Encyclopedia of Digital Transformation
Vol. 2; No. 3; PP. 80 - 95
Fall 2021
DOI: 10.22034/dtj.2023.351292.1074

Institute of Higher Education
Mehr Alborz

Improving the performance of collaborative filtering recommender systems using deep features extraction

Nayer Chavoshi[1], Sina Dami[2]

Abstract

The growing volume of information on the Web and the Internet has made difficult the process of deciding and selecting the information, data, or products needed for many web users. This study proposed a novel method to improve the performance of collaborative filtering recommender systems. The goal is to use deep feature extraction to provide more effective and desirable recommendations to the user. In the preprocessing step, the input data is first entered into the primary processing system and the feature values are normalized. Then, in order to perform more accurate calculations and reduce computation time, the size of the data is reduced using a deep belief network (DBN), while extracting deep features. Afterwards, using the collaborative filtering technique, the recommended items are offerd to the user. Finally, according to the system outputs in the recommendation to the user, the accuracy of the proposed items is evaluated. To evaluate the proposed method, a comparison of its performance on the real-world MovieLens dataset with basic methods has been used. Experimental results showed that the proposed method has better performance in terms of coverage and support than other compared methods.

Keywords: *Recommender system, feature engineering, deep learning, collaborative filtering*

1. M.Sc. Student, Department of Computer Engineering, West Tehran Branch, Islamic Azad University,Tehran, Iran

2. Assistant Professor, Department of Computer Engineering, West Tehran Branch, Islamic Azad University, Tehran, Iran

Submitted: 2022-07-11
Accepted: 2023-01-31
Corresponding Author: Sina Dami
Email: dami@wtiau.ac.ir

دانشنامهٔ تحول دیجیتال

مؤسسه آموزش عالی مهر البرز

دورهٔ ۲، شمارهٔ ۳

پاییز ۱۴۰۰

صص. ۸۰ - ۹۵

Encyclopedia of Digital Transformation

DOI: 10.22034/dtj.2023.351292.1074

بهبود عملکرد سیستم‌های توصیه‌گر پالایش مشارکتی با استخراج عمیق ویژگی‌ها

نیر چاوشی[۱]، سینا دامی[۲]

چکیده

حجم فراوان و روبه رشد اطلاعات بر روی وب و اینترنت، فرایند تصمیم‌گیری و انتخاب اطلاعات، داده یا کالاهای موردنیاز را برای بسیاری از کاربران وب دشوار کرده است. سیستم‌های توصیه‌گر با تحلیل رفتار کاربر خود، اقدام به پیشنهاد مناسب‌ترین اقلام داده، اطلاعات یا کالا می‌کنند. در این پژوهش، با بهره‌گیری از تکنیک‌های یادگیری عمیق، روشی برای بهبود عملکرد سیستم‌های توصیه‌گر پالایش مشارکتی ارائه شده است. هدف، استفاده از استخراج عمیق ویژگی‌ها در جهت ارائه پیشنهادهای مؤثرتر و مطلوب‌تر به کاربر سیستم موردنظر است. در بخش پیش‌پردازش، ابتدا داده‌های ورودی در سیستم پردازش اولیه قرار وارد می‌شوند و مقادیر ویژگی‌ها نرمال‌سازی می‌شوند. سپس، برای اینکه محاسبات دقیق‌تر انجام گیرد و زمان محاسبات نیز کاهش پیدا کند، با استفاده از یک شبکه باور عمیق (DBN)، ضمن استخراج عمیق ویژگی‌ها، ابعاد داده‌ها کاهش پیدا می‌کند. سپس، با استفاده از تکنیک پالایش مشارکتی، اقلام پیشنهادی به کاربر ارائه می‌شوند. در پایان، با توجه به خروجی‌های سیستم در توصیه به کاربر، ارزیابی صحت اقلام پیشنهادی صورت می‌گیرد. جهت ارزیابی روش پیشنهادی از مقایسه عملکرد آن بر روی مجموعه داده دنیای واقعی MovieLens با روش‌های پایه استفاده شده است. نتایج تجربی نشان داده است که روش پیشنهادی از نظر پوشش و حمایت نسبت به سایر روش‌های مورد مقایسه، عملکرد بهتری دارد.

کلمات کلیدی: *سیستم توصیه گر، مهندسی ویژگی‌ها، یادگیری عمیق، فیلترینگ مشارکتی.*

---

۱. استادیار گروه مهندسی کامپیوتر، واحد تهران غرب دانشگاه آزاد اسلامی، تهران، ایران

۲. دانشجو کارشناسی ارشد گروه مهندسی کامپیوتر، واحد تهران غرب دانشگاه آزاد اسلامی، تهران، ایران

تاریخ دریافت مقاله: ۱۴۰۱/۰۴/۲۰

تاریخ پذیرش نهایی مقاله: ۱۴۰۱/۱۱/۱۱

نویسندهٔ مسئول مقاله: سینا دامی

Email: dami@wtiau.ac.ir

مقدمه

اینترنت با سرعت زیادی شروع به رشد کرده و افزایش حجم اطلاعات موجود در وب‌سایت‌ها، مشکلات متعددی برای کاربران ایجاد کرده است. بسیاری از سایت‌ها دارای حجم زیادی از اقلام هستند که کاربران در بررسی و انتخاب آن‌ها دچار سردرگمی می‌شوند. اخیراً سیستم‌های توصیه‌گر، به دلیل توانایی در تصمیم‌گیری برای کاربران در انتخاب‌های مناسب، در عصر توسعه سریع فناوری اینترنت از اهمیت بیشتری برخوردار شده‌اند. به همین دلیل، ضرورت استفاده از سیستم‌های توصیه‌گر جهت فیلتر کردن و پالایش اطلاعات غیرقابل‌چشم‌پوشی است. برخی شرکت‌های بزرگ ارائه‌دهنده خدمات، از این نوع سیستم برای یافتن آسان‌تر اقلام موردعلاقه کاربرانشان استفاده می‌کنند. یک سیستم توصیه‌گر می‌تواند با فیلتر کردن و نمایش فهرست اقلامی که موردعلاقه کاربران است، از هدر رفتن زمان آن‌ها و همچنین ایجاد سردرگمی جلوگیری کند و بهره‌وری کاربران در استفاده مؤثر از این نوع سایت‌ها را افزایش دهد.

سیستم‌های توصیه‌گر[1] به‌موازات وب توسعه‌یافته‌اند. آن‌ها در ابتدا بر اساس فیلترهای جمعیت شناختی[2] مبتنی بر محتوا[3] و پالایش مشارکتی[4] کار می‌کردند (شائو، لی و بیان[5] ۲۰۲۱). امروزه، این سیستم‌ها از انواع اطلاع متناظر با تعاملات کاربران استفاده می‌کنند. پالایش مشارکتی یکی از کاربردی‌ترین روش‌ها در طراحی سیستم‌های توصیه‌گر است. در این روش، اقلام بر اساس سابقه رتبه‌بندی گذشته مانند کاربران هم‌فکر، به کاربر فعال توصیه می‌شود. با وجود این، زمانی که تعداد ویژگی‌ها افزایش پیدا می‌کند، پالایش مشارکتی با کاهش در دقت و مطلوبیت پیشنهادهی به کاربران سیستم همراه است. برای رفع این مشکل، از روش‌هایی سنتی یادگیری ماشین استفاده می‌شود که کارایی مناسبی نیز دارند. روش‌های یادگیری عمیق در پژوهش‌های اخیر کاربردهای فراوانی داشته‌اند و در بسیاری از زمینه‌ها توانسته‌اند از روش‌های سنتی یادگیری ماشین بهتر عمل کنند. در این پژوهش، ترکیبی از تکنیک‌های استخراج عمیق و پالایش مشارکتی برای طراحی سیستم توصیه‌گر ارائه‌شده است.

