آن‌جایی‌که مقدار k متغیر است و بر اساس مقادیر مختلف برای k، خوشه‌های مختلف تولید خواهد شد، به همین دلیل نیاز است تا کیفیت خوشه‌ها مورد آزمایش قرار گیرد. با توجه به پژوهش انجام‌شده توسط نیلاشی و همکاران با استفاده از روش Silhouette داریم [78]:

$$s_i = \frac{(b_i - a_i)}{Max\{a_i, b_i\}}$$

$$a_i = \frac{\sum_{i' \in C_i \cdot i \neq i'} dist(i.i')}{|C_i| - 1} \qquad (۱۳)$$

$$b_i = \min_{C_j : 1 \le j \le K \cdot j \neq i} \left\{ \frac{\sum_{i' \in C_j} dist(i.i')}{|C_j|} \right\}$$

a_i میانگین عدم شباهت i و سایر داده‌های متعلق به خوشه دربرگیرنده i و b_i مینیمم میانگین عدم شباهت i با هریک از خوشه‌هایی که به آن تعلق ندارد، است. a_i کوچک به معنی انطباق خوب داده i و خوشه‌اش و b_i بزرگ به معنی انطباق بد داده i با خوشه مجاورش است. میانگین s_i چگونگی گروه‌بندی داده‌ها در خوشه‌ها را نشان می‌دهد. روابط فوق به‌صورت زیر نیز می‌تواند نوشته شود:

$$s_i = \begin{cases} 1 - a_i / b_i & , \text{ if } a_i < b_i \\ 0 & , \text{ if } a_i = b_i \\ b_i / a_i - 1 & , \text{ if } a_i > b_i \end{cases} \qquad (۱۴)$$

ضریب Silhouette می‌تواند مقادیری بین ۱- و ۱ داشته باشد. هر چه مقدار s_i به ۱ نزدیک‌تر باشد نشان‌دهنده خوشه‌بندی خوب و هرچقدر این مقدار به ۱- نزدیک باشد به معنی خوشه‌بندی ضعیف است.

شکل (۱۴) مقادیر متوسط Silhouette را برای پارتیشن‌های حاصل از آنها ارائه می‌دهد. مقدار Silhouette برای یک پارتیشن با ۶ خوشه، کمی بالاتر از ۰/۸ است. همان‌طور که از شکل (۱۴) استنباط می‌شود، با افزایش تعداد خوشه از ۵، روند نمودار Silhouette به‌طورکامل صعودی است. اما این روند از ۲۳ تا ۲۷ در اوج خواهد بود. مقدار اوج آن برای ۲۴ خوشه برابر با ۰/۹ است. برای خوشه بالای ۲۷، نمودار Silhouette در حال نوسان، اما روند آن نزولی است. بهترین مقدار را می‌توان ۲۴ یا ۲۶ در نظر گرفت.

۴-۳- ارزیابی سامانه پیشنهادی

به‌منظور ارزیابی روش پیشنهادی می‌بایست الگوریتم مذکور را در مقایسه با روش‌های حافظه پایه (به‌طور مثال *KNN* پایه) و مدل پایه (روش‌های خوشه‌بندی پایه) از دو منظرِ پیچیدگی زمانی و دقت، مورد بررسی قرار دهیم.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

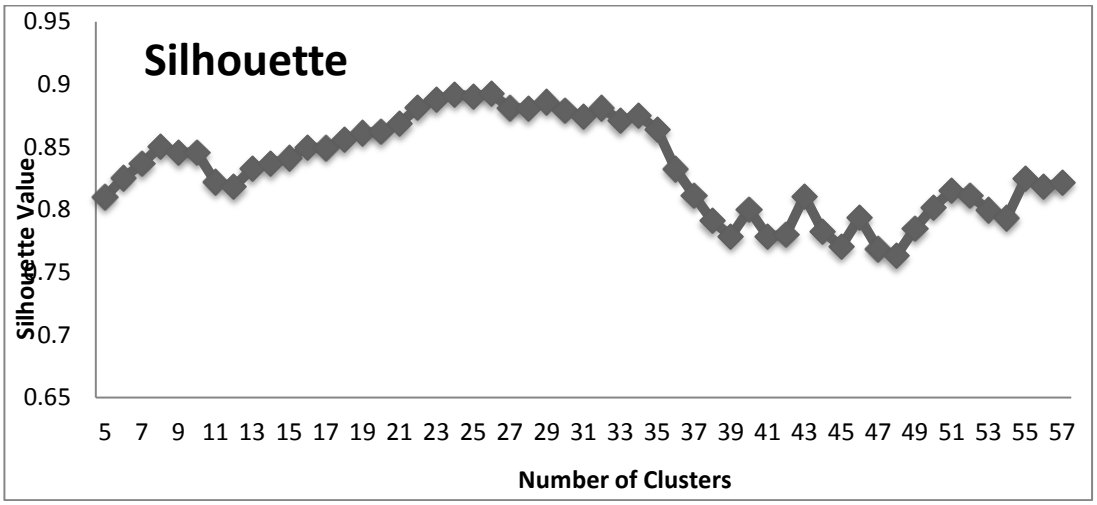
[DOI: 10.52547/jsdp.18.4.89]

روش *KNN* در CF حافظه پایه، استفاده می‌شود و روش خوشه‌بندی در CF مدل پایه استفاده شده است. لازم به ذکر است که سامانه‌های حافظه پایه، دارای دقت بالا و سامانه‌های مدل پایه، پیچیدگی زمانی قابل توجهی دارند. هدف اصلی الگوریتم پیشنهادی در این پژوهش این است که یک سامانه پیشنهادگر ترکیبی با دقتی همانند توصیه‌گرهای حافظه پایه و با مقیاس‌پذیری همانند توصیه‌گرهای مدل پایه ارائه دهد.

روش *KNN* در ColF مبتنی بر حافظه و روش خوشه‌بندی در ColF مبتنی بر مدل استفاده شده است.

قابل توجه است که RS مبتنی بر حافظه از دقت بهینه برخوردار است در حالی‌که سامانه‌های مبتنی بر مدل از نظر پیچیدگی زمانی به‌طور قابل توجهی بهتر هستند. هدف از الگوریتم پیشنهادی RS تهیه یک HRS متشکل از دقت RS مبتنی بر حافظه و مقیاس‌پذیری RS مبتنی بر مدل است.

