Journal of Information Technology Management
ISSN: 2008-5893
EISSN: 2423-5059
Vol. 9, No. 1; PP. 61 - 82
Spring 2017

University of Tehran
Faculty of Management

A New Similarity Method to Optimize Business in the Online Stores Using the Rating Time Technology

Nayereh Zaghari [1], Ardeshir Zamani [2]

Abstract: These days, Emergence of e-commerce web sites is one of the important consequences of the Internet in modern times, but products data is growing exponentially. In such environment, customers face a problem in finding optimized information among huge data bases about the items or desired products. In order to assist buyers, large e-commerce companies are planning to introduce their own recommender systems to help their customers in making a better choice among the items. Due to high percentage error , a basic method to build such systems is not usually being applied. In this essay, two methods have been suggested in order to improve recommendations in recommender systems. Collaborative filtering method is one of the most successful methods used in the system, but using this method that it has common problem the increasing number of users and products, therefore system do not inability to request the requirement of cold start and data sparsity. Two methods have been suggested in order to improve recommendations in recommender systems. To resolve this problem, a new method has been introduced in which by integrating rating time by Pearson also combining semantic technology with social networks offers a solution to reduce issues such as "cold start" and generally "data sparsity" in recommender systems. The result of simulating showed that the proposed approach provided better performance and more accurate predictions in addition of more consistent with user preferences.

Key words: *Data sparsity, Internet store, Recommender systems, Users rating time.*

1. Ph.D. Candidate in Computer Engineering, Azad University, Tehran, Iran
2. Ph.D. Candidate in Business Management, Tehran University, Tehran, Iran

Submitted: 17/ May / 2016
Accepted: 27/ February / 2017
Corresponding Author: Nayereh Zaghari
Email: Nasrin.zaghari@gmail.com

مدیریت فناوری اطلاعات
Journal of Information Technology Management
دانشکدهٔ مدیریت دانشگاه تهران
دورهٔ ۹، شمارهٔ ۱
بهار ۱۳۹۶
صص. ۸۲- ۶۱

ارائهٔ روش جدید شباهت‌یابی برای بهینه‌سازی کسب‌وکار در فروشگاه‌های اینترنتی با استفاده از تکنولوژی‌های زمان نرخ‌دهی

نیره زاغری[۱]، اردشیر زمانی[۲]

چکیده: یکی از پیامدهای مهم گسترش اینترنت در عصر حاضر، پیدایش وبگاه‌های تجارت الکترونیکی است؛ اما تنوع اقلام عرضه‌شده می‌تواند انتخاب محصولات مناسب را برای مشتریان به فرایند پیچیده‌ای تبدیل کند. شرکت‌ها برای رفع این مشکل استفاده از سیستم‌های پیشنهاددهنده را توصیه می‌کنند. به‌دلیل درصد زیاد خطا در پالایش و ارائهٔ پیشنهادها، معمولاً در این‌گونه سیستم‌ها از چند روش پایه برای پیشنهاد اقلام مورد علاقهٔ کاربر استفاده می‌شود. روش پالایش همکارانه، یکی از موفق‌ترین روش‌های به‌کاررفته در این سیستم‌هاست؛ اما استفاده از این روش با افزایش تعداد کاربران و محصولات، مجموعه را با مشکلاتی مانند ناتوانی در پاسخگویی به نیازهای کاربران جدید و تُنُکی داده مواجه می‌کند. به‌منظور رفع مشکل یادشده، در این پژوهش روش جدیدی معرفی شده است که با ادغام زمان نرخ‌دهی کاربران با معیار شباهت‌یابی پیرسون و همچنین ادغام تکنولوژی‌های معنایی و شبکهٔ اجتماعی، راه‌حلی برای کاهش مشکل کاربران جدید و تنکی داده‌ها ارائه می‌کند. نتایج پیاده سازی الگوریتم‌ها نشان می‌دهد رویکرد ارائه‌شده عملکرد بهتر و صحت بیشتری دارد و پیش‌بینی‌های آن با سلیقهٔ کاربران همخوانی بیشتری دارد.

واژه‌های کلیدی: تُنُکی داده، زمان نرخ‌دهی کاربر، سیستم‌های پیشنهاددهنده، فروشگاه‌های اینترنتی.

۱. دانشجوی دکتری مهندسی کامپیوتر، گرایش هوش مصنوعی، دانشگاه آزاد، تهران، ایران
۲. دانشجوی دکتری مدیریت بازرگانی، گرایش استراتژیک، دانشگاه تهران، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۵/۰۲/۲۸
تاریخ پذیرش نهایی مقاله: ۱۳۹۵/۱۲/۰۹
نویسندهٔ مسئول مقاله: نیره زاغری
E-mail: Nasrin.zaghari@gmail.com

مقدمه

با گسترش و پویاترشدن فعالیت‌ها در قرن بیست‌ویک، دیگر استفاده از شیوه‌های سنتی مدیریت فرایندها پاسخگوی نیاز شرکت‌ها نیست و شرکت‌ها باید از سیستم‌های نوینی برای بهینه‌سازی فرایندهای کسب‌وکار خود استفاده کنند (زمانی و رحمتی، ۲۰۱۴). سیستم‌های پیشنهاددهنده از جمله این سیستم‌های نوین است. سیستم‌های پیشنهاددهنده، نوع ویژه‌ای از سیستم‌های پالایش اطلاعات‌اند که آیتم‌ها را بر اساس این که کدام‌یک برای کاربر جذاب است، از مجموعهٔ بزرگی از آیتم‌ها و کاربران پالایش می‌کنند. این سیستم رویکردی است که برای مواجهه با مشکلات حجم فراوان و رو به رشد اطلاعات، ارائه شده است و به کاربر کمک می‌کند تا در میان حجم عظیم اطلاعات، سریع‌تر به هدف خود نزدیک شود. برخی سامانهٔ پیشنهادگر را معادل پالایش گروهی می‌دانند (هیل، استید، رسنستین و فورناس، ۱۹۹۵).

سیستم پیشنهاددهنده یا سامانهٔ پیشنهادگر با تحلیل رفتار کاربر خود، مناسب‌ترین آیتم‌ها (داده، اطلاعات، کالا و...) را پیشنهاد می‌دهد. به زبان ساده‌تر، در سیستم‌های پیشنهاددهنده تلاش می‌شود با پیش‌بینی شیوهٔ تفکر کاربر (به‌کمک اطلاعاتی که از نحوهٔ رفتار کاربر یا کاربران مشابه وی و نظر آنها ذخیره شده است)، مناسب‌ترین و نزدیک‌ترین کالا به سلیقهٔ او شناسایی و پیشنهاد شود. سیستم‌های پیشنهاددهنده براساس تخمین نرخ و چگونگی پیشنهادها به چهار گروه دسته‌بندی می‌شوند: ۱. محتوامحور؛ ۲. هم‌بستگی (یا اشتراکی)؛ ۳. پالایش ترکیبی و ۴. دانش‌محور.

به‌طور معمول، سیستم پیشنهاددهنده پروفایل کاربر را با برخی ویژگی‌های مرجع مقایسه می‌کند و به‌دنبال پیش‌بینی بهایی است که کاربر به کالایی که هنوز در نظر گرفته نشده است، می‌دهد (شایند و کولکارنی، ۲۰۱۲). از طرفی می‌توان با استفاده از دو روش پایهٔ محتوامحور و اشتراکی، به پیش‌بینی نرخ پرداخت. روش محتوامحور، براساس میزان شباهت محتوایی آیتم هدف با آیتم‌های دیگری که کاربر قبلاً نرخ داده است، نرخ پیش‌بینی‌شدهٔ کاربر به آیتم مدنظر را محاسبه می‌کند. در روش اشتراکی، می‌توان از هر دو مورد کمک گرفت؛ به بیان دیگر، هم با استفاده از بهای آیتم‌های همسایه و هم براساس تشخیص کاربران همسایه، به پیش‌بینی بها پرداخت. نام این دو انتخاب به‌ترتیب «اشتراکی آیتم‌محور» و «اشتراکی کاربرمحور» است. در نوع آیتم‌محور، باید آیتم‌های نزدیک به آیتم مدنظر را از لحاظ نزدیکی بهایی که کاربران دیگر به آنها داده‌اند، تشخیص داد؛ سپس مقدار میانگین بهای آنها را با توجه به میزان شباهتشان به آیتم هدف، به‌عنوان نرخ پیش‌بینی‌شده‌ای که در اختیار کاربر هدف قرار می‌گیرد، درنظر گرفت. درنوع کاربرمحور، نیز از میانگین‌گیری استفاده می‌کنیم، اما میانگین نرخ کاربران همسایه با تأثیرپذیری از شباهت هر یک به کاربر هدف، به‌منزلهٔ بهای پیش‌بینی‌شده تلقی می‌شود (شارداناد و مائز، ۱۹۹۵).