پیشینه تحقیق

در عصر کلان داده‌ها، سیستم‌های توصیه‌گر نقش اساسی در مدیریت داده‌ها و فیلتر کردن اطلاعات ایفا می‌کنند. در این زمینه، فیلتر مشارکتی (CF) به‌عنوان یکی از برجسته‌ترین استراتژی‌ها برای مقابله مؤثر با مجموعه داده‌های بزرگ به کار گرفته می‌شود تا محتوای مطلوبی را در قالب توصیه به کاربران، ارائه دهد. با این وجود، مشخص شده است که توصیه‌گرهای CF از پراکندگی داده‌ها، عمدتاً در سناریوهای شروع سرد رنج می‌برند. این حقیقت، کیفیت توصیه‌ها را به‌طور قابل‌توجهی کاهش می‌دهد. در ادبیات تحقیق در مورد موضوع فوق، راه‌حل‌های متعددی پیشنهاد شده است که در آن‌ها مشارکت‌های پیشرفته، به‌نوعی مشروط یا مرتبط با روش‌های سنتی CF مانند فاکتورسازی ماتریسی (MF) (اویانگ و همکاران[6] ۲۰۲۲) اعمال‌شده، یعنی برای مدل‌سازی کاربران و آیتم‌ها در قالب‌های با ابعاد کم، بر بهینه‌سازی خطی تکیه می‌شود. برای غلبه بر چالش‌های ذکر شده، تعداد فزاینده‌ای از مطالعات انجام شده است که تکنیک‌های یادگیری عمیق را در زمینه CF برای مدل‌سازی عامل پنهان بررسی می‌کنند.

با پیشرفت‌های مداوم و سریع در فناوری ارتباطات و ذخیره‌سازی داده‌ها، کاربران اینترنت با حجم زیادی از داده در بستر اینترنت مواجه می‌شوند. پدیده کلان داده چه از طریق تجارت الکترونیک (دامی و یاحقی زاده[7] ۲۰۱۸)، شبکه‌های اجتماعی

1. Recommender Systems
2. Demographic filtering
3. Content-based filtering
4. Collaborative filtering
5. Shao, Li & Bian
6. Ou-Yang et al
7. Dami & Yahaghizadeh

(فن و مصطفوی[1]، ۲۰۱۹)، یا خدمات استریم، چالش‌های قابل‌توجهی را در حفظ بازیابی مؤثر و کارآمد اطلاعات به همراه داشته است. از طرفی، این چالش‌ها فرصت‌هایی را برای ظهور راه‌حل‌های جدید در مدیریت داده‌ها و فیلتر کردن اطلاعات فراهم کرده‌اند (لیو و چونگ[2]، ۲۰۱۸).

با در نظر گرفتن سناریوی کنونی، سیستم‌های توصیه‌گر بدون شک به کاربران کمک کرده‌اند تا با پیشنهاد محتوا یا خدماتی که به‌خوبی با ترجیحات و علایق خاص کاربران مطابقت دارند، تجربه آنلاین رضایت‌بخش‌تری داشته باشند (کاماچو و همکاران[3]، ۲۰۱۸)؛ بنابراین، سیستم‌های توصیه‌گر خود را به‌عنوان تکنیک‌های مناسبی برای مدیریت اضافه بار اطلاعات معرفی کردند؛ سیستم‌هایی که محبوبیت کنونی خود را توجیه می‌کنند، به این معنی که بسیاری از خدمات آنلاین در سال‌های گذشته از آن‌ها استفاده کرده‌اند. به‌عنوان نمونه‌های شاخص، می‌توان شبکه‌های اجتماعی مانند توییتر، اینستاگرام و فیس‌بوک، اسپاتیفای، یوتیوب، نتفلیکس و آمازون پرایم را به‌عنوان سرویس‌های استریم و آمازون، نت‌شویز و گوگل ادز را به‌عنوان کانون‌هایی که خدمات تجارت الکترونیک را نمایندگی می‌کنند، ذکر کرد.

برای توصیه محتوای بالقوه به یک کاربر، یک سیستم توصیه‌گر به داده‌های کلیدی متکی است. به‌عنوان‌مثال داده‌هایی که از نمایه کاربر و اولویت‌های تاریخی استخراج‌شده‌اند، ارتباطات شبکه‌های اجتماعی و هر نوع تعامل دیگری که می‌تواند الگوی مرور کاربر را نشان دهد. به‌طورکلی، روشی که یک توصیه مبتنی بر آن انجام می‌شود، طبقه‌بندی‌های اساسی از انواع سیستم‌های توصیه‌گر را تعریف می‌کند که به‌طور سنتی به‌عنوان رویکردهای مبتنی بر محتوا (CB)، فیلتر مشارکتی (CF) و رویکردهای ترکیبی طبقه‌بندی می‌شوند. در این میان، CF به‌عنوان یکی از محبوب‌ترین و موفق‌ترین استراتژی‌ها برای توصیه معرفی می‌شود که در سال‌های گذشته توجه زیادی از بخش‌های آکادمی و صنعت را به خود جلب کرده است.

یک استراتژی CF برای توصیه می‌تواند تحت عنوان، استفاده و دست‌کاری ترجیحات تاریخچه‌ای از رفتار کاربران قبلی، برای پیش‌بینی ترجیحات ناشناخته کاربران جدید، تعریف شود. این تعریف معمولاً بر اساس این فرض بیان می‌شود که اگر دو کاربر با آیتم‌های مشابهی تعامل داشته باشند یا رفتار تعاملی مشابهی از خود نشان دهند، احتمالاً تعاملات آتی آن‌ها نیز مشابه خواهد بود.

در زمینه فوق، داده‌ها به‌صورت یک ماتریس کاربر × آیتم‌ها (به‌عنوان‌مثال، یک آرایه دوبعدی) هستند که در آن ردیف‌های ماتریس معمولاً کاربران را نشان می‌دهند و ستون‌های آن آیتم‌ها را بیان می‌کنند. مثالی از این مدل در شکل ۱ نشان داده شده است. هنگامی‌که کاربر u با یک آیتم i تعامل می‌کند، یک مقدار یا وزن قدرت آن تعامل را تعیین می‌کند، وزنی که اساساً به کاربرد دامنه (به‌عنوان‌مثال، فیلم، موسیقی، رسانه‌های اجتماعی، اخبار) و نوع توصیه‌ای که باید انجام شود (یعنی پیش‌بینی یا رتبه‌بندی) بستگی دارد. دو شکل اساسی تعامل را می‌توان صریح (یعنی رتبه‌بندی، بررسی یا هر نوع امتیازی که کاربر به‌صراحت به یک مورد داده است) یا ضمنی (به‌عنوان‌مثال، تعداد کلیک‌ها، زمان صرف شده برای تماشای یک ویدیو، گوش دادن به موسیقی یا مرور یک وب‌سایت که از اقدامات کاربر محاسبه می‌شود) دانست.

به‌طورکلی، تعامل کاربر با تمام آیتم‌های یک پایگاه داده غیرممکن است؛ بنابراین، تنها بخش کوچکی از آیتم‌ها به ازای هر کاربر رتبه‌بندی شده، یا در تعامل هستند. همچنین، با افزایش حجم داده‌ها، رویکردهای مبتنی بر CF از مقیاس‌پذیری، پراکندگی داده‌ها و مسائل شروع سرد رنج می‌برند، مسائلی که باعث کاهش عملکرد مستقیم به علایق توصیه‌ها می‌شود. پراکندگی داده‌ها و شروع سرد، مسائل مرتبطی هستند؛ زیرا هر دوی آن‌ها از کمبود داده ناشی می‌شوند. مسئله اول با تعداد کم رتبه‌بندی برای موارد موجود مشخص می‌شود که یافتن رابطه بین کاربران و موارد را دشوار می‌کند. مسئله دوم زمانی اتفاق می‌افتد که فقدان داده‌های کاربر و تاریخچه رتبه‌بندی آیتم، سیستم را از ارائه توصیه‌های قابل‌اعتماد باز می‌دارد.