از روش‌های مختلف دیگری برای مقایسه با روش پیشنهادی استفاده شده است. این روش‌ها عبارتند از: ConF RS غیرنرمال [82]، RS مبتنی بر تجزیه ارزش منفرد (با SVD نشان داده شده است) [83]، RS مبتنی بر محبوبیت (با Pop مشخص می‌شود) [84]، Top-N RS مبتنی بر هستان‌شناسی با استفاده از فاکتورسازی ماتریس (با OTopN نشان داده شده است [89].



(شکل-۱۴): پارتیشن‌های مقادیر Silhouette به‌طور متوسط با مقادیر مختلف k برای پارامتر تعداد خوشه‌ها در الگوریتم خوشه‌بندی Kmeans به‌دست آمده است.

(Figure-14): The average Silhouette values partitions obtained by different k values for the parameter of clusters' number in the Kmeans clustering algorithm

به‌منظور ارزیابی سامانه پیشنهادگر پیشنهادی از منظر پیچیدگی زمانی، می‌توان پیچیدگی زمانی الگوریتم پیشنهادی را با پیچیدگی زمانی محاسبه شده برای الگوریتم‌های حافظه پایه (هم‌چون *KNN*) و مدل پایه (روش‌های خوشه‌بندی) مقایسه کرد. در این میان، پارامترهایی همچون: تعداد کل کاربران، تعداد خوشه‌ها، تعداد اقلامی که مشابه پروفایل کاربر هدف هستند و تعداد کاربرانی که به اقلام مشابه پروفایل کاربر هدف امتیاز مشابهی داده‌اند، حائز اهمیت می باشند.

تعداد کل کاربران، تعداد خوشه‌ها، تعداد اقلامی که مشابه پروفایل کاربر هدف هستند و تعداد کاربرانی که به اقلام مشابه پروفایل کاربر هدف امتیاز مشابهی داده‌اند به‌ترتیب عبارتند از: m، N_1، N_2 و $|G|$. پیچیدگی زمانی الگوریتم پیشنهادی برابر است با $O(N_2 \times |G|)$. این در مقایسه با روش مرسوم ColF مبتنی بر *KNN*، که $O(m)$ است، افزایش یافته و نزدیک به پیچیدگی زمانی مربوط به سامانه خوشه‌بندی مبتنی بر مدل است، که $O(N_1)$ گزارش شده است.

به‌منظور شناسایی پیچیدگی زمانی الگوریتم، شکل (۱۵) نمودار زمان (در ثانیه) را نشان می‌دهد. این الگوریتم‌ها تحت شرایط یکسان مقایسه می‌شوند و در یک رایانه اجرا می‌شوند. شکل (۱۵) نشان می‌دهد که زمان محاسبه روش پیشنهادی کمتر از روش *KNN* است. درواقع، شکل (۱۵) نشان می‌دهد که زمان اجرای روش پیشنهادی به‌طور قابل توجهی کمتر از زمان اجرای روش *KNN* است اما از زمان اجرای روش مبتنی بر خوشه-بندی بدتر است. زمان محاسبه روش *KNN* در مقایسه با خوشه‌بندی و روش پیشنهادی به‌طور قابل توجهی زیاد است. علاوه‌بر این، روشی که فقط از خوشه‌بندی استفاده می‌کند، در مقایسه با سایر روش‌های استاندارد، زمان محاسبه بهینه را به‌دست آورده است. همچنین، زمان محاسبه روش ما نسبتاً نزدیک به روش خوشه‌بندی است.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]
[DOI: 10.52547/jsdp.18.4.89]

همان‌طور که در شکل (۱۵) مشاهده می‌شود، روش KNN به‌طور متوسط بیش از ۲ (حدود ۲/۴۳) ثانیه زمان می‌برد. اما روش پیشنهادی به‌طور متوسط کمتر از ۱ (حدود ۰/۴) ثانیه برای اجرا مصرف می‌کند. روش مبتنی بر k–میانگین به‌طور متوسط کمتر از ۱ (حدود ۰/۱۱) ثانیه برای اجرا مصرف می‌کند. اما تعداد همسایگان کمترین تأثیر را در زمان اجرای روش‌های مختلف دارد.

به‌وسیله معیارهای آماری هم‌چون میانگین خطای مطلق (MAE)، می‌توان تفاوت میان پیش‌بینی‌های انجام شده توسط الگوریتم پیشنهادی و رتبه‌بندی‌های واقعی را اندازه‌گیری نمود. میانگین خطای مطلق از طریق رابطه زیر محاسبه می‌شود.

$$MAE = \frac{\sum_i \sum_j |R_{ij} - \hat{R}_{ij}|}{TestSize} \quad (۱۵)$$

به این‌صورت که $TestSize$ مقدار آیتم‌هایی است که کاربران آزمایش رتبه‌بندی کرده‌اند و R_{ij} امتیاز پیش‌بینی‌شده‌ای است که کاربر i به آیتم j ارائه می‌دهد.

معیارهای Decision – Support نقش مهمی در ارزیابی سامانه‌های پیشنهادگر خواهند داشت. بسیاری از معیارهای مربوط به این حوزه، پارامترهایی هستند که در ارزیابی الگوریتم‌های بازیابی اطلاعات بسیار شناخته شده و حائز اهمیت می‌باشند. از جمله این معیارها می توان به دقت و فراخوانی اشاره کرد.

معیار دقت بر اساس نسبت تعداد اقلام مرتبط انتخاب شده به تعداد اقلام انتخاب شده طبق رابطه زیر تعریف می‌شود:

$$Pre = \frac{TPR}{TPR + FPR} \quad (۱۶)$$

در حقیقت، معیار دقت، بیانگر احتمال اینکه قلم انتخاب شده، مرتبط باشد است. معیار فراخوانی نیز همان‌طور که در رابطه زیر نشان داده شده است، برابر است با نسبت تعداد داده‌های مرتبط انتخاب شده به کل داده‌های مرتبط موجود:

$$Rec = \frac{TPR}{TPR + FNR} \quad (۱۷)$$

معیار فراخوانی، احتمال این‌که یک قلم مرتبط انتخاب شود، را نشان می‌دهد.

به این ترتیب‌که FNR مقدار پیش‌بینی‌های اشتباه غیرمرتبط است، TPR مقدار پیش‌بینی‌های درست مرتبط و FPR مقدار پیش‌بینی‌های اشتباه مرتبط است.