سیستم پیشنهاددهنده محصولات فروشگاه‌های آنلاین را با توجه به سلیقهٔ مشتری غربال می‌کند. سیستم بر پایهٔ پروفایل مشتری ساخته می‌شود و به همین دلیل، امکان پیداکردن محصول مدنظر کاربر را افزایش می‌دهد. از این رو، برای درک نیاز مشتریان و پاسخ مناسب به این نیازها، ضروری است از الگوریتم‌های کارآمدی استفاده شود. بر همین اساس، پژوهش حاضر قصد دارد با مطالعه و تحقیق در این حوزه و ارائهٔ الگوریتمی برای بهبود سیستم پیشنهاددهندهٔ مبتنی بر راهکار پالایش همکارانهٔ کاربر، به این مسئله پاسخ دهد و از این طریق پیشنهادهایی ارائه دهد که با سلیقه و ترجیح مشتریان همخوانی بیشتری داشته باشد. استفاده از روش پالایش همکارانهٔ مبتنی بر حافظه، یکی از روش‌های موفق و شناخته‌شده در توسعهٔ سیستم‌های پیشنهاددهنده است (شامبور و لو، ۲۰۱۲). در این پژوهش برای پیشنهاد کالا به مشتری، از نظر کاربرانی استفاده می‌شود که آن کالا را دیده‌اند یا تهیه کرده‌اند. یکی از چالش‌های پیش روی این روش، ورود کاربران جدید و نداشتن پیشینهٔ اطلاعاتی از آنها و یافتن کاربرانی است که سلیقه و ترجیح مشابهی با کاربران جدید دارند؛ در این صورت، سیستم نمی‌تواند پیشنهادی در اختیار آنها قرار دهد. از این رو پژوهش حاضر قصد دارد از اطلاعات فردی کاربران، به‌منظور بهبود نتایج سیستم پیشنهاددهنده استفاده کند. بدین منظور، بر اساس اطلاعات فردی کاربران ماتریسی تشکیل داده و معیار جدیدی برای محاسبهٔ شباهت با ترکیب درصد وزن‌دار زمان و معنای شباهت پالایش همکارانه ایجاد می‌شود. برای پیشگیری از مشکل یادشده، می‌توان شباهت کاربران جدید را با درصد وزن بزرگ‌تری همراه کرد و برعکس آن، برای کاربران قدیمی‌تر درصد وزن شباهت را کاهش داد. روش کاربرمحور پیشنهادی در این پژوهش، از تأثیر زمان نرخ‌دهی بر علاقهٔ کاربران در گذشت زمان برای بهبود برخورد با مشکل تُنُکی داده و روابط اعتماد بین کاربران در شبکه‌های اجتماعی، به‌منظور تشخیص هرچه بهتر علایق کاربر و در نتیجه بهبود توصیه‌ها، استفاده می‌کند. برای تشخیص بهتر شباهت میان آیتم‌ها، روش آیتم‌محور پیشنهادی از طبقه‌بندی موضوعی و معنایی آیتم‌ها نیز بهره می‌برد. همچنین از کلمات کلیدی استخراج‌شده با روش‌های معنایی از محتوای آیتم‌ها در جهت بهبود توصیه‌ها، به‌خصوص هنگام برخورد با مشکل «کاربر جدید»، استفاده شده است. نوآوری این پژوهش در این است که با ادغام زمان نرخ‌دهی کاربران با معیار شباهت‌یابی پیرسون برای کاهش تُنُکی داده و ادغام تکنولوژی‌های معنایی و شبکهٔ اجتماعی، مشکلات «کاربر جدید» و به‌طور کلی «تُنُکی داده‌ها» در سیستم‌های پیشنهاددهنده حل شده است. این نوشتار تلاش می‌کند سیستم‌های پیشنهاددهنده را بررسی کند و نسل حاضر از روش‌های سیستم‌های پیشنهاددهنده را که به چهار گروه کلی دسته‌بندی می‌شوند، معرفی نماید. همچنین پژوهش حاضر درصدد است با ارائهٔ روشی بتواند ترافیک داده‌ها را با توجه به سیستم‌های ذکر شده برطرف کند.

پیشینهٔ پژوهش

کاهش پیوستهٔ هزینه‌های ذخیره‌سازی و پردازش داده‌ها موجب شده است کسب‌وکارهای برخط بتوانند به مجموعهٔ گسترده‌ای از اطلاعات که از طریق تعاملات مجازی یا انواع گوناگون تراکنش مشتریان قابل استخراج است، دست یابند؛ این اطلاعات عبارت‌اند از: دفعات تکرار خرید مشتری و وفاداریش، زمان شروع یا پایان مشاهدهٔ وب‌سایت فروشگاه، نوع کالاهایی که قبلاً کاربر پسندیده، زمان خرید کالا توسط کاربر و همچنین اطلاعاتی مانند رتبه‌بندی، پیشینهٔ خرید و اطلاعات کالاها. ایواتا در سال ۲۰۰۸ نشان داد فروشگاه با درک علاقه و سلیقهٔ اخیر کاربران، می‌تواند پیشنهادهای مؤثرتری ارائه دهد؛ زیرا علایق و سلایق کاربران روزبه‌روز تغییر می‌کند. با پیشنهاد محصولاتی که اغلب توسط کاربران وفادار خریداری شده است، فروشگاه می‌تواند فروش را افزایش دهد؛ چرا که به شمار کاربران وفادار افزوده می‌شود. همچنین فروشندگان می‌توانند در تعیین قیمت محصولات انعطاف‌پذیری بیشتری داشته باشند (گوپال، تریپادی و والتر، ۲۰۰۶).

در پژوهشی دیگر که کی‌پور و همکارانش با عنوان «ارائهٔ روشی جدید برای پیشگویی پیوند بین رأس‌های موجود در شبکه‌های اجتماعی» انجام دادند، دو رویکرد سراسری و محلی پیمایش گراف شبکه را پیشنهاد دادند و با توجه به مطالعات خود بر دو شبکهٔ اجتماعی فیس‌بوک و اپ. نیونز و استفاده از معیار آدامیک ـ آدار، به این نتیجه رسیدند که رویکرد محلی به‌دلیل محلی‌بودن می‌تواند پیشگویی خوبی برای یال‌هایی انجام دهد که قرار است در آینده شکل بگیرند (کی‌پور، براری و شیرازی، ۱۳۹۳).

کرامتی و خالقی (۱۳۹۳) در مقالهٔ خود با استفاده از شیوه‌های داده‌کاوی، به ارائهٔ مدلی برای توسعهٔ سیستم‌های پیشنهاددهندهٔ محصول به مشتریان در سطح خرده‌فروشی اقدام کردند. در این مقاله با استناد به چارچوب پیشنهادی مدل، نخست مشتریان با تکیه بر رویکرد بخش‌بندی مبتنی بر ارزش طول عمر و با لحاظ نسبی ترجیحات، بر اساس مشخصه‌های مدل RFM، خوشه‌بندی شدند؛ سپس با بهره‌مندی از ساختار پیشنهاددهی دومرحله‌ای، پیشنهادهای گوناگونی در دو سطح متمایز از رده‌بندی محصول به هر یک از مشتریان هدف ارائه دادند. نتیجهٔ این پژوهش نشان داد استفاده از مدل‌های ترکیبی در مقایسه با مدل‌های مشابه سنتی، موجب کاهش مشکلات عدم تراکم و مقیاس‌پذیری می‌شود و عملکرد بهتری دارد.

دسته‌ای از پژوهش‌ها نیز بر مبنای اطلاعات زمانی اجرا شده‌اند؛ برای نمونه، لی، پارک و پارک (۲۰۰۸) در پژوهشی با استفاده از دو شاخص زمان خرید کاربر و زمان عرضهٔ آیتم که دقت توصیه‌کننده‌ها را افزایش می‌دهند، دو تابع رتبه‌بندی برای محاسبهٔ وزن مبتنی بر اطلاعات زمانی پیشنهاد دادند. آنها پیش‌تر یک نوع اطلاعات زمانی شامل زمان عرضهٔ آیتم و زمان خرید کاربر و

تفاوت زمانی بین این دو را بررسی کردند و نتیجه گرفتند این‌گونه اطلاعات زمانی می‌تواند دقت توصیه‌گرها را در سیستم‌های توصیه‌گر مبتنی بر فیلتر همبستگی برای کاراکترهای شرح داده‌شده در محیط تجارت الکترونیک، افزایش دهد.

کریمی، عسگری و پیراسته (۱۳۹۴) با توجه به اطلاعات جمعیت‌شناختی جامعهٔ آماری و خوشه‌بندی و همچنین با استفاده از پالایش همکارانهٔ مبتنی بر کاربر، به ارائهٔ الگوهایی برای پیشنهادهای نزدیک‌تر به سلیقهٔ مشتریان در فروشگاه‌های آنلاین و فیلم پرداختند. آنها بر این عقیده بودند که استفاده از یکی از روش‌ها به تنهایی نمی‌تواند پیشنهادهایی نزدیک به نظر کاربران ارائه کند و در کنار پالایش همکارانه، اطلاعات جمعیت‌شناختی هم تأثیر شایان توجهی در شناسایی اولویت‌های کاربران دارد.