1. Fan & Mostafavi
2. Liu & Cheung
3. Camacho et al

Users \ Items	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	4		3	1	5			
u_2		5	3	4		2	3	
u_3		1		5	4		3	5
u_4	5			2		4		
u_5	2					3	1	
u_6		5	4		1			4

شکل ۱: نمونه‌ای از ماتریس رتبه‌بندی

از زمان پیدایش CF، الگوریتم‌ها، روش‌ها و رویکردهای توصیه‌ای زیادی پیشنهاد شده‌اند. از تکنیک‌های سنتی تا جدیدترین روش‌های همسایگی و مدل‌های عامل پنهان مانند الگوریتم‌های مبتنی بر فاکتورسازی ماتریسی (MF)، در حال حاضر محبوب‌ترین آن‌ها هستند. در حالی که روش اولی یک روش آماری آسان برای پیاده‌سازی است که توصیه‌های منطقی و قابل تفسیر ارائه می‌کند، دومی یک رویکرد مقیاس‌پذیر است که برای دستیابی به نتایج پیش‌بینی سریع و دقیق حتی برای پایگاه‌های داده بزرگ مؤثر است.

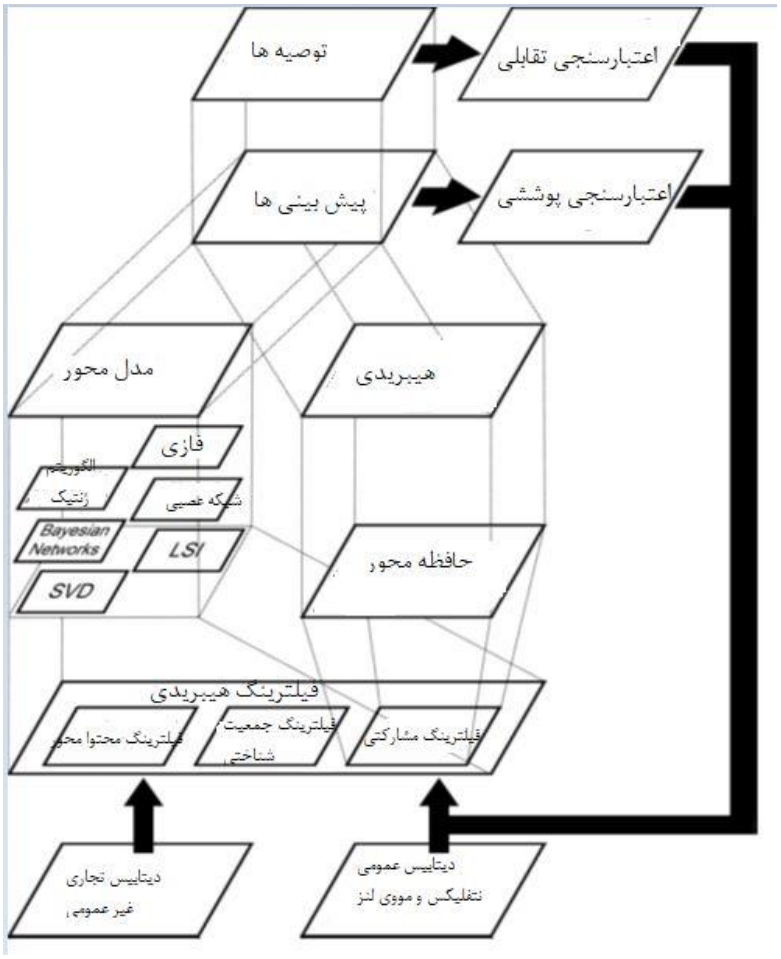
علی‌رغم شیوع آن‌ها در ادبیات CF، تکنیک‌های فوق هنوز محدودیت‌های خود را دارند؛ اگرچه روش‌های همسایگی برای پیاده‌سازی و ارائه توصیه‌های معقول ساده هستند؛ اما از مشکلات مقیاس‌پذیری رنج می‌برند، مشکلاتی که طبیعتاً در سناریوهای کلان داده حیاتی‌تر هستند.

به‌طورکلی، مدل‌های عامل پنهان نشان داده‌اند که توصیه‌های کارآمدی را در شرایط پراکندگی بالا انجام می‌دهند. اساساً، آن‌ها قبل از ارائه توصیه به یک مرحله یادگیری از قبل برای کشف پارامترهای مدل بهینه در فضای ویژگی‌های کم بعدی، هم برای کاربران و هم برای آیتم‌ها نیاز دارند. بااین‌حال، این روش‌های خطی معمولاً تفسیرپذیری نتایج توصیه شده را محدود می‌کنند؛ زیرا توانایی پیش‌بینی آن‌ها به عوامل ابعاد پایین تعداد کاربران و آیتم‌ها بستگی دارد. علاوه بر این، آن‌ها برای یادگیری بازنمایی‌های پنهان پیچیده‌تر و انتزاعی‌تر از داده‌های موجود تلاش می‌کنند و همچنین از نظر محاسباتی گران هستند.

عمدتاً به دلیل چالش‌های ذکر شده، الگوریتم‌های یادگیری عمیق در چند سال گذشته مورد توجه خیلی زیادی قرار گرفته‌اند (هوآ و همکاران[1] ۲۰۱۹). این روش‌ها برای یادگیری پارامترهای مدل، بر تکنیک‌های بهینه‌سازی غیرخطی تکیه می‌کنند. این تکنیک‌ها، الگوریتم‌ها را برای نمایش داده‌ها، به شکل قابل ملاحظه‌ای قوی‌تر می‌کند و قدرت تعمیم بالای آن‌ها نیز برخورد کارآمد با داده‌های خام یا غیرساخت‌یافته را تسهیل می‌کند. به این ترتیب، شبکه‌های عصبی عمیق قادر به دستیابی به انتزاعات عمیق‌تر از داده‌ها هستند که به طور مؤثر پیشرفت‌های قابل‌توجهی را در انواع وظایف مانند تشخیص (دامی، شیرازی و حسینی[2]، ۲۰۱۳) پیش‌بینی (دامی[3] ۲۰۱۶)، خوشه‌بندی (کریمی و همکاران[4] ۲۰۲۱) و طبقه‌بندی (خلف، ۲۰۲۱) نشان می‌دهند.

1. Hua et al
2. Dami, Shirazi & Hoseini
3. Dami
4. Karim et al

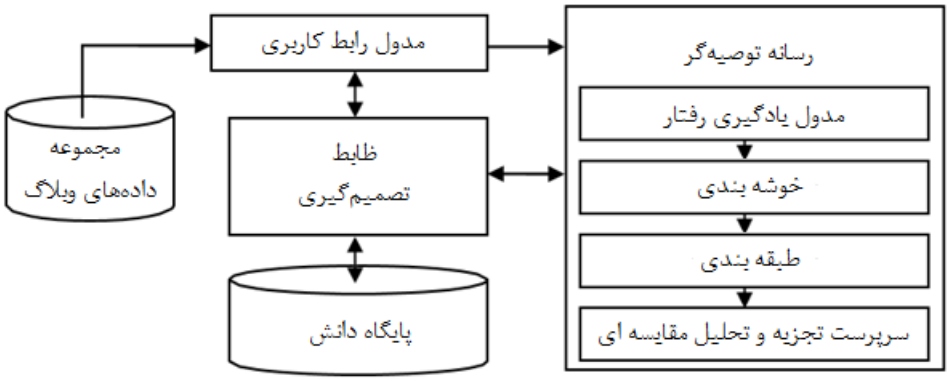
تأکید می‌کنیم که بسیاری از تکنیک‌های طبقه‌بندی دیگر را می‌توان برای مقابله با مشکل CF در نظر گرفت، مانند شبکه‌های عصبی احتمالی پیشرفته، طبقه‌بندی دینامیک عصبی و طبقه‌بندی‌کننده ماشین بولتزمن محدود (ژانگ و همکاران[1] ۲۰۱۸).
در نتیجه، مطالعه و توسعه تکنیک‌های یادگیری عمیق، پیشرفت‌های مهمی را در بسیاری از زمینه‌های تحقیقاتی علوم رایانه تسهیل کرده است که از آن‌ها، تشخیص اشیا، تشخیص گفتار، بینایی رایانه و پردازش زبان طبیعی را نقل می‌کنیم. از این نظر، مطالعات مربوط به یادگیری عمیق در سیستم‌های توصیه‌گر نیز مورد توجه قرار گرفته‌اند که به راحتی از افزایش تعداد مقالات منتشر شده در سال‌های گذشته قابل مشاهده است.
شکل ۲، نمونه‌هایی از روش‌های مورد استفاده در سیستم‌های توصیه‌گر را نشان می‌دهد (بابادیلا و همکاران[2]، ۲۰۱۳). این شکل، مهم‌ترین روش‌های مرسوم، تکنیک‌ها و الگوریتم‌های ارائه‌ی پیشنهاد و رابطه و گروه‌بندی‌های آن‌ها را ارائه می‌دهد.