لازم به‌ذکر است باتوجه به این‌که با افزایش تعداد اقلام بازیابی شده، فراخوانی افزایش یافته و دقت کاهش می‌یابد، لذا می‌بایست از معیاری استفاده شود که قادر باشد هر دو معیار دقت و فراخوانی را به‌صورت مشترک با هم در نظر بگیرد. برای این منظور می‌توان از معیار F که درواقع یک میانگین هارمونیک از دقت و فراخوانی است استفاده کنیم [80] (رابطه ۱۸).

پارامتر Φ ممکن است برای وزن‌دادن به تأثیر یک یا هر دو مورد استفاده قرار گیرد، به طوری که $\Phi > 1$ اهمیت دقت را افزایش می‌دهد و $\Phi < 1$، اثر فراخوان را افزایش می‌دهد. برای دست‌یابی به یک F – Measure متعادل، $\Phi = 1$ فرض می‌شود.

$$F_{\Phi} = \frac{(1 + \Phi^2) \times Pre \times Rec}{\Phi^2 \times (Pre + Rec)} \quad (۱۸)$$

علاوه‌براین، برای ارزیابی روش پیشنهادی از طریق معیارهای Decision – Support، فراخوان و دقت با استفاده از انواع مختلف:

Top-N 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20

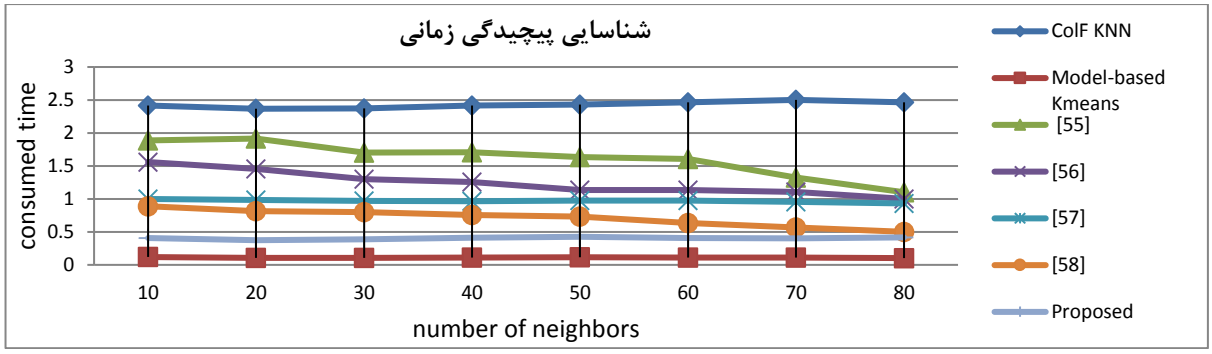
حاصل می‌شود.

مقادیر فراخوانی و دقت برای مقادیر مختلف Top-N برای محاسبه متریک F استفاده می‌شود و همچنین متریک F در شکل (۱۶-ب) نشان داده شده است. روش پیشنهادی همچنین با استفاده از MAE ارزیابی می‌شود تا عملکرد آن تعیین شود و با RSهای نوع مبتنی بر حافظه (مانند KNN) و از نوع مبتنی بر مدل، مانند رویکردهای خوشه‌بندی مقایسه می‌شود. در جدول (۳)، MAE برای اندازه‌های مختلف همسایه k در مجموعه داده‌های معیار ارائه شده است.

شکل (۱۶) روش پیشنهادی را باسایر روش‌های مشابه با دو معیار فراخوانی و معیار F1 مورد مقایسه قرار داده است. نمودار(۱۶-الف) بیانگر معیار فراخوانی و نمودار (۱۶-ب) معیار رتبه F1 را نشان می‌دهند. نتایج حاصله نشان‌دهنده برتری روش پیشنهادی نسبت به سایر روش‌های مشابه است.

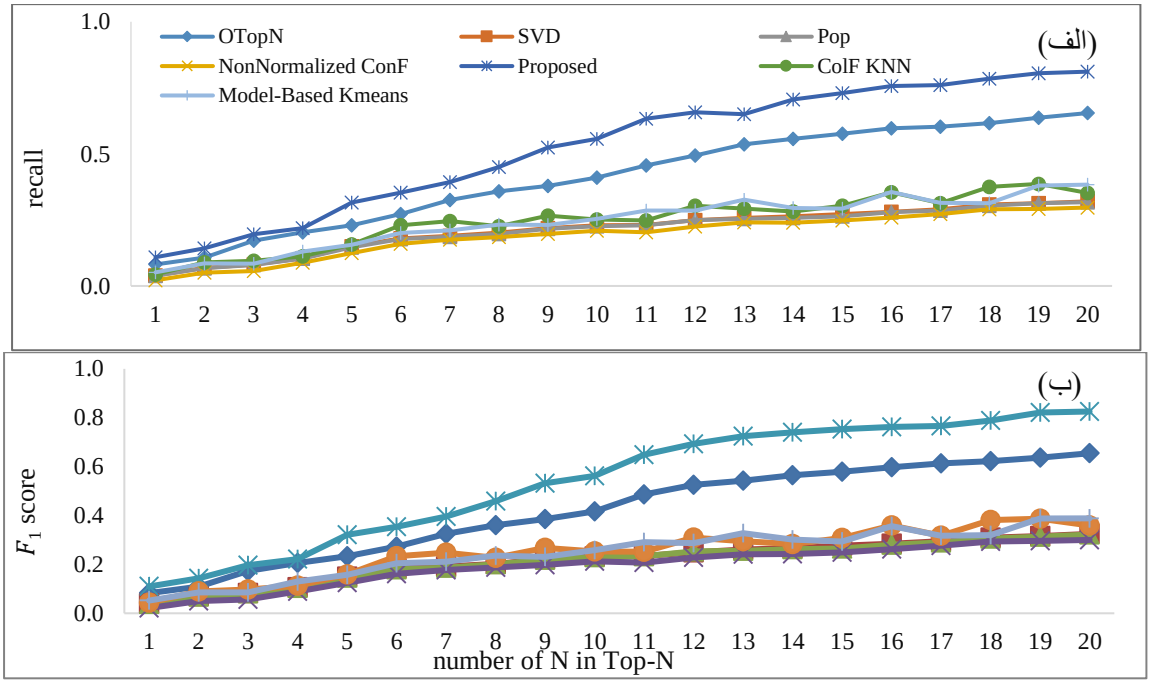
[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]



(شکل-۱۵): محور عمودی به معنای زمان مصرف‌شده و محور افقی به معنای تعداد همسایگان است

(Figure-15): The vertical axis stands for the consumed time in terms of second and the horizontal axis stands for the number of neighbors



(شکل-۱۶): (الف) محور عمودی به معنای فراخوان و محور افقی به معنای تعداد N در Top-N است. (ب) محور عمودی به معنای رتبه F_1 و محور افقی به معنای تعداد N در Top-N است.