در پژوهشی دیگر، حسنقلی‌پور و همکارانش در مقاله‌ای با عنوان «بررسی تأثیر خصوصیات مشتریان بر تمایل آنها به پذیرش خرید اینترنتی» نشان دادند ویژگی‌های مصرف‌کنندگان، یکی از مهم‌ترین عوامل پذیرش خرید اینترنتی آنان است. روش استفاده‌شده در پژوهش، توصیفی از نوع همبستگی بود و داده‌ها از طریق توزیع پرسشنامه بین ۲۳۱ نفر از دانشجویان دانشگاه تهران گردآوری شد. نتایج تحلیل رگرسیون چندگانه نشان داد ویژگی‌های مصرف‌کنندگان بر اساس کالاها و خدمات متفاوت، تأثیر متفاوتی دارد. پذیرش نوآوری در حوزهٔ فناوری اطلاعات، خودکفایتی اینترنتی، ادراک از امنیت وب، نگرانی‌های حریم خصوصی و درخور بودن محصول، ویژگی‌هایی هستند که بر اساس نوع کالا یا خدمت، بر نگرش افراد به پذیرش خرید اینترنتی آنان تأثیرگذارند (حسنقلی‌پور، امیری، فهیم و قادری عابد، ۱۳۹۲).

از میان مطالعات انجام‌شده در این حوزه، تنها اندکی از آنها بر مبنای اطلاعات زمانی بوده‌اند؛ مثل زمان خرید کاربر و زمان عرضه که دقت توصیه‌کننده‌ها را افزایش می‌دهند. از یافته‌های این پژوهش‌ها می‌توان به این نتیجه رسید که اطلاعات زمانی می‌تواند دقت پیشنهاددهنده‌ها را در سیستم‌های پیشنهاددهندهٔ مبتنی بر پالایش مشارکتی، برای ویژگی‌های شرح داده‌شده در محیط تجارت الکترونیک سیار بهبود بخشد؛ ولی در مقالات بررسی‌شده، از واحد زمان و تأثیر آن در معیارهای شباهت‌یابی و تأثیر روابط معنایی در شبکه‌های اجتماعی استفاده نشده که مقالهٔ حاضر بر این موضوع تمرکز کرده است. با توجه به توضیحاتی که بیان شد، می‌توان گفت هدف اصلی در این مقاله، ارائهٔ روشی جدید برای مشابهت‌یابی در سیستم‌های پیشنهاددهنده و بررسی و تشخیص میزان علاقهٔ کاربران مشابه کاربر هدف با معیار زمان است. همچنین، ویژگی‌هایی که می‌تواند در افزایش دقتِ سیستم‌های پیشنهاددهنده اهمیت داشته باشد و مطالعات دیگر کمتر به آنها توجه کرده‌اند را بررسی می‌کند. از این رو، بررسی و تشخیص میزان تشابه آیتم هدف با آیتم‌هایی که

کاربر هدف از قبل به آنها نرخ داده با استفاده از معیار زمان، یکی از اهداف این پژوهش است. در این پژوهش تلاش می‌شود تأثیر افزایش دقت و زمان بر بهبود کلی عملکرد پیشنهادها و کاهش تُنُکی داده و تأثیر تکنیک ادغام اطمینان در سیستم پیشنهاددهنده بر تُنُکی داده‌ها مشخص شود. در این رابطه به‌منظور افزایش دقت، تأثیر معیار زمان بر تشخیص میزان تشابه آیتم هدف با آیتم‌هایی که کاربر هدف از قبل به آنها نرخ داده و تأثیر معیار زمان بر تشخیص میزان علاقهٔ کاربران مشابه کاربر هدف با درنظر گرفتن ویژگی‌های جانبی که کمتر به آنها توجه شده، در روش پیشنهادی دخالت داده شده است تا مشخص شود، تکنیک ادغام اطمینان بر کمتر شدن تُنُکی داده‌ها مؤثر است.

روش‌شناسی پژوهش

در این پژوهش هم از سیستم‌های پیشنهاددهندهٔ پالایش مشارکتی مبتنی بر حافظه و هم مبتنی بر مدل استفاده شده است. در بخشی که مبتنی بر مدل[1] است، از روش پالایش مشارکتی پیشنهادی ادغام اعتماد صریح و پالایش مشارکتی کاربرمحور بهبودیافته استفاده می‌شود که برای حل مشکل کاربر جدید در خصوص کاربران و اقلام پیشنهاد داده شده به‌کار می‌رود. به‌منظور حل مشکل تُنُکی داده در خصوص کاربران و اقلام نیز از روش مبتنی بر حافظهٔ[2] پالایش مشارکتی استفاده می‌شود.

در روش اول، شاخص زمان با ضریب پیرسونی که مبتنی بر حافظهٔ پالایش مشارکتی سیستم‌های پیشنهاددهنده است (در این بخش نیز از زمان‌های نرخ‌دهی کاربران به آیتم‌ها بهره استفاده شده)، ادغام شده‌اند. سیستم‌های پیشنهاددهنده به‌طور وسیعی در سایت‌های تجارت الکترونیک استفاده می‌شوند. هدف از طراحی این سیستم‌ها، کمک به کاربران برای یافتن اقلام مدنظرشان است. یکی از موفق‌ترین و کاراترین الگوریتم‌های ایجاد پیشنهاد، روش پالایش گروهی است. پالایش گروهی به‌منظور پیشنهاد یک کالا به کاربر هدف، از سوابق امتیازدهی کاربران مشابه به آن کالا استفاده می‌کند. پالایش گروهی ضمن برخورداری از مزیت‌های بسیار و کارایی زیاد، چندین مشکل عمده دارد. از جمله محدودیت‌های پالایش گروهی می‌توان به مشکل کاربر جدید و تُنُک‌بودن ماتریس رتبه‌دهی اشاره کرد. کاربر جدید، به‌معنای ورود محصول جدید به سیستم است که هنوز هیچ رتبه‌ای دریافت نکرده یا به‌طور کلی این محصول در نظر گرفته نشده است. منظور از خلوت‌بودن یا تُنُکی داده این است که در ماتریس قلم ـ کاربر، تعداد

1. Model-base
2. Memory-base

رتبه‌هایی که وجود دارد نسبت به کل رتبه‌هایی که می‌تواند وجود داشته باشد بسیار کم است؛ به بیان ساده‌تر، نسبت این دو مقوله به هم اساساً منطقی نیست. در این پـژوهش، بـه‌منظـور بهبـود سیستم‌های پیشنهاددهنده در موقعیتی که با مشکلات یادشده مواجه‌انـد، از داده‌هـای موجـود در شبکه‌های اجتماعی استفاده شده است. اطلاعات موجود در شبکه‌های اجتماعی، شـامل پروفایـل کاربران و روابـط بـین آنهاسـت. در روش پیشـنهادی اول، ابتـدا کـاربران براسـاس دو مـاتریس رتبه‌دهی و زمان رتبه‌دهی خوشه‌بندی می‌شوند؛ سپس به‌منظور یافتن نزدیک‌تـرین همسـایه‌هـا، اطلاعات مربوط به محتوای اقلام و ماتریس‌های جدید ترکیب خواهند شد.

روش پیشنهادشدهٔ اول

در روش پیشنهادشده، استراتژی مبتنی بر زمان و نرخ‌دهی مدل‌های کلاسیک و مدل‌های پیشین مقایسه شدند و مشخص شد در حالت‌های قبلی، تنها یک ماتریس به نام مـاتریس کاربرــ آیـتم وجود دارد، اما در اینجا از دو ماتریس کاربرـ آیتم و کاربرـ زمان استفاده شده است کـه مـاتریس کاربرـ آیتم نرخ‌دهی و ماتریس کاربرـ زمان تاریخ نرخ‌دهی را نشان می‌دهد.

مدل کاری مبتنی بر زمان

مراحل تولید پیشنهاد توسط سیستم پیشنهاددهندهٔ مبتنی بـر زمـان در شـبکه‌هـای اجتمـاعی را می‌توان به سه مرحلهٔ زیر دسته‌بندی کرد:

مرحلهٔ نخست

در این مرحله با توجه به اطلاعات کاربران و زمان امتیازدهی هـر منبـع از نظـر کـاربر، مـاتریس کاربرـ منبع شکل می‌گیرد. ماتریس رتبه‌بندی که ماتریسی $m \times n$ است، ارتبـاط بـین منـابع و کاربران را نشان می‌دهد و در آن m تعداد کاربران و n تعداد منابع را مشخص می‌کنـد. مقـداری که در سلول $i \times j$ ماتریس قرار می‌گیرد، عددی است که میـزان علاقـهٔ کـاربر i بـه منبـع j را نشان می‌دهد. برای تشکیل ماتریس رتبه‌بنـدی بایـد ابتـدا مـاتریس منبـع ـ کـاربر هـر یـک از استراتژی‌های مبتنی بر زمان تشکیل شود.