شکل ۲: نمونه‌های مرسوم توصیه‌ها و رابطه‌ی آن‌ها (بابادیلا و همکاران، ۲۰۱۳)

معماری نهایی سیستم پیشنهادی در شکل ۳ نشان داده شده است. معماری سیستم پیشنهادی، از هفت مؤلفه اصلی مانند مجموعه داده‌های وبلاگ، مدول رابط کاربری، مدول ویژگی‌گزینی، چارچوب سیستم توصیه‌گر و نتایج، تشکیل شده است. چارچوب سیستم توصیه‌گر از دو مدول به نام‌های خوشه‌بندی و انتخاب منفی تشکیل شده است.

1. Zhang et al
2. Bobadilla et al



شکل ۳: معماری سیستم پیشنهادی

مجموعه داده‌های وبلاگ، از مجموعه داده استاندارد کاربران شبکه اجتماعی که از سراسر جهان جمع‌آوری می‌شود، تشکیل شده است. مدول رابط کاربر، اطلاعات ضروری را از مجموعه داده‌های استاندارد جمع‌آوری می‌کند و آن را به منظور پردازش بیشتر، به سیستم توصیه‌گر می‌فرستد. سیستم توصیه‌گر از چهار مدول فرعی مانند مدول تحلیل رفتار، مدول خوشه‌بندی، مدول طبقه‌بندی و عامل تحلیل مقایسه‌ای تشکیل شده است. در بین این موارد، مدول تحلیل رفتار، ویژگی‌های ضروری را از مجموعه داده مورد نظر، بر مبنای عوامل تعیین شده، انتخاب می‌کند و به منظور پردازش بیشتر، موارد ثبت شده را به مدول طبقه‌بندی ارسال می‌کند. مدول طبقه‌بندی، داده‌ها را با کمک ضابط تصمیم‌گیری و پایگاه قوانین، طبقه‌بندی می‌کند.

در نهایت، عامل تحلیل مقایسه‌ای، با کمک ضابط تصمیم‌گیری و پایگاه قوانین، تصمیم نهایی را می‌گیرد و مناسب بودن محصول فعلی را برای جامعه هدف می‌سنجد. ضابط تصمیم‌گیری، برای همه کاربران شبکه اجتماعی، بر مبنای این مسئله که چه رفتاری دارند و با استفاده از قوانین و نتایج طبقه‌بندی، تصمیم‌گیری می‌کند. این دانش، تجربه پیشین کاربران شبکه اجتماعی و علایق فعلی نوجوانان در شبکه‌های اجتماعی را در برمی‌گیرد و به منظور تصمیم‌گیری در مورد کاربران، با استفاده از ضابط تصمیم‌گیری، کارآمد است.

روش پژوهش

روش انجام تحقیق در بخش اول به‌صورت کتابخانه‌ای، از طریق مطالعه مقالات و تحقیقات علمی، گزارش‌های فنی، کتب، پایان‌نامه‌ها و پروژه‌های تحقیقاتی صورت گرفته و مطالب موجود در سایت‌های اینترنتی علمی و فنی در زمینه سیستم‌های توصیه‌گر و به ویژه سیستم‌های توصیه‌گر پالایش مشارکتی بررسی می‌شوند. در بخش بعدی، به‌صورت تحلیلی از طریق مقایسه روش‌های موجود، شناخت مزایا و معایب روش‌های کنونی و شناسایی خلأهای موجود در این زمینه انجام می‌گیرد. ارائه روش پیشنهادی برای بهبود سیستم توصیه‌گر پالایش مشارکتی با استفاده استخراج عمیق ویژگی‌ها آخرین مرحله تحقیق خواهد بود.

نمودار شکل ۴، مراحل انجام پژوهش را نشان می‌دهد. در بخش پیش‌پردازش، ابتدا داده‌های ورودی به سیستم، مورد پردازش اولیه قرار می‌گیرند و مقادیر ویژگی‌ها نرمال‌سازی می‌شوند. سپس، برای اینکه محاسبات را دقیق‌تر کنیم و زمان محاسبات نیز کاهش پیدا کند، اقدام به کاهش تعداد ابعاد داده‌های ورودی با روش DBN خواهیم. در نهایت تنها ویژگی‌هایی از داده‌های ورودی باقی می‌مانند که در انجام محاسبات مؤثر هستند.

بالاخره، با استفاده از فیلترینگ مشارکتی، اقدام به پیشنهاد اقلام به کاربران صورت می‌گیرد. در پایان، با توجه به خروجی‌های سیستم در پیشنهاددهی به کاربر، صحت پیشنهاد ارائه شده به کاربر را بازبینی می‌کنیم. در ادامه این بخش در مورد مراحل فوق، به‌تفصیل صحبت خواهیم کرد و آن‌ها را با جزییات بیشتری بررسی می‌کنیم.

شکل ۴: فرآیند انجام پژوهش

پیش‌پردازش داده‌ها

تمرکز روش ما بر بهره‌برداری از اطلاعات موجود ماتریس کاربر-آیتم تمرکز دارد که برای بهبود توصیه‌ها مفید است. ماتریس کاربر- کالا، پایه روش پالایش گروهی است که با اطلاعات جانبی همراه با کالا، ویژگی‌هایی مانند برچسب‌ها، سال انتشار، محتوای چندرسانه‌ای و غیره را بررسی می‌کند. برای کاربران، شغل، سن، جنسیت و غیره در نظر گرفته می‌شود. در نهایت، اطلاعات تعاملی مربوط به تعامل کاربر و کالا، مانند قیمت، خرید، کلیک بررسی می‌شوند. در این تحقیق، ما با جمع‌آوری اطلاعات جانبی غنی درباره آیتم‌ها، ماتریس کاربر-آیتم را افزایش دادیم. بعلاوه، اطلاعات تعاملی را به جای نرخ صریح کاربر از تحلیل بازخورد ضمنی و فعالیت‌های کاربر برای ساخت مشخصات شخصی کاربران به کار گرفتیم.

با این فرض که ویژگی‌های آیتم‌ها، به خصوص ویژگی‌های محتوایی آن مانند برچسب‌ها، یکی از دلایل اصلی برای علاقه‌مندی کاربر به اقلام هستند، اقلام را در ساختاری مناسب و بر اساس ویژگی‌هایی مانند برچسب‌ها و دیگر مشخصات برای استنتاج میزان ترجیحات کاربر در گروهی که آیتم به آن تعلق دارد، دسته‌بندی کردیم.

برچسب‌ها در واقع اقدامات کاربر در تعریف ویژگی‌های محتوایی است تا به‌طور ضمنی علایق کاربر با اطلاعاتی مهم درک شود. اولاً، برچسب‌های اجتماعی منابع ضمنی ارزشمندی از محتوای ژانر هر آهنگ هستند که بدون پرسش صریح، چیزی که فرد می‌خواهد را به او ارائه می‌دهند، ثانیاً، برچسب‌ها به دلیل درک کاربران از ویژگی‌های محتوا محور اقلام، دیدگاه مشابهی را از آن‌ها استخراج می‌کنند.