(Figure-16): (a) The vertical axis stands for recall and the horizontal axis stands for the number of N in Top-N. (b) The vertical axis stands for F_1 score and the horizontal axis stands for the number of N in Top-N.

(جدول-۳): مقایسه روش‌های مختلف از نظر معیار MAE

(Table-3): Comparison of different methods in terms of MAE criteria

	Number of *KNN*						
	5	**10**	**15**	**20**	**50**	**60**	**80**
Method							
LMR	1.1075541	1.1512821	1.0064977	1.0599112	1.0462548	1.0075896	0.9201686
GMR	1.0075541	1.0612821	0.9787374	1.0742112	0.9049137	0.9004336	0.8792492
LMR+	0.9411229	0.9120105	0.9089575	0.8760508	0.8639927	0.8199077	0.8004988
LULCS	1.0062576	0.9747877	0.9578965	0.912306	0.8620667	0.8525098	0.824147
User- and Item-based + SVD + EM + Ontology [85]	0.93532	0.88901	0.81480	0.80042	0.75672	0.73342	0.63422
[55]	0.8102836	0.8068856	0.7905467	0.790009	0.7783540	0.7506440	0.7503783
[56]	0.7802768	0.7769650	0.7690481	0.766301	0.7085576	0.7006634	0.6800476
[57]	0.6172945	0.5517851	0.5426238	0.5369432	0.52845077	0.5176570	0.4949458
Proposed	**0.5502844**	**0.5569947**	**0.5005291**	**0.558579**	**0.4583577**	**0.4006688**	**0.4000473**

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

روش پیشنهادی علاوه‌بر روش‌های به نسبت مشابه قبلی با ۴ روش نسبتا جدید دیگر از نظر معیار MAE نیز مورد مقایسه قرار گرفته است (در جدول (۳) روش پیشنهادی با این ۴ روش مقایسه شد). روش اول در مرجع [85] آورده شده است. روش [85] یک روش توصیه ترکیبی مبتنی بر رویکردهای پالایش مشارکتی (CF) است. بر این اساس، در این تحقیق با استفاده از روش‌های کاهش ابعاد و هستان‌شناسی، سعی شده تا دو عیب اصلی سامانه‌های پیشنهادگر شامل پراکندگی داده‌ها و مقیاس-پذیری حل شود. سپس، از هستان‌شناسی برای بهبود دقت توصیه‌ها در بخش CF استفاده شده است. در بخش CF، همچنین از یک روش کاهش ابعاد، تجزیه ارزش منفرد (SVD) برای یافتن مشابه‌ترین اقلام و کاربران در هر خوشه از اقلام و کاربران استفاده شده که می‌تواند قابلیت‌های روش توصیه را بهبود بخشد. در این روش، فقط از خوشه‌بندی EM و SVD غیرافزایشی برای کاهش ابعاد استفاده شده است. این در حالی است که به‌کارگیری SVD افزایشی می‌تواند توصیه‌هایی با مقیاس‌پذیری مناسب‌تری ارائه دهد. در روش پیشنهادی ما از خوشه‌بندی بهبودیافته‌ای استفاده کرده‌ایم که چالش هم‌پوشانی در خوشه‌بندی را برطرف کرده و دیگر مشکلات خوشه‌بندی سنتی را ندارد. خوشه‌بندی، هستان‌شناسی پروفایل کاربر، و الگوریتم پیشنهادی حافظه پایه (الگوریتم بهبود یافته KNN) در بخش پالایش مشارکتی استفاده می‌شوند. ۳ روش دیگر قبلا نیز در بخش مرور منابع آورده شده‌اند، شامل مراجع [55]، [56] و [57] هستند.

نتایج جدول (۳) حاکی از برتری دقت روش پیشنهادی نسبت به ۴ روش رقیب می‌باشد. به‌طور کلی، می‌توان نتیجه گرفت که این روش ترکیبی به‌طور قابل توجهی زمان محاسبه و همچنین دقت و پیش‌بینی توصیه را بهبود می‌بخشد، همچنین در موضوع مقیاس‌پذیری و مسأله پراکندگی داده‌ها در مقایسه با KNN استاندارد و روش‌های مبتنی بر مدل، مؤثر است. با توجه به سامانه‌های پیشنهادی، مصالحه بین دقت و زمان محاسبه حیاتی است، از این‌رو روش پیشنهادی یک سامانه هوشمند دلگرم‌کننده و تأثیرگذار برای توصیه فیلم‌ها است.

از جمله دیگر بهبودهایی که در حوزه سامانه‌های پیشنهادگر اخیراً توسط پژوهش‌گران انجام گرفته، می‌توان به روش‌های یادگیری عمیق اشاره کرد. از جمله این روش‌ها می‌توان به [86-88] مراجعه کرد.

۵- نتیجه‌گیری

در این پژوهش، سامانه پیشنهادگر مبتنی بر پالایش مشارکتی (CF) و پالایش محتوا پایه (CBF) پیشنهاد شده است. به منظور غلبه بر مسأله شروع سرد (قلم جدید) ایده استفاده از روش هستان‌شناسی در CBF مطرح شده؛ همچنین جهت رفع مشکلاتی همچون پراکندگی داده‌ها و مقیاس‌پذیری در CF، از روش‌های خوشه‌بندی استفاده شده است.

در روش پیشنهادی، روش حافظه پایه و مدل پایه با استفاده از هستان‌شناسی ترکیب شده است. با توجه به راه‌کارهای ارائه‌شده و نتایج به‌دست‌آمده، می‌توان گفت روش ترکیبی پیشنهادی، در مقایسه با رویکردهای استاندارد حافظه پایه و مدل پایه، دارای مقیاس‌پذیری و دقت خوبی است و بهبود قابل توجهی در مورد زمان محاسبات، دقت پیش‌بینی رتبه‌بندی و تطابق پیشنهادات ارائه‌شده به کاربر فعال با سلایق وی، ایجاد کرده است. بهبود پارامترهای مذکور نشان‌دهنده اثربخشی روش پیشنهادی در کاهش پراکندگی داده‌ها و افزایش مقیاس‌پذیری، در مقایسه با روش‌های مبتنی بر مدل و حافظه پایه است.