مرحلهٔ دوم

در این مرحله، میزان شباهت بین کاربر فعال و سایر کـاربران کـه همسـایه‌هـای آن را تشـکیل می‌دهند، به‌دست می‌آید که برای این کار از ضریب شباهت پیرسونی استفاده مـی‌شـود. یکـی از انواع روش پالایش مشارکتی، پژوهش همبستگی است. در این نوع پژوهش رابطهٔ میان متغیرهـا براساس هدف پژوهش تحلیل می‌شود و اصولاً هدف پژوهش، تعیین رابطه میان متغیرهاست.

مرحلهٔ سوم

در این مرحله که آخرین مرحلهٔ کار محسوب می‌شود، باید منابعی را به کاربر فعال پیشنهاد داد. برای این منظور از (رابطهٔ ۱) استفاده می‌شود.

$$Score\ (u,r) = \frac{\sum v \in neighbor(u) M_{v,r} \times cov\ (u,v) \times T}{|\sum v \in neighbor\ cov\ (u,v) \times T|}$$ رابطهٔ ۱)

در این رابطه، V همسایه‌های کاربر فعال u است و $cov\ (u,v) \times T$ هممیزان شباهت بین کاربر u وv را با تأثیر زمان مشخص می‌کند.

روش پیشنهادشدهٔ دوم

روش پیشنهادشدهٔ دوم در این پژوهش از دو جزء کاربرمحور و آیتم‌محور تشکیل می‌شود. به بیان دیگر، در این روش با توجه به موقعیت و خصوصیات کاربر هدف و آیتم هدف، از طریق دو الگوریتم کاربرمحور و آیتم‌محور به پیش‌بینی نرخ پرداخته می‌شود. به‌طور مشخص، در سیستم توصیه‌گر نهایی که قرار است از این الگوریتم‌های پیشنهادشده به‌منظور پیش‌بینی نرخ‌های کاربر هدف به آیتم‌ها استفاده کند، این دو الگوریتم با هم ادغام خواهند شد که طریقهٔ ادغام در رابطهٔ ۲ مشاهده می‌شود. در این رابطه مشخص است که در صورت صفر نبودن هر دو نرخ پیش‌بینی‌شده توسط دو الگوریتم نام‌برده، میانگین سازگاری این دو نرخ به‌عنوان نرخ نهایی پیش‌بینی‌شده در نظر گرفته می‌شود. دلیل انتخاب این نوع میانگین این است که دو الگوریتم پیشنهادشده، موازی هم عمل می‌کنند؛ بنابراین میانگین سازگاری آنها بهترین گزینه برای ادغام نتایجشان است.

رابطهٔ ۲)

$$Pred_{u,b} = \begin{cases} -1 if\ P_{u,b}^{EpTeUCF} = 0\ and\ P_{u,b}^{CbShoShcSeICF} = 0 \\ P_{u,b}^{EpTeUCF}\ if\ P_{u,b}^{EpTeUCF} \neq 0\ and\ P_{u,b}^{CbShoShcSeICF} = 0 \\ P_{u,b}^{CbShoShcSeICF}\ if\ P_{u,b}^{EpTeUCF} = 0\ and\ P_{u,b}^{CbShoShcSeICF} \neq 0 \\ \frac{2 \times P_{u,b}^{EpTeUCF} \times P_{u,b}^{CbShoShcSeICF}}{P_{u,b}^{EpTeUCF} + P_{u,b}^{CbShoShcSeICF}}\ if\ P_{u,b}^{EpTeUCF} \neq 0\ and\ P_{u,b}^{CbShoShcSeICF} \neq 0 \end{cases}$$

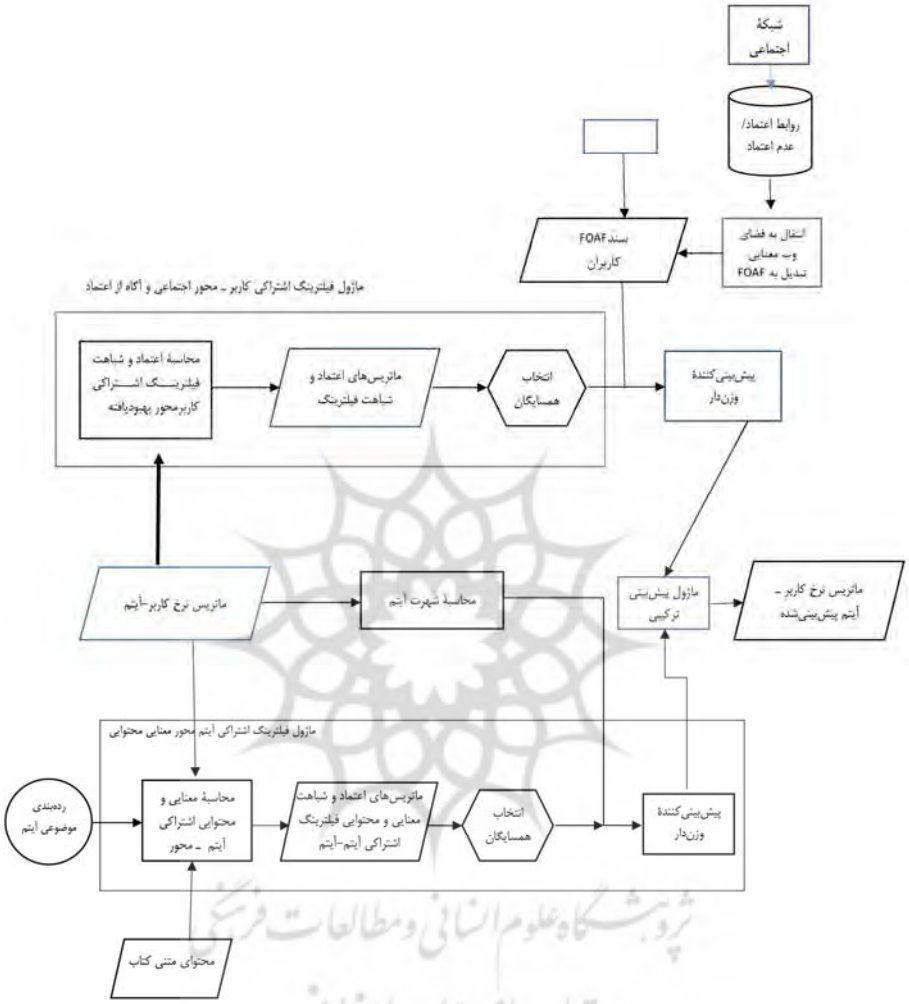
با توجه به این رابطه می‌توان به ساختار سیستم توصیه‌گر نهایی و روند پیش‌بینی نرخ در آن پی برد. این ساختار نوعی روش ترکیبی را نشان می‌دهد که در آن نرخ‌های پیش‌بینی‌شده توسط دو الگوریتم پیشنهادی با هم ادغام می‌شوند تا نتیجهٔ بهتری در روند کاهش خطای سیستم

کسب شود. شکل ۱، مراحل پیش‌بینی نرخ در سیستم توصیه‌گر پیشنهادشده را نمایش می‌دهد. در این شکل دو ماژول زیر مشخص شده است:

- ماژول فیلترینگ همکارانه (اشتراکی) کاربرمحور اجتماعی و آگاه از اعتماد؛ این ماژول همان الگوریتم کاربرمحور پیشنهادی با نام EpTeUCF، است.
- ماژول فیلترینگ همکارانه (اشتراکی) آیتم‌محور معنایی و محتوایی؛ این ماژول نیز در واقع الگوریتم آیتم‌محور پیشنهادی است که با نام CbShoShcSeICF در ارزیابی‌ها مشخص می‌شود.