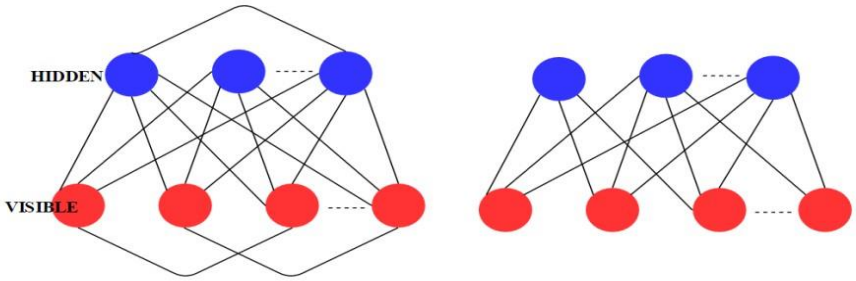
اگر آیتمی هیچ برچسبی نداشته باشد، سیستم توصیه پیشنهاد محتوای توصیفی هر آیتم را بازیابی می‌کند. در اینجا، هیچ رتبه‌بندی صریحی وجود ندارد؛ اما رفتار کاربر به شکل چندتایی ذخیره می‌شود که در آن، آیتم نشانگر چیزی است که کاربر به آن علاقه نشان داده است. برای غلبه بر مشکل پراکندگی و اندازه بزرگ مجموعه داده و پالایش آن برای تولید قوانین انجمنی، ما از ویژگی‌های آهنگ برای مدل‌سازی تعاملات کاربران با اقلام استفاده کردیم تا ترجیحات کاربر را نشان داده و روابط ضمنی بین کاربران را تشخیص دهیم.

استخراج ویژگی‌ها با استفاده از شبکه باور عمیق

شبکه باور عمیق (DBN)، یک مدل شبکه عصبی عمیق است که با انباشته سازی ماشین‌های بولتزمن محدود (RBM) شکل گرفته است. خروجی هر RBM قبلی، ورودی RBM بعدی است. داده‌ها از طریق RBM پایینی به شبکه وارد می‌شوند، دو لایه

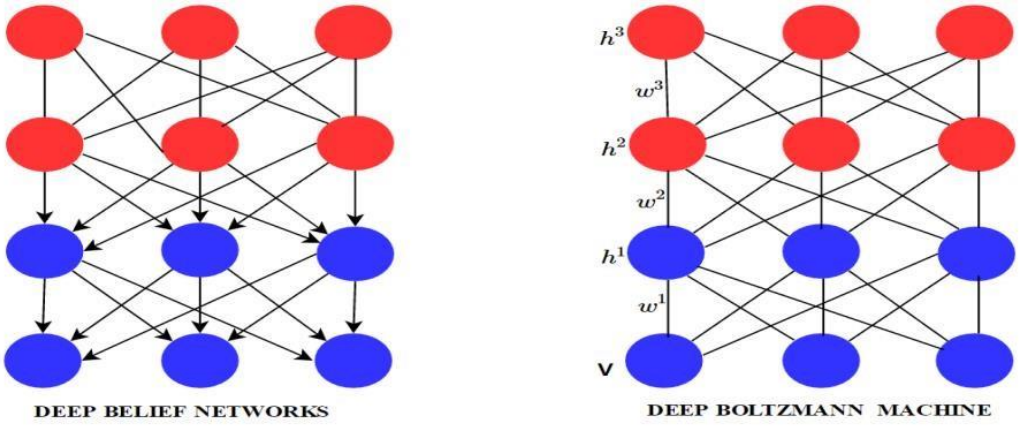
بالا یک اتصال دوتایی دارند، در بقیه لایه‌ها، اتصال یک جهته است. توجه دارید که DBM یک مدل گرافیکی بدون جهت است. در حال حاضر هیچ ارتباطی بین واحدهای یک‌لایه وجود ندارد؛ ولی بین واحدهای ورودی و واحدهای نهان ارتباط تعریف شده است. DBM در هنگامی‌که با منبع بزرگی از داده‌های بدون برچسب آموزش می‌بیند و با داده‌های برچسب دار تنظیم می‌شود به‌عنوان یک دسته‌بندی کننده، خوب عمل می‌کند.

ساختار این ماشین، از ماشین بولتزمن (BM) عمومی الهام گرفته است که عبارت است از شبکه‌ای از واحدهای مبتنی بر تصمیمات احتمالی جهت تعیین حالت‌های فعال و غیرفعال آن‌ها. آموزش الگوریتم BM آسان است، ولی در فرایند کند نمایش داده می‌شود. کاهش تعداد لایه‌های نهان یک DBM به یک ماشین بولتزمن محدود (RBM) را ایجاد می‌کند. DBM در هنگامی‌که با منبع داده بزرگ بدون برچسب آموزش داده می‌شود و با داده‌های برچسب دار خوب تنظیم می‌شود به‌عنوان یک دسته‌بندی کننده خوب عمل می‌کند. آموزش پشت‌هایی از RBMها و لایه‌های نهان بسیار، با استفاده از فعال‌سازی ویژگی روی یک RBM به‌عنوان ورودی برای لایه بعدی منجر به تشکیل شبکه باور عمیق (DBN) می‌شود. شکل ۵، ساختار یک RBM را نشان می‌دهد.



شکل ۵: چپ) یک ماشین بولتزمن با ویژگی‌های نهان روشن و خاموش و لایه متغیر نشان‌دهنده بردار احتمالی حالت‌های روشن و خاموش است؛ راست) یک RBM با واحدهای آشکار و غیرآشکار متصل و واحدهای نهان و غیرنهان متصل.

شکل ۶ یک DBN را نشان می‌دهد که از پشته‌هایی از ماشین‌های بولتزمن محدود (RBM) و یک لایه اضافی برای عمل تفکیک تشکیل شده است. RBM ها مدل‌های تولیدی احتمالی هستند که توزیع احتمال مشترک داده‌های مشاهده شده (آموزشی) را بدون استفاده از برچسب داده‌ها یاد می‌گیرند. هنگامی‌که ساختار یک DBN مشخص شد، هدف آموزش یادگیری وزن w بین لایه‌ها است. هر گره از گره دیگر در همان لایه مستقل است و با در نظر داشتن این خاصیت برای تمامی گره‌ها می‌توان وزن‌های تولیدی هر RBM را آموزش داد. این الگوریتم پس از آن به سراغ الگوریتم یادگیری لایه به لایه حریصانه می‌رود که هر پشته لایه RBM را در هر دفعه یاد بگیرد. در شکل زیر در سمت چپ، لایه بالا به رنگ قرمز یک RBM و لایه پایین به رنگ آبی شبکه باور حلقوی جهت‌دار را شکل می‌دهد.



شکل ۶: شبکه باور عمیق (DBN) سه لایه و ماشین بولتزمن عمیق سه لایه

توصیه به کاربر با استفاده از فیلترینگ مشارکتی

سیستم‌های توصیه‌گر با تحلیل و بررسی داده‌های متعلق به کاربران، یک سری آیتم‌های خاص را بر مبنای علایق به کاربران پیشنهاد می‌کنند. هدف از آنالیز داده‌های مربوط به کاربران، استخراج الگوهای هر کاربر به منظور پیش‌بینی آیتم‌ها است. یکی از مهم‌ترین روش‌ها در سیستم‌های توصیه‌گر، روش فیلترینگ مشارکتی است. در سیستم‌های توصیه‌گر مبتنی بر فیلترینگ مشارکتی از معیارهای شباهت جهت کشف کردن کاربران مشابه با کاربر جدید برای ارائه پیشنهاد استفاده می‌شود. از چالش‌های سیستم‌های توصیه‌گر مبتنی بر فیلترینگ مشارکتی می‌توان به فاکتورهای شباهت و تشخیص همسایگی اشاره کرد.

فیلترینگ مشارکتی، تشابه بین کاربر و گزینه را به‌صورت هم‌زمان در نظر می‌گیرد. مدل‌های فیلترینگ مشارکتی می‌تواند گزینه‌ای به کاربر الف پیشنهاد دهد که بر اساس تشابه‌های بین کاربر الف و کاربر ب است. برای مثال کاربر الف وارد سایت نمایش فیلم می‌شود، مدل فیلترینگ مشارکتی باید هم فیلم‌هایی مشابه فیلم‌هایی که کاربر الف در گذشته نگاه کرده است را پیشنهاد دهد و هم فیلم‌هایی که کاربران مشابه به کاربر الف تماشا کرده‌اند.