علاوه‌بر راه‌کارهای ارائه‌شده در بالا، به منظور بهبود دقت پیشنهادات (در مقایسه با روش مدل پایه) از روش‌های خوشه‌بندی استفاده شده است. همچنین به‌منظور ارائه میزان مقیاس‌پذیری قابل قبول در روش پیشنهادی (در مقایسه با روش‌های حافظه پایه) از نوعی الگوریتم حافظه پایه (KNN بهبود یافته) استفاده می‌شود.

به‌طور خلاصه می‌توان گفت در روش ترکیبی پیشنهادی، از هستان‌شناسی و خوشه‌بندی پیشرفته برای تولید نتایج دقیق‌تر استفاده شده است. به‌منظور تجزیه وتحلیل روش پیشنهادی و کاربردی‌بودن آن، دقت پیش-بینی انجام شده و میزان پیچیدگی زمانی (مقیاس‌پذیری) روش پیشنهادی در یک پایگاه داده واقعی در زمینه سامانه پیشنهادگر فیلم (ارائه‌شده توسط Movielens) مورد ارزیابی قرار گرفت. همچنین به منظور بررسی عملکرد روش ارائه‌شده در این پژوهش (با استفاده از معیارهای دقت، فراخوانی، MAE و F1) الگوریتم پیشنهادی با الگوریتم‌های پژوهش‌های گذشته مقایسه شد. با توجه به راه‌کارهای ارائه‌شده مشخص شد روش‌های ترکیبی مبتنی بر پالایش مشارکتی (حافظه پایه و مدل پایه) به همراه روش‌های محتوا پایه (بر اساس هستان‌شناسی)، دقت

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

پیش‌بینی و مقیاس‌پذیری سامانه پیشنهادگر را بهبود می‌دهند؛ به‌منظور اثبات این قضیه با استفاده از اندازه گیری سه معیار: MAE، دقت تصمیم‌گیری (اندازه‌گیری پارامتر F1) و مقیاس‌پذیری (پیچیدگی زمانی) نشان داده شد روش پیشنهادی در مقایسه با الگوریتم‌های پیشین دارای عملکرد بهتری (دقت پیشنهادات ارائه شده به کاربر فعال) هستند.

در آخر ذکر یک نکته الزامی است، همان‌طور که در اکثر سامانه‌های پیشنهادگر، رابطه متقابل بین پارامترهای زمان محاسبات (بهبود مقیاس‌پذیری) و دقت (کاهش پراکندگی داده‌ها) از اهمیت خاصی برخوردار است، ولی یک شرط حیاتی برای آن وجود دارد. با توجه به آنکه اغلب سامانه‌های پیشنهادگر به‌عنوان بخشی از یک سامانه جامع‌تر به شکل برخط[1] مورد استفاده قرار می‌گیرند و با درنظرگرفتن این اصل که اساساً سامانه پیشنهادگر برای بهبود تجربه کاربر در استفاده از کل سامانه به‌کار می‌رود، فرایند ارائه پیشنهادات بایستی در زمان منطقی انجام شود. بنابراین با توجه به مطالب بالا، در این پژوهش نیز سعی شده است در کنار ارائه راه‌کارهایی جهت بهبود عملکرد سامانه پیشنهادگر، شرط حیاتی بالا که همانا بهبود تجربه کاربر در استفاده از کل سامانه است را مد نظر قرار داده و هدف اصلی و کاربردی روش پیشنهادی خود را ارائه سامانه پیشنهادگری که بتواند به‌عنوان یک سامانه هوشمند مؤثر برای توصیه اقلام (فیلم‌ها) عمل کند، قرار دهیم.

در این پژوهش، صرفاً از روش خوشه‌بندی k-means، برای خوشه‌بندی در سامانه پیشنهادگر استفاده شده است. علاوه بر این، روش پیشنهادی در حوزه توصیه فیلم مورد ارزیابی قرار گرفت. از این‌رو در کارهای آتی می‌توان، سایر روش‌های خوشه‌بندی به ویژه روش‌های خوشه‌بندی جمعی را به‌منظور ارتقاء سامانه پیشنهادگر، مورد ارزیابی قرار داد. همچنین می‌توان برای ارزیابی سامانه پیشنهادگر، انواع دیگری از مجموعه‌داده‌ها در حوزه گردشگری و خرید، مورد استفاده قرار گیرد.

6- References ۶- مراجع

[1] E. Rich, "User modeling via stereotypes," *Cognitive science*, vol. 3, pp. 329−354 , 1979.

[2] G. Salton, "Automatic text processing: The transformation, analysis, and retrieval of," *Reading: Addison-Wesley*, 1989.

[3] B. Murthi and S. Sarkar, "The role of the management sciences in research on personalization," *Management Science*, vol. 49, pp. 1344-1362, 2003.

[4] M. J. D. Powell, *Approximation theory and methods*: Cambridge university press, 1981.

[5] G. Lilien, P. Kotler, and K. Moorthy, "Marketing models prentice-hall," *Englewood Cliffs, NJ*, 1992.

[6] J. S. Armstrong, *Principles of forecasting: A handbook for researchers and practitioners* vol. 30: Springer Science & Business Media, 2001.

[7] F. McSherry and I. Mironov, "Differentially private recommender systems: Building privacy into the netflix prize contenders," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 627-636.

[8] S. S. Anand and B. Mobasher, "Intelligent techniques for web personalization," in *Proceedings of the 2003 international conference on Intelligent Techniques for Web Personalization*, 2003, pp. 1-36.

[9] P. Lops, M. De Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender systems handbook*, ed: Springer, 2011, pp. 73-105.

[10] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, 1998, pp. 43-52.

[11] L. Iaquinta, M. De Gemmis, P. Lops, G. Semeraro, M. Filannino, and P. Molino, "Introducing serendipity in a content-based recommender system," in *Hybrid Intelligent Systems, 2008 . HIS'08. Eighth International Conference on*, 2008, pp. 168-173.

[12] B. M. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering," in *Proceedings of the fifth international conference on computer and information technology*, 2002, pp. 291-324.