روش کاربرمحور پیشنهادشده در این پژوهش، از دو طریق به محاسبهٔ فهرست کاربران مشابه با کاربر هدف می‌پردازد. در روش نخست، شباهت فیلترینگ همکارانهٔ بهبودیافته با استفاده از داده‌های نرخ‌دهی کاربران به آیتم‌ها محاسبه می‌شود. روش دوم نیز تشخیص کاربران مورد اعتماد کاربر هدف بر اساس روابط موجود در شبکهٔ اعتماد کاربران یا اسناد FOAF آنان است. گفتنی است در این‌جا منظور از سند FOAF کاربری، سند روابط معمولی کاربران نیست؛ بلکه روابط اعتماد به نظر کاربران دیگر است که خود کاربران به صراحت در شبکهٔ اجتماعی بیان کرده‌اند. روش آیتم‌محور پیشنهادی نیز از طریق سه رویکرد به محاسبهٔ شباهت بین آیتم هدف با سایر آیتم‌ها می‌پردازد و در نهایت با ادغام موازی این سه رویکرد، آیتم‌هایی که بیشترین شباهت را دارند (یا همان آیتم‌های همسایهٔ آیتم هدف)، تشخیص می‌دهد. رویکرد نخست، محاسبهٔ شباهت معنایی بین آیتم‌ها از طریق اندازه‌گیری مشابهت محتوایی آنهاست. رویکرد دوم، محاسبهٔ شباهت بین آیتم هدف با سایر آیتم‌های نرخ داده‌شده توسط کاربر هدف، از طریق اندازه‌گیری بیشترین همپوشانی موجود بین سر موضوعاتی است که دو آیتم به آنها تعلق دارد و رویکرد سوم، محاسبهٔ شباهت بین دو آیتم از طریق تعداد سر موضوعات مشترک آنهاست. ورودی اصلی روش دوم و سوم، طبقه‌بندی موضوعی آیتم‌هاست که حاوی فهرست سلسله‌مراتبی[1] سرموضوعاتی است که آیتم‌ها به آنها تعلق دارند. شهرت آیتم‌های همسایهٔ آیتم هدف نیز از روی داده‌های نرخ‌دهی کاربران به آیتم‌ها محاسبه می‌شود و علاوه‌بر نرخ کاربر هدف به هر آیتم همسایه، وزن دوم پیش‌بینی نرخ احتمالی کاربر هدف به آیتم هدف، محسوب می‌شود. در پایان نیز دو نرخ پیش‌بینی‌شده از دو روش پیشنهادی فوق، با هم ترکیب شده (با استفاده از رابطهٔ ۲) و نرخ پیش‌بینی‌شدهٔ سیستم توصیه‌گر پیشنهادی به‌دست می‌آید. حال از این نرخ می‌توان آیتم‌های احتمالی مورد علاقهٔ کاربر هدف (آیتم‌های با بیشترین نرخ پیش‌بینی‌شدهٔ احتمالی) را تشخیص داد و به ساخت مجموعهٔ توصیه‌های کاربر هدف پرداخت.

1. Hierarchical



شکل ۱. ساختار کلی روند پیش‌بینی نرخ در سیستم توصیه‌گر پیشنهادی

یافته‌های پژوهشی

تجزیه و تحلیل داده‌ها در روش پیشنهادشدهٔ اول

در حوزهٔ داده‌کاوی و الگوریتم‌های یادگیری، بهره‌مندی از مجموعهٔ دادهٔ مناسب، اهمیت زیادی دارد. مطالعهٔ مجموعهٔ دادهٔ انتخاب‌شده از نظر کمی و کیفی و نیز، انتخاب ویژگی‌های مناسب برای

توصیف نمونه‌ها، یکی از گام‌های ضروری و مفصل در این حوزه محسوب می‌شود. مجموعهٔ دادهٔ استفاده‌شده برای تحلیل نتایج این پژوهش، علاقهٔ کاربران به مجموعه‌ای از فیلم‌هاست. این مجموعهٔ داده، شامل اطلاعات فردی کاربران، فیلم‌ها، انواع و تعداد آنها و اطلاعات امتیازدهی است. اطلاعات مجموعهٔ داده در جدول ۱ آورده شده است. مجموعهٔ داده‌های مووی لنز[1] حاوی فایل فشردهٔ tar. برای بازسازی فایل‌های داده است و همچنین دربردارندهٔ مجموعهٔ کاملی از داده و نیز ۱۰۰,۰۰۰ رتبه‌بندی توسط ۹۴۳ کاربر در ۱۶۸۲ اقلام است. این فایل از نظر نرخ‌دهی کاربران، رتبهٔ نسبتاً خوبی دارد؛ به‌طوری که در این فایل هر کاربر حداقل به ۲۰ فیلم امتیاز داده است. این فایل همچنین بخش جداگانه‌ای دارد که شامل فهرستی از شناسهٔ کاربر، کد آیتم، امتیاز، نرخ زمان می‌شود. جدول ۱ نمونه‌ای از اطلاعات ۴ کاربر و زمان نرخ‌دهی را نشان می‌دهد که از مجموعهٔ دادهٔ مووی لنز ارزیابی و سنجیده شده است. در این جدول، زمانی که هر کاربر برای نرخ‌دهی صرف کرده، در یک دورهٔ زمانی سه‌ماهه در نظر گرفته شده است و ارزیابی علایق کاربران در زمان‌های متفاوت نسبت به اقلامی که نرخ‌دهی شده‌اند، نشان داده است.

جدول ۱. اطلاعات مربوط به کاربر ـ منبع

کاربر ۱	زمان	۱۹۹۷/۵/۱۲		۱۹۹۷/۶ /۱۸	۱۹۹۸/۱/۰۳
	رتبه	۳		۱	۵
کاربر ۲	زمان		۱۹۹۷ /۷/۱۲		
	رتبه		۴		
کاربر ۳	زمان	۱۹۹۷/۵/۱۹		۱۹۹۷/۱۱ /۲	
	رتبه	۳		۲	
کاربر ٤	زمان	۱۹۹۷/۱۰/۵	۱۹۹۷/۱۱ /۸		۱۹۹۸ /۲ /۷
	رتبه	۵	۲		۳

ارزیابی روش پیشنهادی اول

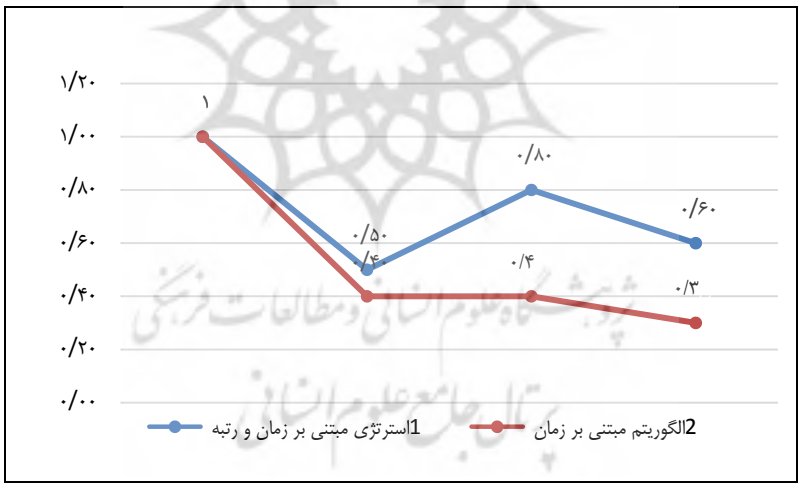
همان‌طور که می‌دانیم، دو پارامتر دقت[1] و پوشش در میزان کیفیت اطلاعات استخراجی بسیار مؤثرند. از پارامتر دقت و پوشش که هر دو معیار از شناخته‌شده‌ترین معیارهای رده‌بندی هستند،

1. Movie lens
1. Precision
2. Coverage

عموماً برای اندازه‌گیری کیفیت اعمال استخراج اطلاعات استفاده می‌شود؛ رابطهٔ دقت، بیان می‌کند که چند درصد از مجموعه پیشنهادهای ارائه‌شده، واقعاً درست هستند. این رابطه در واقع درستی و دقت پیشنهادهای ارائه‌شده توسط سیستم را می‌سنجد؛ بنابراین هرچه مقدار این معیار بیشتر باشد، نشان‌دهندهٔ تعداد کم اشتباه‌های محاسباتی سیستم است. این معیار معمولاً مستقل از رابط کاربری است و می‌تواند به صورت برون‌خط هم محاسبه شود (رابطهٔ ۳ را مشاهده کنید).

$$precision = \frac{|recomended\ items \cap favorite\ item|}{recomended\ items}$$ رابطهٔ ۳)

با محاسبهٔ دقت و درستی پیشنهادها و استراتژی مبتنی بر زمان و رتبه و الگوریتم مبتنی بر زمان، می‌توان نمودار مقایسه‌ای میزان پارامتر دقت با توجه به تعداد منابع پیشنهادی را ترسیم کرد. مطابق شکل ۲ که در زیر مشاهده می‌شود، دقت و درستی استراتژی مبتنی بر زمان و رتبه در مقایسه با الگوریتم مبتنی بر زمان، مقدار بالاتری دارد. در واقع، هرچه مقدار به‌دست‌آمده از این محاسبه بیشتر باشد، از دقت و درستی بیشتری برخوردار است؛ یعنی دو آیتم زمان و رتبه‌بندی به مراتب بر دقت کار سیستم می‌افزایند و خطاهای احتمالی را به میزان شایان توجهی کاهش می‌دهند.