فیلتر مشارکتی روشی برای پیش‌بینی خودکار (فیلتر کردن) در مورد علایق یک کاربر با جمع‌آوری اولویت‌ها یا اطلاعات سلیقه از بسیاری از کاربران (همکاری) است. فرض اساسی رویکرد فیلترینگ مشارکتی این است که اگر یک فرد A نظر یکسانی با شخص B در مورد یک موضوع داشته باشد، احتمال دارد که A نظر B را در مورد موضوعی متفاوت نسبت به نظر شخصی که به‌طور تصادفی انتخاب شده است داشته باشد. به‌عنوان‌مثال، یک سیستم توصیه فیلتر مشترک برای ترجیحات در تلویزیون برنامه‌نویسی می‌تواند پیش‌بینی کند که کاربر کدام برنامه تلویزیونی را دوست دارد، با توجه به فهرستی جزئی از سلایق آن کاربر (دوست داشتن یا دوست نداشتن). باید توجه داشت که این پیش‌بینی‌ها مختص کاربر است؛ اما از اطلاعات جمع‌آوری شده از بسیاری از کاربران استفاده می‌شود. این با رویکرد ساده‌تر یعنی دادن میانگین امتیاز (غیراختصاصی) برای هر آیتم موردعلاقه، به‌عنوان‌مثال بر اساس تعداد آرا قابل انجام است.

در مفهوم کلی‌تر، فیلتر مشترک روند فیلتر برای اطلاعات و یا الگوهای استفاده از تکنیک‌های مربوط به همکاری میان عوامل متعدد، نظرات، منابع داده و غیره است. کاربردهای فیلتر مشارکتی معمولاً شامل مجموعه داده‌های بسیار بزرگی است. روش‌های فیلتر مشترک برای انواع مختلفی از داده‌ها از جمله: داده‌های سنجش و پایش، مانند اکتشاف مواد معدنی، سنجش محیطی در مناطق بزرگی ا حسگرهای متعدد به کار گرفته شده‌اند.

داده‌های مالی، مانند مؤسسات خدمات مالی که بسیاری از منابع مالی را ادغام می‌کنند یا در تجارت الکترونیک و برنامه‌های کاربردی وب که تمرکز بر روی داده‌های کاربر و غیره است. این بخش بر فیلتر کردن مشترک برای داده‌های کاربر متمرکز است، اگرچه برخی از روش‌ها و رویکردها ممکن است برای سایر برنامه‌های کاربردی اصلی نیز کاربرد داشته باشد.

در روش فیلترینگ مشارکتی بر اساس شباهت رفتاری و الگوهای عملکردی کاربرانی که شباهت‌های رفتاری و الگوهای مشابه با کاربر فعلی در گذشته داشته‌اند، پیشنهادهای ارائه می‌شود. شاید تعریف آن کمی پیچیده باشد؛ ولی به‌طور ساده روش فیلترینگ مشارکتی بر این فرض استوار است که کاربرانی که یک سری نظرهای مشابه درباره یک آیتم منظور از آیتم، فیلم، عکس، موزیک یا هر چیز دیگری است که پیشنهاد می‌شود دارند، درباره آیتم‌های دیگر هم نظرات مشابه دارند.

تجزیه‌وتحلیل یافته‌ها

مجموعه داده‌ها

در این مقاله، جهت ارزیابی مدل پیشنهادی از مجموعه داده دنیای واقعی MovieLens استفاده شده است. این مجموعه داده‌ها حاوی رأی‌هایی هستند که کاربران به فیلم‌های مختلف داده‌اند. اولویت‌دهی در مجموعه داده‌های MovieLens با افزایش یک ستاره، توسط مقیاس ۵ ستاره در نظر گرفته می‌شود که هر کاربر می‌تواند به هر فیلم رأی‌دهی کند. سه نسخه از این مجموعه داده‌ها به شرح زیر استفاده شده است که همگی این مجموعه داده‌ها آز آدرس https://grouplens.org/datasets/movielens/ قابل‌دسترسی هستند.

۱) مجموعه داده‌های MovieL ens Small شامل ۱۲ هزار رأی کاربران است که در آن به تعداد ۱۳۰۰ کاربر در مورد 2300 فیلم رأی‌دهی کرده‌اند.

۲) مجموعه داده‌های MovieL ens Medium شامل حدود ۲ میلیون اولویت ناشناس است و اولویت‌دهی ۷۰۰۰ فیلم توسط ۱۱۰۰۰ کاربر را شامل می‌شود که هر کاربر حداقل ۱۸ مورد را رتبه‌بندی کرده است.

۳) مجموعه داده‌های MovieLens L arge شامل ۱۸ میلیون رأی‌دهی است که ۱۶۸۰۰۰ کاربر در مورد ۱۹۰۰۰ فیلم انجام داده‌اند.

شبیه‌سازی روش پیشنهادی

برای شبیه‌سازی روش پیشنهادی از جعبه‌ابزار یادگیری عمیق در ابزار MATLAB بهره گرفته شده است. این نرم‌افزار بسیار مناسب انجام محاسبات عددی و یک زبان برنامه‌نویسی نسل چهارم است که دارای کاربری‌های اثبات شده حوزه پژوهش جاری و بسیاری از زمینه‌های دیگر مربوط به هوش مصنوعی و یادگیری ماشین و یادگیری عمیق است. جعبه‌ابزار یادگیری عمیق MATLAB، یک چارچوب یکپارچه و منسجم را برای طراحی و بهینه‌سازی انواع الگوریتم‌های یادگیری عمیق تهیه و ارائه کرده است. می‌توان از شبکه‌ها به انواع شبکه‌های برای آماده‌سازی و بهبودبخشی فرآیندهای رده‌بندی و رگرسیون اطلاعات، سری‌های زمانی و نوشتاری استفاده کرد و کاربردهای بسیار وسیعی در سایر حوزه‌های هوش مصنوعی و یادگیری ماشینی دارا است. برنامه‌های کاربردی و نمودارها در بصری‌سازی فعالیت‌ها، طراحی ساختارهای مرتبط با الگوریتم‌ها و نمایش آموزش در این جعبه‌ابزار موجود و قابل استفاده است.

معیارهای ارزیابی

بعد از بررسی در مراجع مختلفی که در زمینه کیفیت‌سنجی قوانین انجمنی وجود دارند ۲ متریک مورد انتخاب واقع شدند که در مورد آن‌ها، شرح ذکر می‌شود. قبل از ورود به بحث متریک‌های مذکور، یک سری پیش‌نیازها بیان می‌شود که در تمامی روش‌های سنجش کیفیت قواعد انجمنی، استفاده می‌شود. قواعد انجمنی به‌صورت $H \leftarrow B$ بیان می‌شوند و به این معنی است که اگر H در یک مجموعه آیتم ظاهر شود، آنگاه B نیز در آن مجموعه آیتم حضور پیدا می‌کند. شکل ۷ حالات حضور بخش مقدم و تالی قواعد را نشان می‌دهد.

	B	$\overline{B}$	
H	$n(HB)$	$n(H\overline{B})$	$n(H)$
$\overline{H}$	$n(\overline{H}B)$	$n(\overline{H}\overline{B})$	$n(\overline{H})$
	$n(B)$	$n(\overline{B})$	N

شکل ۷: حالات حضور بخش مقدم و تالی قواعد انجمنی

که در آن B شامل مواردی است که مقدم قانون، ظهور پیدا می‌کند؛ H شامل مواردی است که تالی قانون، ظهور پیدا می‌کند؛ $\overline{B}$ شامل مواردی است که مقدم قانون، ظهور پیدا نمی‌کند و $\overline{H}$ شامل مواردی است که تالی قانون، ظهور پیدا نمی‌کند. $n(X)$ تعداد اعضای موجود در مجموعه X را نشان می‌دهد و N نیز بیان‌گر تعداد کل داده‌های موجود در مجموعه داده‌ها است. بر این اساس، متریک‌های پوشش[1] و حمایت[2] به‌صورت زیر تعریف می‌شوند.