[13] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: A constant time collaborative filtering algorithm," *information retrieval*, vol. 4, pp. 133-151, 2001.

[14] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: A constant time collaborative filtering algorithm," *information retrieval*, vol. 4, pp. 133-151, 2001.

[15] M. Nilashi, O. bin Ibrahim, and N. Ithnin, "Hybrid recommendation approaches for multi-criteria collaborative filtering," *Expert Systems with Applications*, vol. 41, pp. 3879-3900, 2014.

[1] Online

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

[28] R. Burke, "Hybrid recommender systems: Survey and experiments," *User modeling and user-adapted interaction*, vol. 12, pp. 331-370, 2002.

[29] J. C, Flinn, & G. L. Denning, “Interdisciplinary challenges and opportunities in international agricultural research”, IRRI research paper series-International Rice Research Institute, 1982.

[30] Chen, S., Peng, Y, “*Matrix factorization for recommendation with explicit and implicit feedback. Knowl*”, Based Syst. 158: 109-117, 2018.

[31] G. Adomavicius, & Y. Kwon, “*New recommendation techniques for multicriteria rating systems*”, IEEE Intelligent Systems, 22(3), 2007.

[32] G. Adomavicius, & A. Tuzhilin, “*Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions*”, Knowledge and Data Engineering, IEEE Transactions on, 17(6), 734-749, 2005.

[33] D. H. Park, H. K. Kim, I. Y. Choi, & J. K. Kim, “*A literature review and classification of recommender systems research*”, *Expert Systems with Applications*, *39*(11), 10059-10072, 2012.

[34] M. Montaner, B. López, and J. L. De La Rosa, "*A taxonomy of recommender agents on the internet*", *Artificial intelligence review*, vol. 19, pp. 285-330, 2003.

[35] J.-S. Lee and S. Olafsson, "Two-way cooperative prediction for collaborative filtering recommendations," *Expert Systems with Applications*, vol. 36, pp. 5353-5361, 2009.

[36] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The adaptive web*, ed: Springer, 2007, pp. 291-324.

[37] R. Van Meteren and M. Van Someren, "Using content-based filtering for recommendation," in *Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop*, 2000, pp. 47-56.

[38] K. Verbert, N. Manouselis, X. Ochoa, M.Wolpers, H. Drachsler, I. Bosnic & Duval, E, “*Context-aware recommender systems for learning: a survey and future challenges*”, IEEE Transactions on Learning Technologies, vol. *5*(4), pp.318-335, 2012.

[39] Y. Koren, R. Bell & C. Volinsky, “*Matrix factorization techniques for recommender systems*”, Computer, vol. 42(8), pp. 30-37, 2009.

[16] G. Linden, B. Smith, and J. York, "Amazon. Com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, vol. 7, pp. 76-80, 2003.

[17] M. C. Pham, Y. Cao, R. Klamma, and M. Jarke, "A clustering approach for collaborative filtering recommendation using social network analysis," *J. UCS*, vol. 17, pp. 583-604, 2011.

[18] S. Gong, "A collaborative filtering recommendation algorithm based on user clustering and item clustering," *JSW*, vol. 5, pp. 745-752, 2010.

[19] Y. He, S .Yang, and C. Jiao, "A hybrid collaborative filtering recommendation algorithm for solving the data sparsity," in *Computer Science and Society (ISCCS), 2011 International Symposium on*, 2011, pp. 118-121.

[20] S. K. Shinde and U. Kulkarni, "Hybrid personalized recommender system using centering-bunching based clustering algorithm," *Expert Systems with Applications*, vol. 39, pp. 1381-1387, 2012.

[21] K. Truong, F. Ishikawa, and S. Honiden, "Improving accuracy of recommender system by item clustering," *IEICE TRANSACTIONS on Information and Systems*, vol. 90, pp. 1363-1373, 2007.

[22] P. Wang, "A personalized collaborative recommendation approach based on clustering of customers," *Physics Procedia*, vol. 24, pp. 812-816, 2012.

[23] Z. K. Zhang, T. Zhou, and Y.-C. Zhang, "Tag-aware recommender systems: A state-of-the-art survey," *Journal of computer science and technology*, vol. 26, p. 767, 2011.

[24] M. C. Pham, Y. Cao, R. Klamma, and M. Jarke, "A clustering approach for collaborative filtering recommendation using social network analysis," *J. UCS*, vol. 17, pp. 583-604, 2011.

[25] J. Breese, D. Heckerman, and C. Kadie,. “An experimental comparison of collaborative filtering methods”, Technical Report MSR-TR 98-12, Microsoft Research, Redmond, WA, 1998.

[26] Y. He, S .Yang, and C. Jiao, "A hybrid collaborative filtering recommendation algorithm for solving the data sparsity," in *Computer Science and Society (ISCCS), 2011 International Symposium on*, 2011, pp. 118-121.

[27] S. K. Shinde and U. Kulkarni, "Hybrid personalized recommender system using centering-bunching based clustering algorithm," *Expert Systems with Applications*, vol. 39, pp. 1381-1387, 2012.

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

[40] Y. Koren, "*Factorization meets the neighborhood: a multifaceted collaborative filtering model*", KDD 2008, 426-434, 2008.

[41] L. C. Cheng, , & H. A. Wang, "*A fuzzy recommender system based on the integration of subjective preferences and objective information*", Applied Soft Computing, vol. 18, .290-301, 2014.

[42] S. Chen, Peng, Y. "*Matrix factorization for recommendation with explicit and implicit feedback. Knowl*", Based Syst. Vol. 158, pp. 109-117, 2018.

[43] M. Y. H. Al-Shamri, "*User profiling approaches for demographic recommender systems. Knowl*", Based Syst. Vol. 100, pp. 175-187, 2016.

[44] G. Jawaheer, M. Szomszor, and P. Kostkova, "Comparison of implicit and explicit feedback from an online music recommendation service," in *proceedings of the 1st international workshop on information heterogeneity and fusion in recommender systems*, 2010 ,pp. 47-51.

[45] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, vol. 2009, pp. 4, 2009.

[46] J. Bobadilla, A. Hernando, F. Ortega, and J. Bernal, "A framework for collaborative filtering recommender systems," *Expert Systems with Applications*, vol. 38, pp. 14609-14623, 2011.