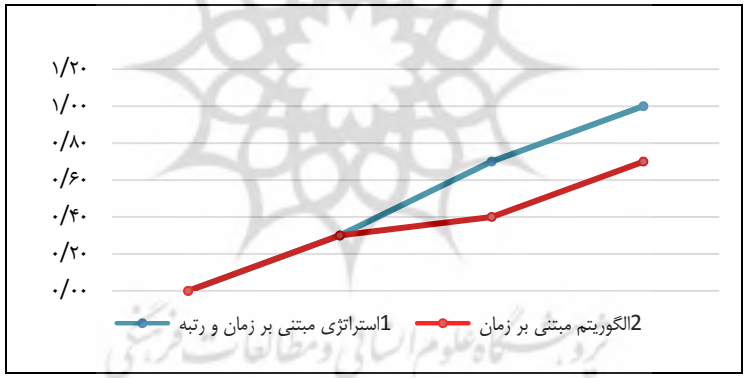
شکل ۲. مقایسهٔ میزان پارامتر دقت با توجه به تعداد منابع پیشنهادی

اکنون می‌خواهیم به بحث دربارهٔ پارامتر پوشش بپردازیم و بیان کنیم که پوشش اصولاً چیست و چه رابطه‌ای با سایر آیتم‌هایی دارد که در الگوریتم یادشده لحاظ شده است. پوشش به این معناست که چند درصد از آیتم‌هایی که واقعاً مورد علاقهٔ کاربر است، توسط سیستم به او پیشنهاد

شده است. در این دو رابطه، منظور از آیتم‌های مورد علاقه[1]، مجموعه‌ای است که کاربر به آن علاقه دارد و در واقع باید به کاربر پیشنهاد داده شود و موارد پیشنهادی[2] نیز مجموعهٔ آیتم‌هایی است که توسط سیستم به کاربر توصیه شده است.

$$covarage = \frac{|recomended\ items \cap favorite\ items|}{favorite\ items}$$ رابطهٔ ۴)

با توجه به رابطهٔ ۴ که پوشش پیشنهادها را محاسبه می‌کند و نیز، استراتژی مبتنی بر زمان و رتبه و الگوریتم مبتنی بر زمان، می‌توان نمودار مقایسهٔ میزان پارامتر پوشش را با توجه به تعداد منابع پیشنهادی ترسیم کرد (شکل ۳). همان‌طور که مشاهده می‌شود، خطوط آبی نسبت به خطوط قرمز، در جایگاه بالاتری قرار دارند و این امر نشان‌دهندهٔ این است که پوشش استراتژی مبتنی بر زمان و رتبه در مقایسه با الگوریتم مبتنی بر زمان، مقدار بیشتری دارد؛ هر چه مقدار این محاسبه بیشتر باشد، نشان‌دهندهٔ پوشش بیشتر است و اقلام مورد علاقهٔ بیشتری به کاربر پیشنهاد شده است.



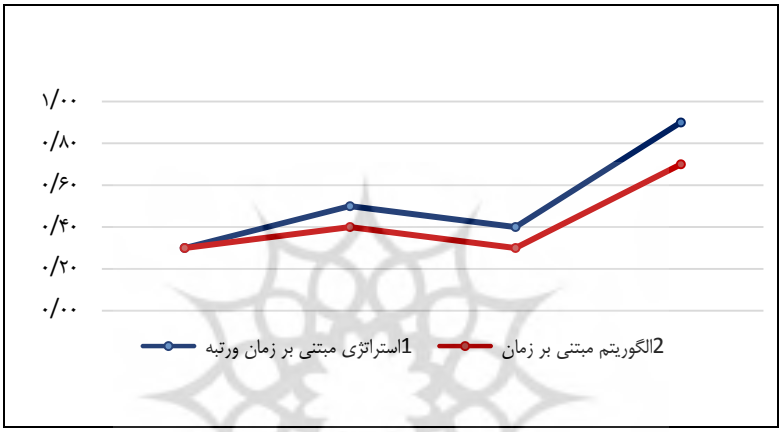
شکل ۳. مقایسهٔ میزان پارامتر پوشش باتوجه به تعداد منابع پیشنهادی

تا این‌جا دیدیم که این معیارها به‌طور موازی عمل می‌کنند و بدین ترتیب نمی‌توان آنها را جمع‌بندی کرد. به‌دلیل عملکرد موازی این معیارها، برای تجمیع آنها باید از معیاری به نام $F_{measure}$ استفاده کرد. این معیار میانگین سازگاری دو معیار دقت و پوشش است که در قالب رابطهٔ ۵ بیان شده است.

1. Favorite items
2. Recommended items

$$F_{measure} = \frac{2 \times (covarage \times precision)}{covarage + precision}$$ رابطهٔ ۵)

با توجه به رابطهٔ ۵ که به‌منظور محاسبهٔ میانگین سازگاری دو معیار دقت و پوشش پیشنهادها طراحی شده است و استراتژی مبتنی بر زمان و رتبه و الگوریتم مبتنی بر زمان، می‌توان شکل ۴ را ترسیم کرد که به مقایسه میزان پارامتر F۱ با توجه به تعداد منابع پیشنهادی می‌پردازد.



شکل ٤. مقایسهٔ میزان پارامتر F۱ باتوجه به تعداد منابع پیشنهادی

ارزیابی روش‌های تأثیر زمان در معیار شباهت‌یابی پیرسون

در این روش، برای ساخت الگوریتم جدید از شباهت پیرسون استفاده شده است؛ زیرا نسبت به روش‌های دیگر مثل کوساین و... مزیت‌هایی دارد؛ برای مثال نرمالایز می‌شود و نسبت به نویز مقاومت بیشتری نشان می‌دهد. در شکل ۵ خط قرمز بیان‌کنندهٔ پیرسون ادغام‌شده با زمان است؛ خط آبی KNN را نشان می‌دهد که به‌وسیلهٔ زمان تغییر کرده است؛ خط سبز معرف تغییرات الگوریتم KNN است؛ خط بنفش کمرنگ پیرسون کامل را نشان می‌دهد. در واقع این نمودار به مقایسهٔ الگوریتم‌های روش تأثیر زمان و مقایسهٔ الگوریتم‌های KNN، پیرسون کامل، پیرسون ادغام با زمان و KNN که به‌وسیلهٔ زمان تغییر یافته، پرداخته است. همان‌طور که مشاهده می‌شود، KNN در سه حالت۱. ساده‌شده؛ ۲. پیچیدگی‌های اضافه که زمان در آن تأثیر داده‌شده و ۳. در مدلی که بهترین خطا به‌کمک آن به‌دست آمده، بررسی شده است. خروجی این سه الگوریتم در دو ماتریس ترکیب‌شده و برای پیش‌بینی استفاده می‌شود.

شکل ۵. مقایسهٔ الگوریتم‌ها با روش تأثیر زمان

روش کار بدین صورت است که ابتدا با استفاده از ۲۰درصد دادهٔ آزمایشی که در هر جدول وجود دارد، پیش‌بینی آغاز می‌شود؛ سپس جدولی ساخته می‌شود که ورودی آن شامل رگرسیون به‌دست‌آمده از ماتریس ترکیبی سه الگوریتم است، در اینجا از ضرایبی که در دادهٔ آموزش به‌دست آمده است، برای پیش‌بینی استفاده می‌شود. سپس با این روش، شروع به پیش‌بینی کرده و از پیش‌بینی‌ها روش محاسبهٔ RMSE به‌دست می‌آید که در شکل ۵ نشان داده شده است. همان‌طور که در شکل مشاهده می‌شود، RMSE آن نسبت به بقیهٔ حالت‌ها کمتر است.

تجزیه و تحلیل داده‌ها در روش پیشنهادی دوم

مجموعه دادهٔ به‌کاررفته در ارزیابی‌ها

در ارزیابی‌های آفلاین، از مجموعه دادهٔ اپینیونز[1] استفاده کردیم. ابتدا قصد استفاده از مجموعه دادهٔ جمع‌آوری‌شدهٔ پاولو ماسا را داشتیم، اما در این پژوهش از سرموضوع[2] کتاب هم استفاده شده است. برای دسترسی به سرموضوع هر کتاب، باید بین مجموعه دادهٔ اپینیونز با محیط داده‌های بیرونی آن، از جمله طبقه‌بندی موضوعی کتابخانهٔ کنگره و آمازون، ارتباطی برقرار شود و برای این کار به شمارهٔ کتاب سریال یکتای جهانی[3] نیاز بود. از سویی مجموعه دادهٔ یادشده، برای کتاب‌ها شمارهٔ

1. Epinions
2. Subject Heading
3. International Serial Book Number (ISBN)

مستقلی دارد. در نتیجه تصمیم به خزش[1] دوباره در سایت اپینیونز گرفته شد و طی پنجاه روز اطلاعات کاربران و کتاب‌ها از قسمت کتب سایت اپینیونز جمع‌آوری شدند. این مجموعه دادهٔ جدید، مزیت‌های بسیار زیادی را نسبت به نمونهٔ قبلی دارد.

ارزیابی روش‌های پیشنهادی دوم

در این بخش به نمایش نتایج ارزیابی‌ها از الگوریتم‌های پیشنهادی در مقایسه با الگوریتم‌های رقیب می‌پردازیم. ابتدا نتایج ارزیابی الگوریتم کاربرمحور پیشنهادی را بررسی می‌کنیم و پس از آن نوبت به الگوریتم آیتم‌محور پیشنهادی خواهد رسید.