- **پوشش**

این متریک می‌سنجد که چه تعدادی از نمونه‌های ورودی به سیستم که برای کشف قواعد در نظر گرفته شده‌اند، توسط تالی قانون پوشش داده می‌شوند. به‌عبارت‌دیگر با این روش سنجیده می‌شود که قانون موردنظر تا چه حد دارای عمومیت[3] است. بیان ریاضی این روش به‌صورت رابطه (۱) تعریف می‌شود.

$$Cov(H \leftarrow B) = p(B) \qquad (۱)$$

- **حمایت**

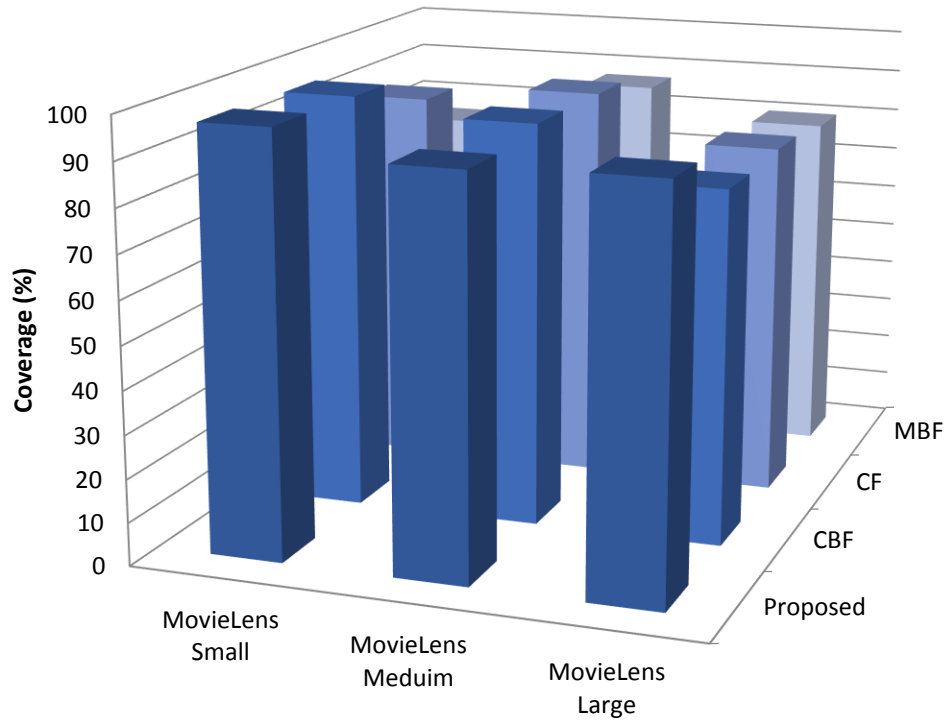
این متریک، تعداد حضور و نرخ آن را برای مقدم و تالی قانون را با هم و به‌عنوان یک آیتم ست در مجموعه داده‌های ورودی می‌سنجد. در واقع این روش مشخص می‌کند که این قانون چقدر از مجموعه آیتم ست‌ها را حمایت می‌کند. بیان ریاضی این روش به شرح زیر است. نکته مهم اینکه این رابطه نسبت به مقدم و تالی قانون حالت تقارن دارد یا به‌عبارت‌دیگر Support یک قانون با معکوس آن مساوی است.

$$Sup(H \leftarrow B) = p(HB) \qquad (۲)$$

نتایج ارزیابی

برای سنجش نتایج حاصله از مقایسه روش پیشنهادی (Proposed) با سه روش فیلترینگ مبتنی بر محتوی (CFB)، فیلترینگ مبتنی بر حافظه (MBF) و فیلترینگ مشارکتی (CF) استفاده شده است. شکل ۸، مقایسه متریک پوشش نتایج به‌دست‌آمده از روش پیشنهادی نسبت به سایر روش‌های مورد مقایسه را به ازای هر مجموعه داده نشان می‌دهد. همان‌طور که قابل مشاهده است، روش پیشنهادی از نظر پوشش نسبت به سایر روش‌های مورد مقایسه، عملکرد بهتری را به نمایش گذاشته است. البته در بعضی در حالات نیز، سایر روش‌های مورد مقایسه تا حدی بهتر عمل کرده‌اند. در مجموعه داده‌های MovieLens Small روش پیشنهادی به همراه CBF بهترین عملکرد را داشته‌اند. در مجموعه داده‌های MovieLens Medium روش‌های CF و CBF عملکرد بهتری داشته‌اند و روش MBF کمی ضعیف‌تر عمل کرده است. در مجموعه داده‌های MovieLens Large روش پیشنهادی با فاصله از سایر روش‌های مورد مقایسه بهتر عمل کرده و کارایی بهتری داشته و سایر روش‌های مورد مقایسه به میزان قابل‌توجهی عقب‌تر از روش پیشنهادی هستند.

1. Coverage
2. Support
3. Generality



	MovieLens Small	MovieLens Meduim	MovieLens Large
Proposed	96.97467	90.93349	92.13778
CBF	96.3824	92.90773	81.38804
CF	88.60389	92.52276	82.14813
MBF	75.55416	87.07386	80.0653

شکل ۸: مقایسه متریک پوشش روش پیشنهادی با روش‌های مورد مقایسه

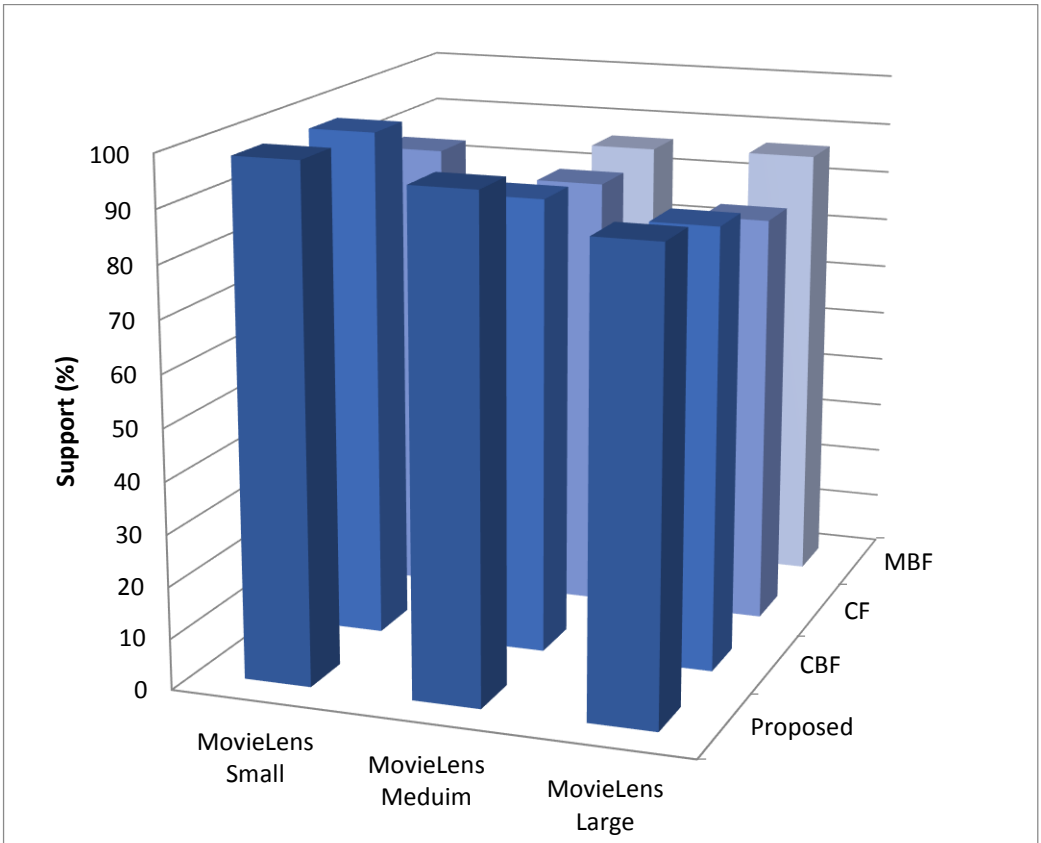
جدول ۱، میانگین پوشش را برای روش‌های مختلف نشان می‌دهد و همان‌طور که مشهود است، روش پیشنهادی دارای میانگین پوشش بهتری بوده است؛ روش CBF در رده بعدی عملکرد از نظر متریک پوشش قرار می‌گیرد و سایر روش‌ها نیز پس از این دو قرار گرفته‌اند.