[47] A. M. Acilar and A. Arslan, "A collaborative filtering method based on artificial immune network," *Expert Systems with Applications*, vol. 36, pp. 8324-8332, 2009.

[48] J. Borràs, A. Moreno, and A. Valls, "Intelligent tourism recommender systems: A survey," *Expert Systems with Applications*, vol. 41, pp. 7370-7389, 2014.

[49] S. Gong and H. Ye, "An item based collaborative filtering using bp neural networks prediction," in *Industrial and Information Systems, 2009. IIS'09. International Conference on*, 2009, pp. 146-148.

[50] A. Abdelwahab, H. Sekiya, I. Matsuba, Y. Horiuchi, S. Kuroiwa, and M. Nishida, "An efficient collaborative filtering algorithm using svd-free latent semantic indexing and particle swarm optimization," in *Natural Language Processing and Knowledge Engineering, 2009. NLP-KE 2009. International Conference on*, 2009, pp. 1-4.

[51] X. Zhou, J. He, G. Huang, Y. Zhang, "*SVD-based incremental approaches for recommender systems*", J. Comput. Syst. Sci. 81(4): 717-733, 2015.

[52] R. Baeza-Yates and B. Ribeiro-Neto, *Modern information retrieval* vol. 463: ACM press New York, 1999.

[53] C. Basu, H. Hirsh, and W. Cohen , "Recommendation as classification: Using social and content-based information in recommendation," in *Aaai/iaai*, 1998, pp. 714-720.

[54] K. Lang, "Newsweeder: Learning to filter netnews," in *Proceedings of the 12th international conference on machine learning*, 1995, pp. 331-339.

[55] K. Bagherifard, M. Rahmani, M. Nilashi, V. Rafe. "Performance improvement for recommender systems using ontology", *Telematics Informatics*, vol. 34(8), pp. 1772-1792, 2017.

[56] K. Bagherifard, M. Rahmani, M. Nilashi, V. Rafe, M. Nilashi, "*A Recommendation Method Based on Semantic Similarity and Complementarity Using Weighted Taxonomy: A Case on Construction Materials Dataset*", J. Inf. Knowl. Manag,vol. 17(1), pp. 1-26, 2018.

[57] M. Nilashi, O. Ibrahim, K. Bagherifard, "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques", Expert Syst. Appl. 92: 507-520, 2018.

[58] J. Alspector, A. Kolcz, and N. Karunanithi, "Comparing feature-based and clique-based user models for movie selection," in *Proceedings of the third ACM conference on Digital libraries* ,1998 ,pp. 11-18.

[59] N. Guarino, D. Oberle, and S. Staab" ,What is an ontology?," in *Handbook on ontologies*, ed: Springer, 2009, pp. 1-17.

[60] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing?," *International journal of human-computer studies*, vol. 43, pp. 907-928, 1995.

[61] W. Borst, "Construction of engineering," ed: Ontologies, University of Tweenty, Enschede, NL-Center for Telematica and Information Technology, 1997.

[62] A. Flahive, B. O. Apduhan, J. W. Rahayu, and D. Taniar, "Large scale ontology tailoring and simulation in the semantic grid environment," *International Journal of Metadata, Semantics and Ontologies*, vol. 1, pp. 265-281, 2006.

[63] G. Antoniou and F. Van Harmelen, *A semantic web primer*: MIT press, 2004.

[64] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure, "Ontological user profiling in recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, pp. 54-88, 2004.

[65] N. Guarino, C. Masolo, and G. Vetere , "Ontoseek: Content-based access to the web,"

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

government-to-business e-services," *Internet Research*, vol. 20, pp. 342-365, 2010.

[76] O. Daramola, M. Adigun, and C. Ayo, "Building an ontology-based framework for tourism recommendation services," *Information and communication technologies in tourism 2009*, pp. 135-147 ,2009.

[77] R. Q. Wang and F.-S. Kong, "Semantic-enhanced personalized recommender system," in *Machine Learning and Cybernetics, 2007 International Conference on*, 2007, pp. 4069-4074.

[78] S. Trewin, "Knowledge-based recommender systems," *Encyclopedia of library and information science*, vol. 69, p. 180, 2000.

[79] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," in *Proceedings of the 14th international conference on World Wide Web* ,2005 , pp. 22-32.

[80] T. N. Pham, T.-H. Vuong, T.-H. Thai, M.-V. Tran, and Q.-T. Ha, "Sentiment analysis and user similarity for social recommender system: An experimental study," in *Information science and applications (icisa) 2016*, ed: Springer, 2016 ,pp. 1147-1156.

[81] P. Buitelaar, P. Cimiano, and B. Magnini, *Ontology learning from text: Methods, evaluation and applications* vol. 123: IOS press, 2005.

[82] B. Sarwar, G. Karypis, J. Konstan, & J. Riedl, "Analysis of recommendation algorithms for e-commerce*",* Paper presented at the Proceedings of the 2nd ACM conference on Electronic commerce, 2000.

[83] P. Cremonesi, Y. Koren, and R. Turrin, *"Performance of recommender algorithms on top-n recommendation tasks,"* in Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 39-46.

[84] R. Bambini, P. Cremonesi, R. Turrin, "*A Recommender System for an IPTV Service Provider: a Real Large-Scale Production Environment*", In book: Recommender Systems Handbook (pp.299-331), DOI: 10.1007/978-0-387-85820-3_9.

[85] M. Nilashi, O. Ibrahim, K. Bagherifard, "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques", *Expert Systems with Applications*, vol. 92, pp. 507-520, 2015.

[86] Wang H, Wang N (2015) Collaborative deep learning for recommender systems. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge

IEEE Intelligent Systems and their Applications, vol. 14, pp. 70-80, 1999.

[66] M. Craven, A. McCallum, D. PiPasquo, T. Mitchell, and D. Freitag, "Learning to extract symbolic knowledge from the world wide web," Carnegie-mellon univ pittsburgh pa school of computer Science1998.

[67] D. Godoy and A. Amandi, "User profiling for web page filtering," *IEEE Internet computing*, vol. 9, pp. 56-64, 2005.

[68] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-based personalized search and browsing," *Web Intelligence and Agent Systems: An international Journal*, vol. 1, pp. 219-234, 2003.

[69] M. I. Martín-Vicente, A. Gil-Solla, M. Ramos-Cabrer, J. J. Pazos-Arias, Y. Blanco-Fernández, and M. López-Nores, "A semantic approach to improve neighborhood formation in collaborative recommender systems," *Expert Systems with Applications*, vol. 41, pp. 7776-7788, 2014.