معرفی الگوریتم‌های کاربرمحور رقیب

در این بخش به ارزیابی تعدادی از الگوریتم‌های رقیب در مقایسه با الگوریتم‌های کاربرمحور پیشنهادی پرداخته می‌شود. فیلترینگ همکارانهٔ ساده[2]، الگوریتم پایهٔ معروف و مناسبی برای تشکیل مجموعه توصیه‌های سیستم توصیه‌گر است. فیلترینگ همکارانهٔ ساده برای نمایش برتری سیستم‌های توصیه‌گر معنایی که بر خلاف سیستم‌های معمولی از دانش خارج از خود نیز برای بهبود بهره می‌برند، استفاده می‌شود. الگوریتمی با نام TeCF پیشنهاد شده که از نوع ترکیبی است (شامبور و لو، ۲۰۱۲). این الگوریتم از روابط اعتماد ضمنی بین کاربران و همچنین از فیلترینگ همکارانهٔ بهبودیافتهٔ کاربری، به‌طور ترکیبی استفاده می‌کند. این الگوریتم در نمودارهای زیر با نام اختصاری ImpTeCF شناخته می‌شود که همانند روش TeCF نوعی روش ترکیبی است و از روابط اعتماد صریح بین کاربران استفاده می‌کند. این روش پیشنهادی در مقالهٔ حاضر، ادغامی از روابط اعتمادهای صریح بین کاربران و روش فیلترینگ همکارانهٔ بهبودیافته کاربری است و با نام اختصاری EpTeCF در نمودارهای زیر مشخص شده است.

یکی از رقبای الگوریتم پیشنهادی این پژوهش، الگوریتم اولیهٔ (ICF) در نظر گرفته شده است. این الگوریتم که توسط ساروار، کاریپیس، کنستاین و رید (۲۰۰۱) معرفی شده است، با نام Sarwar-ICF در نمودارهای ارزیابی ظاهر می‌شود (کاندلیر، فسانت و مایر، ۲۰۰۸). به‌دلیل معرفی اولیهٔ این الگوریتم توسط آقای رسنیک این الگوریتم در نمودارهای زیر با نام UCF[3] شناخته می‌شود. جزئیات این الگوریتم پایه، ضریب پیرسون شباهت بین دو کاربر را بر اساس تغییرات نرخ‌دهی آنان فقط روی آیتم‌های مشترک اندازه می‌گیرد و این یک ضعف اساسی محسوب

1. Crawl
2. Simple Collaborative Filtering (sCF)
3. User Collaborative Filtering (UCF)

می‌شود؛ زیرا دو کاربر که تنها در نرخ‌دهی یکسان به یک آیتم اشتراک دارند، می‌توانند به‌طور کلی مشابه تشخیص داده شوند (رسنیک و همکاران، ۱۹۹۴). الگوریتم آیتم‌محور، نام ShoShcSeICF [1] را به خود گرفته است. روش کار در آن شبیه الگوریتم SeCF است با این تفاوت که مشابهت‌های معنایی در این‌جا علاوه‌بر مقایسهٔ تعداد اعضای مجموعهٔ سرموضوع کتاب‌ها، با استفاده از حداکثر میزان هم‌پوشانی سرموضوعات نیز محاسبه می‌شود (شامبور و لو، ۲۰۱۲).

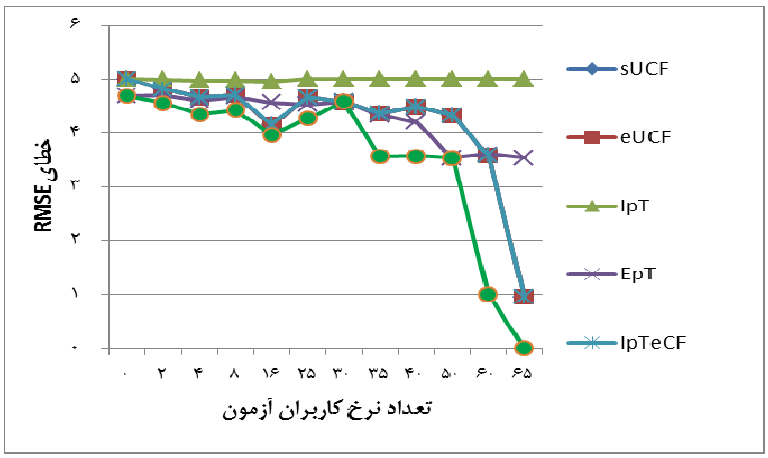
با ادغام معیار فیلترینگ همکارانهٔ بهبودیافته و معیار شباهت معنایی هم‌پوشانی سرموضوعات برای دو آیتم و درکنار محبوبیت آیتم‌های همسایه به‌عنوان وزن شباهت هر آیتم، می‌توان از تأثیر هر دو معیار در روند پیش‌بینی نرخ کاربر هدف به آیتم هدف استفاده کرد. این ادغام از طریق رابطهٔ ۱ انجام می‌شود. در این رابطه، $ShoSSim_{i,j}$ حداکثر میزان هم‌پوشانی سرموضوعات آیتم همسایهٔ j و آیتم هدف i، $eICF_{i,j}$ میزان شباهت کسینوسی در فیلترینگ همکارانهٔ دو آیتم، IR_j محبوبیت آیتم همسایهٔ j، $r_{a,j}$ مقدار نرخی است که کاربر هدف a به آیتم همسایه j داده و $P_{a,i}^{CbShoShcSeICF}$ نرخ پیش‌بینی‌شدهٔ کاربر هدف a است که پس از بازدید، ممکن است به آیتم i بدهد. از این نرخ پیش‌بینی‌شده در تولید مجموعه آیتم‌های پیشنهادی استفاده می‌شود. نام این روش برگرفته از اجزای آن است و همانند روش پیشنهادی آیتم‌محور اول عمل می‌کند، با این تفاوت که روش محتوامحور نیز به آن اضافه شده است[2]. نام این الگوریتم پیشنهادی در نمودارهای ارزیابی برگرفته از جزءهای گوناگون آن است و به‌طور اختصاری با CbShoShcSeCF[3] مشخص می‌شود (دونوان و اسمیت، ۲۰۰۵).

ارزیابی الگوریتم کاربرمحور پیشنهادی

در شکل ۶ می‌توان با استفاده از خطای جذر میانگین مربعات، الگوریتم کاربرمحور پیشنهادشدهٔ این پژوهش را با رقیب‌هایش مقایسه کرد. همان‌طور که در شکل مشاهده می‌شود، از لحاظ این نوع خطا نیز الگوریتم پیشنهادشده نسبت به الگوریتم‌های دیگر، برتری نسبتاً زیادی دارد. به بیانی دیگر، خطای این روش پیشنهادی کاهش شایان توجهی را نسبت به سایر الگوریتم‌های مشابه پیشین خود نشان می‌دهد.

---

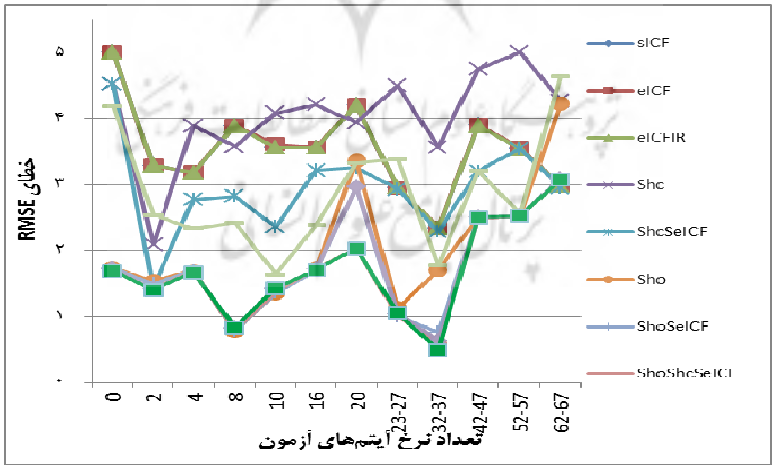
1. ShoShcSeICF = Subject heading overlapping (Sho) + Subject heading count (Shc) + enhanced Item-based Collaborative Filtering (ICF)
2. CbShoShcSeICF = Content based ShoShcSeICF
3. CbShoShcSeICF = Content based ShoShcSeICF



شکل ٦. مقایسهٔ الگوریتم کاربرمحور پیشنهادی با الگوریتم‌های رقیب (خطای جذر میانگین مربعات RMSE)

ارزیابی الگوریتم آیتم‌محور پیشنهادی

الگوریتم آیتم‌محور پیشنهادشده نیز در ارزیابی‌ها، برتری شایان توجهی نشان می‌دهد. دلیل این برتری چشمگیر را می‌توان علاوه‌بر استفادهٔ مناسب از هم‌پوشانی سرموضوعات در روند پیش‌بینی نرخ، بررسی و مقایسهٔ محتوای کتاب‌ها در تعیین شباهت بین آنها، عنوان کرد. مقایسهٔ این الگوریتم‌ها از دید خطای جذر میانگین مربعات، در شکل ۷ نشان داده شده است.