جدول ۱: مقایسه میانگین پوشش روش پیشنهادی با سایر روش‌های مورد مقایسه

Proposed	CBF	CF	MBF
93.34864667	90.22605667	87.75826	80.89777333

شکل ۹، مقایسه متریک حمایت نتایج حاصله از روش پیشنهادی را نسبت به سایر روش‌های مورد مقایسه به ازای هر مجموعه داده نشان می‌دهد. در مجموعه داده‌های MovieLens Small روش پیشنهادی به همراه CBF بهترین عملکرد را داشته‌اند و دو روش دیگر با فاصله پس‌ازاین دو قرار گرفته‌اند در مجموعه داده‌های MovieLens Medium روش پیشنهادی با فاصله از سایر روش‌های مورد مقایسه بهتر عمل کرده است. هر سه روش دیگر به میزان قابل‌توجهی عقب‌تر از روش پیشنهادی هستند. در

مجموعه داده‌های MovieLens Large روش پیشنهادی عملکرد بهتری را نشان داده و با فاصله کمی از آن، روش MBF قرار دارد و دو روش دیگر نیز در رتبه‌های بعدی قرار می‌گیرند.



	MovieLens Small	MovieLens Meduim	MovieLens Large
Proposed	98.49620926	95.39080721	88.65505059
CBF	98.08480929	87.67099391	84.88673663
CF	89.08804943	84.55902436	79.63723649
MBF	73.92897335	85.97546722	86.41544425

شکل ۹: مقایسه متریک حمایت روش پیشنهادی با روش‌های مورد مقایسه

جدول ۲، میانگین متریک حمایت را برای هریک از روش‌ها نمایش می‌دهد و همان‌طور که واضح است روش پیشنهادی نسبت به سایر روش‌ها عملکرد بهتری داشته است؛ پس از آن روش CBF و در رتبه‌های بعدی روش‌های CF و MBF عملکرد ضعیف‌تری داشته‌اند.

جدول ۲: مقایسه میانگین حمایت روش پیشنهادی با سایر روش‌های مورد مقایسه

Proposed	CBF	CF	MBF
94.18068902	90.21417994	84.42810343	82.10662827

نتیجه‌گیری و پیشنهادها

امروزه زندگی انسان در جوامع مختلف با انواع سیستم‌های تحت وب پیوند خورده است و بسیاری از امور روزمره مردم در تعامل با این سیستم‌ها شکل می‌گیرد. واضح است که حجم بسیار زیادی از اطلاعات در این سیستم‌ها موجود هستند که هر کاربر بسته به نیاز خود صرفاً به بخش مشخص و گزینش شده‌ای از این اطلاعات نیاز دارد تا امور روزمره خود را به انجام برساند. درواقع نیاز به بهبود روش پالایش مشارکتی به‌وضوح احساس می‌شود؛ بنابراین در این پژوهش با بهره‌گیری از تکنیک‌های یادگیری عمیق، روشی برای بهبود سیستم توصیه گر پالایش مشارکتی، با استفاده استخراج عمیق ویژگی‌ها، ارائه شد تا بتواند پیشنهادهای مؤثرتر و مطلوب‌تری را به کاربر سیستم ارائه دهد. در این پژوهش، جهت ارزیابی مدل پیشنهادی از مجموعه داده دنیای واقعی MovieLens استفاده شده است. این مجموعه داده‌ها حاوی رأی‌هایی هستند که کاربران به فیلم‌های مختلف داده‌اند. اولویت‌دهی در مجموعه داده‌های MovieLens با افزایش یک ستاره، توسط مقیاس ۵ ستاره در نظر گرفته می‌شود که هر کاربر می‌تواند به هر فیلم رأی‌دهی کند. سه نسخه از این مجموعه داده‌ها استفاده شده است. برای سنجش نتایج حاصله، از مقایسه روش پیشنهادی با سه روش دیگر CBF، MBF و CF استفاده شده است. روش پیشنهادی از هر دو منظر پوشش و حمایت نسبت به سایر روش‌های مورد مقایسه، عملکرد بهتری را به نمایش گذاشته است. یکی از پژوهش‌های مرتبط که به‌عنوان پیشنهاد برای کارهای آتی این پژوهش می‌تواند مطرح باشد این است که در فیلترینگ تعاملی به‌صورت ترکیبی (هیبرید) با روش‌های دیگر استفاده شود. همچنین توسعه روش پیشنهادی جهت کاربرد در سیستم‌های موازی و توزیع‌شده نیز می‌تواند به‌عنوان نقاط گسترش روش پیشنهادی باشد.

منابع

Bobadilla, J, Ortega, F, Hernando, A, & Gutiérrez, A. (2013). Recommender systems survey. Knowledge-based systems, 46, 109-132.

Camacho, L. A. G, & Alves-Souza, S. N. (2018). Social network data to alleviate cold-start in recommender system: A systematic review. Information Processing & Management, 54(4), 529-544.

Dami, S, & Yahaghizadeh, M. (2018, February). Efficient event prediction in an IOT environment based on LDA model and support vector machine. In 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS) (pp. 135-138). IEEE.

Dami, S, Shirazi, H, & Hoseini, M. (2013). A Data Mining Model for Anomaly Detection of Satellite Launch Vehicle. journal of Advanced Defense Science and Technology, 4(1), 51-63.

Dami, S. (2016). News Events Prediction Based on Casual Inference in First-Order Logic (FOL). Journal of Soft Computing and Information Technology, 5(4), 11-25.

Fan, C, & Mostafavi, A. (2019). A graph-based method for social sensing of infrastructure disruptions in disasters. Computer-Aided Civil and Infrastructure Engineering, 34(12), 1055-1070.

Hua, C, Wang, H, Wang, H, Lu, S, Liu, C, & Khalid, S. M. (2019). A novel method of building functional brain network using deep learning algorithm with application in proficiency detection. International journal of neural systems, 29(01), 1850015.

Karim, M. R, Beyan, O, Zappa, A, Costa, I. G, Rebholz-Schuhmann, D, Cochez, M, & Decker, S. (2021). Deep learning-based clustering approaches for bioinformatics. Briefings in Bioinformatics, 22(1), 393-415.

Khallaf, R, & Khallaf, M. (2021). Classification and analysis of deep learning applications in construction: A systematic literature review. Automation in construction, 129, 103760.

Koren, Y, Rendle, S, & Bell, R. (2022). Advances in collaborative filtering. Recommender systems handbook, 91-142.

Liu, X, & Cheung, Y. M. (2018). On incremental collaborative appearance model and regional particle filtering for lip region tracking. Integrated Computer-Aided Engineering, 25(1), 63-80.

Ou-Yang, L, Lu, F, Zhang, Z. C, & Wu, M. (2022). Matrix factorization for biomedical link prediction and scRNA-seq data imputation: an empirical survey. Briefings in Bioinformatics, 23(1), bbab479.

Sethi, R, & Mehrotra, M. (2021). Cold start in recommender systems—a survey from domain perspective. In Intelligent Data Communication Technologies and Internet of Things (pp. 223-232). Springer, Singapore.

Shao, B, Li, X, & Bian, G. (2021). A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. Expert Systems with Applications, 165, 113764.

Zhang, N, Ding, S, Zhang, J, & Xue, Y. (2018). An overview on restricted Boltzmann machines. Neurocomputing, 275, 1186-1199.