[70] A. Sieg, B. Mobasher, and R. Burke, "Improving the effectiveness of collaborative recommendation with ontology-based user profiles," in *proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, 2010, pp. 39-46.

[71] I. Cantador, A. Bellogín, and P. Castells, "A multilayer ontology-based hybrid recommendation model," *Ai Communications*, vol. 21, pp. 203-210, 2008.

[72] Y. Deng, Z. Wu, C. Tang, H. Si, H. Xiong, and Z. Chen, "A hybrid movie recommender based on ontology and neural networks," in *Proceedings of the 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing*, 2010, pp. 846-851.

[73] L. Zhuhadar, O. Nasraoui, R. Wyatt, and E. Romero, "Multi-model ontology-based hybrid recommender system in e-learning domain," in *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on*, 2009, pp. 91-95.

[74] A. Moreno, A. Valls, D. Isern, L. Marin, and J. Borràs, "Sigtur/e-destination: Ontology-based personalized recommendation of tourism and leisure activities," *Engineering Applications of Artificial Intelligence*, vol. 26, pp. 633-651, 2013.

[75] J. Lu, Q. Shambour, Y. Xu, Q. Lin, and G. Zhang, "Bizseeker: A hybrid semantic recommendation system for personalized

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]

Discovery and Data Mining, pp 1235–1244. https://doi.org/10.1145/2783258.2783273.

[87] Zhang L, Luo T, Zhang F, Wu Y (2018) A recommendation model based on deep neural network. IEEE Access 6:9454–9463. https://doi.org/10.1109/ACCESS.2018.2789866.

[88] Zhang S, Yao L, Sun A, Tay Y (2019) Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Comput. Surv.52:5:1–5:38. https://doi.org/10.1145/3285029.

[89] H. Cui, M. Zhu, and S. Yao, "Ontology-based Top-N Recommendations on new items with matrix factorization," *Journal of Software*, vol. 9, pp. 2026-2032, 2014.

پیوست

$$IaD(\boldsymbol{C_1},\boldsymbol{C_2},\beta,\theta)=\begin{cases}\dfrac{\ddot{L}_1^{\,-1}\times\sum_{d=1}^{\theta}\pi\left(\boldsymbol{C_1}\in SChCSet\left(\boldsymbol{C_2}|RD_d^{C_2}\right)\right)}{\sum_{L=1}^{\beta}\left[\ddot{L}^{-1}\times\sum_{C'\in ChCSet(\boldsymbol{C_2},\ddot{L})}\sum_{d=1}^{\theta}\pi\left(C'\in SChCSet\left(\boldsymbol{C_2}|RD_d^{C_2}\right)\right)\right]}+\ddot{L}_1^{\,-1} & \boldsymbol{C_1}\in ChCSet(\boldsymbol{C_2},\ddot{L}_1),\ddot{L}_1\le\beta\\ 0 & otherwise\end{cases} \quad (۱)$$

پیام بحرانی دانش آموخته دوره دکترای تخصصی دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران در رشته مهندسی کامپیوتر گرایش سامانه‌های نرم افزاری است.

زمینه‌های پژوهشی مورد علاقه ایشان مباحثی نظیر سامانه‌های امنیت اطلاعات، هستان‌شناسی و سامانه‌های پیشنهادگر است.

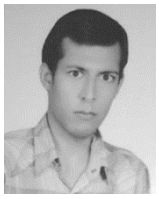
نشانی رایانامه ایشان عبارت است از:

bahranipayam@gmail.com

بهروز مینایی بیدگلی دانش آموخته دانشگاه ایالتی میشیگان آمریکا در رشته علوم و مهندسی کامپیوتر با تخصص هوش مصنوعی و داده‌کاوی است. او در حال حاضر عضو هیأت علمی و دانشیار دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت و رئیس دانشکده مهندسی کامپیوتر است. ایشان سرپرستی گروه پژوهشی فناوری‌های بازی‌های رایانه‌ای و نیز آزمایشگاه داده‌کاوی را به عهده دارد. محاسبات نرم، یادگیری ماشین، بازی‌های رایانه‌ای، داده‌کاوی، متن‌کاوی، و پردازش زبان طبیعی، زمینه‌های پژوهشی مورد علاقه ایشان است.

نشانی رایانامه ایشان عبارت است از:

b_minaei@iust.ac.ir



حمید پروین تحصیلات خود را در مقطع کارشناسی در دانشگاه چمران اهواز به پایان رساند. ایشان مدرک کارشناسی ارشد و دکترا را در دانشگاه علم و صنعت اخذ کردند و پس از آن به عضویت هیأت علمی دانشگاه آزاد اسلامی واحد نورآباد ممسنی درآمدند. وی هم‌اکنون در چندین واحد دانشگاهی در رشته کامپیوتر مشغول به تدریس است. زمینه‌های پژوهشی وی مباحثی نظیر الگوریتم‌های بهینه‌سازی، طبقه‌بندی و خوشه‌بندی داده‌ها است.

نشانی رایانامه ایشان عبارت است از:

parvin@iust.ac.ir

میترا میرزارضایی استادیار دانشکده فنی و مهندسی گروه کامپیوتر دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران هستند. زمینه‌های پژوهشی مورد علاقه ایشان، یادگیری ماشین و شناسایی الگوها است.

نشانی رایانامه ایشان عبارت است از:

mirzarezaee@srbiau.ac.ir

احمد کشاورز مدارک کارشناسی و کارشناسی ارشد خود را به‌ترتیب در سال‌های ۱۳۸۰ و ۱۳۸۳ از دانشگاه شیراز و تربیت مدرس در رشته مهندسی برق و مخابرات سیستم دریافت کرد. ایشان درجه دکترای خود را در سال ۱۳۸۷ از دانشگاه تربیت مدرس در رشته مخابرات سامانه دریافت کرده است. وی هم‌اکنون دانشیار گروه مهندسی برق دانشگاه خلیج فارس است. زمینه‌های پژوهشی مورد علاقه ایشان عبارت است از: سنجش از دور، پردازش تصاویر پزشکی، ماشین بینایی و هوش مصنوعی.

نشانی رایانامه ایشان عبارت است از:

a.keshavarz@pgu.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2024-04-20]

[DOI: 10.52547/jsdp.18.4.89]