شکل ۷. مقایسهٔ الگوریتم آیتم‌محور پیشنهادی با الگوریتم‌های رقیب (خطای جذر میانگین مربعات RMSE)

نتیجه‌گیری و پیشنهادها

با توجه به روش‌های پیشنهادشده در این پژوهش و با استفاده از شبکهٔ اعتماد و زمان نرخ‌دهی، می‌توان به کاهش خطای پیش‌بینی پرداخت. در واقع شباهت کاربران به یکدیگر علاوه‌بر این الگوریتم پایه، به‌وسیلهٔ اعتمادی که خود کاربران به‌صراحت در شبکهٔ اجتماعی ایجاد می‌کنند، بهتر تشخیص داده می‌شود. استفاده از اعتمادهای صریح ایجادشده توسط کاربران انسانی که در شبکه‌های اجتماعی وجود دارد، موجب بهبود چشمگیر توصیه‌های سیستم و به بیان دیگر، کاهش شایان توجه خطا در پیش‌بینی‌های سیستم توصیه‌گر می‌شود. مهم‌تر از همه این که سیستم توصیه‌گر با استفاده از این اعتمادهای صریح، قادر به تولید مجموعه پیشنهادها در شرایط شروع آهستهٔ کاربر هدف خواهد بود.

در روش دوم، از یک منبع اطلاعات معنایی خارج از سیستم استفاده می‌شود. این منبع که طبقه‌بندی مفهومی از سرموضوع کتاب‌هاست، در تشخیص میزان شباهت ضمنی از دو طریق متفاوت کمک‌رسانی می‌کند که عبارت‌اند از: هم‌پوشانی سرموضوع کتاب‌ها و تعداد سرموضوع مشترک کتاب‌ها. به‌هر حال ممکن است کتاب‌هایی باشند که اطلاعات مفهومی آنها در طبقه‌بندی موجود نباشد. بنابراین در این روش از الگوریتم معنایی برای استخراج کلمات کلیدی و اصلی متن کتاب‌ها نیز استفاده شده است. این روش معنایی که از ابرداده‌های[1] پیوندی به‌منظور تشخیص کلمات کلیدی متن ورودی استفاده می‌کند، نتایج بسیار خوبی را در استخراج مهم‌ترین کلمات از خود نشان داده است. در مرحلهٔ بعد نیز این کلمات معنایی، مبنای مقایسهٔ محتوایی کتاب‌ها و تعیین میزان شباهت آنها به یکدیگر قرار گرفتند. در نتیجهٔ به‌کارگیری دو روش معنایی و محتوایی در این پژوهش، روشی برای بهبود پیشنهادها در سیستم پیشنهاددهنده ارائه شد. روش پیشنهادی، نوعی الگوریتم کاربرمحور است و براساس تشخیص شباهت کاربر هدف با همسایگانش و با در نظر گرفتن زمان نرخ‌دهی و تأثیر آن در شباهت‌یابی جهت کاهش تُنُکی داده‌ها به پیش‌بینی نرخ‌ها می‌پردازد. بر اساس یافته‌های پژوهش، برای بهبود و اصلاح مشکل کاربران جدید و تنکی داده، می‌توان در آینده از معیارهای شباهت‌یابی استفاده کرد که در بازهٔ اطمینان با متغیر هدف از بین همسایگان، نزدیک‌ترین همسایه را شناسایی و پیش‌بینی کند. با تأثیر زمان در معیار بازهٔ اطمینان چبیشف و محدود کردن بازهٔ اطمینان، امید است به پیش‌بینی نزدیک‌تری دست پیدا کرد.

در این پژوهش تلاش شد با شناسایی تأثیر بازهٔ زمان نرخ‌دهی (زمانی که کاربر به اقلام نرخ داده است) و با در نظر گرفتن بازهٔ اطمینان برای شباهت‌یابی کاربران، پیشنهادهای نزدیک‌تری

---

1. Mega Data

به‌علاقهٔ کاربران پیش‌بینی شود. با توجه به یافته‌های این پژوهش، می‌توان استنتاج روابط اعتماد ضمنی نزدیک به صریح و بازهٔ اطمینان را یکی از پژوهش‌های آتی تکمیل‌کنندهٔ این مقاله دانست.

فهرست منابع

حسنقلی‌پور، ط.؛ امیری، م.؛ فهیم، ف.؛ قادری عابد، ا.ح. (۱۳۹۲). بررسی تأثیر خصوصیات مشتریان بر تمایل آنها به پذیرش خرید اینترنتی (پیمایشی پیرامون دانشکدهٔ مدیریت دانشگاه تهران). *نشریهٔ مدیریت فناوری اطلاعات*، ۵(۴)، ۸۴-۶۷.

کرامتی، ع.؛ خالقی، ر. (۱۳۹۳). توسعهٔ یک سیستم پیشنهاددهندهٔ محصول طراحی مدل ترکیبی با بهره‌گیری از روش‌های فیلترینگ مشارکت‌محور، کشف قوانین انجمنی، و بخش‌بندی مشتریان. *نشریهٔ مهندسی صنایع*، ۴۸ (۲): ۲۸۰- ۲۵۷.

کریمی علویجه، م. ح.؛ عسگری، ش.؛ پرسته، س. (۱۳۹۴). فروشگاه اینترنتی هوشمند: سیستم پیشنهاددهندهٔ مبتنی بر تحلیل رفتار کاربران. *نشریهٔ مدیریت فناوری اطلاعات*، ۷ (۲)، ۴۰۶- ۳۸۵.

کی‌پور، ا.؛ براری، م. و شیرازی، ح. (۱۳۹۳). ارائهٔ روشی جدید برای پیشگویی پیوند بین رأس‌های موجود در شبکه‌های اجتماعی. *فصلنامهٔ مدیریت فناوری اطلاعات*، ۶ (۳)، ۴۸۶- ۴۷۵.

Candillier, L., Meyer, F. & Fessant, F. (2008). Designing Specific Weighted Similarity Measures to Improve Collaborative Filtering Systems. in *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*. vol. 5077, P. Perner, Ed., ed: Springer Berlin Heidelberg, pp. 242-255.

Gopal, R.D., Tripathi, A.K., Walter, Z.D. (2006). Economics of first-contact e-mail advertising. *Decision Support System*, 42(3),1366–1382.

Hasan Gholipour, T., Amiri, M., Fahim, F. & Ghaderi abed, A. (2013). Effect of customer characteristics on their willingness to adopt Internet shopping. *Quarterly Journal of Information technology management*, 5(4), 67-84. (*in Persian*)

Hill, W., Stead, L., Rosenstein, M. & Furnas, G. (1995). *CHI '95 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA - May 07 - 11, Pages 194-201.

Karimi, M., Askari, SH. & Paraste, S. (2015). Intelligent Internet store: Recommended system based on analysis of user behavior, *Quarterly Journal of Information technology management*, 7(2), 385-406. (*in Persian*)

Keramati, A. & Khaleghi, R. (2013). Developing a design product recommender system utilizes a hybrid model based collaborative filtering methods, the discovery of association rules and customer segmentation. *Specialist Journal of industrial manageent*, 48 (2), 257-280. (*in Persian*)

Keypour, A., Barari, M. & Shirazi, H. (2014). A new method for predict links between nodes in social networks. *Quarterly Journal of Information technology management*, 6(3), 475-486. (*in Persian*)

Lee, T. Q., Park, Y. & Park, Y. T. (2008). A time-based approach to effective recommender systems using implicit feedback. *Expert systems with applications*, 34(4), 3055-3062.

O'Donovan, J. & Smyth, B. (2005). Trust in recommender systems. *presented at the Proceedings of the 10th international conference on Intelligent user interfaces*, San Diego, California, USA.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. & Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. *presented at the Proceedings of the 1994 ACM conference on Computer supported cooperative work*. Oct 22 (pp. 175-186). ACM.

Sarwar, B., Karypis, G. Konstan, J. & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms, *presented at the Proceedings of the 10th international conference on World Wide Web*, Hong Kong, Hong Kong.

Shambour, Q. & Lu, J. (2012). A trust-semantic fusion-based recommendation approach for e-business applications. *Decision Support Systems*, 54(1), 768-80.

Shardanand, U. & Maes, P. (1995). Social information filtering: algorithms for automating "word of mouth". *presented at the Proceedings of the SIGCHI conference on Human factors in computing systems*.1995 May 1 (pp. 210-217). ACM Press/Addison-Wesley Publishing Co.

Shinde, S.K. & Kulkarni, U. (2012). Hybrid personalized recommender system using centering-bunching based clustering algorithm. *Expert Systems with Applications*, 39(1), 1381-1387.

Zamani, A., Rahmati, M. H. (2014). Identifying and Rating Affecting Factors on Business Process Management (Bpm) Successful Execution in Iranian Insurance Companies by Analytical Hierarchical Process (Ahp) Technique. *Journal of Social Issues & Humanities*, 2(11), 121-127.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی