دوره پانزدهم، شماره بهار و تابستان ۱۴۰۱

مجله فناوری اطلاعات در طراحی مهندسی

Information Technology in Engineering Design

http://ited.sinaweb.net

سیستم پیشنهاددهنده مبتنی بر اعتماد با اعمال پارامترهای فاصله اجتماعی، زمان و مکان و ارائه یک معیار جدید برای شباهت زمینه‌ای

سمانه شیبانی[۱] حسن شاکری*[۲] رضا شیبانی[۳]

(۱) گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

(۲) گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران*

(۳) گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

(تاریخ دریافت :۱۴۰۱/۰۶/۲۱ تاریخ پذیرش: ۱۴۰۱/۰۸/۲۷)

چکیده

سیستم‌های پیشنهاددهنده در سال‌های اخیر به طور گسترده در تحقیقات و کاربرد مورد توجه قرار گرفته‌اند. متداول‌ترین رویکرد برای شناسایی اقلام مورد علاقه کاربر در این سیستم‌ها پالایش مشارکتی (CF) است. با وجود این، CF پایه چالش‌های خاصی دارد و در شرایط خلوتی داده‌ها و شروع سرد، کارآمدی و دقت قابل قبولی ارائه نمی‌کند برای مقابله با این مشکلات، راه‌حل‌های مختلفی ارائه شده است. یکی از این راهکارها بهبود رویکرد پالایش مشارکتی سنتی با اعمال پارامتر اعتماد است که منجر به افزایش دقت پیشنهادهای سیستم می‌شود. از آنجا که مقدار اعتماد مستقیم به یک کاربر همواره در دسترس نیست، انتشار اعتماد یعنی تخمین اعتماد غیرمستقیم می‌تواند سودمند باشد. از طرف دیگر در صورتی که پارامترهای مختلف در انتشار اعتماد در نظر گرفته شود، کارایی و کارآمدی سیستم‌های پیشنهاددهنده مبتنی بر اعتماد بهبود می‌یابد. در این مقاله، یک مدل انتشار اعتماد چهاربعدی برای استفاده در سیستم‌های پیشنهاددهنده ارائه شده است. در این مدل، برای تخمین اعتماد غیرمستقیم بین کاربران سیستم، تأثیر چهار پارامتر فاصله اجتماعی، موقعیت مکانی، زمان و زمینه در انتشار اعتماد در نظر گرفته‌شده است. همچنین یک معیار شباهت زمینه‌ای جدید مبتنی بر درخت، پیشنهاد شده است که براساس وزن‌دهی به یال‌ها در درخت زمینه برمبنای دو سناریوی مختلف عمل می‌کند. مدل پیشنهادی بر روی چند مجموعه‌داده شناخته‌شده، اعمال شده و مورد ارزیابی قرار گرفته است. نتایج آزمایش‌های انجام‌شده نشان می‌دهد که مدل پیشنهادی دقت توصیه‌ها را به طور میانگین به میزان ۳,۷ درصد نسبت به بهترین روش‌های موجود افزایش می‌دهد.

واژه‌های کلیدی: سیستم پیشنهاددهنده، انتشار اعتماد، خدمات مبتنی بر مکان، شباهت زمینه‌ای، پیشنهاد آگاه از زمان

* عهده‌دار مکاتبات:

حسن شاکری

نشانی: گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

پست الکترونیکی: shakeri@mshdiau.ac.ir

۱- مقدمه

در سال‌های اخیر کاربرد سیستم‌های پیشنهاددهنده به عنوان یک ابزار کارآمد جهت کمک به کاربران در پیداکردن و انتخاب محصولات و خدمات مورد علاقه آنها گسترش یافته است. این سیستم‌ها از تکنیک‌های مبتنی بر هوش مصنوعی برای پیش‌بینی علایق کاربران استفاده می‌کنند و با پیشنهاد اقلام به کاربر و ارزیابی بازخورد وی در مورد ترجیحات کاربران، یاد می‌گیرند.

یکی از متداول‌ترین رویکردهای مورد استفاده برای شناسایی اقلام مناسب برای پیشنهادکردن، رویکرد فیلترینگ مشارکتی است که بر دو فرض مبتنی است: اول این که به احتمال زیاد اقلامی از سوی کاربر مورد پذیرش قرار می‌گیرد که مشابه با مواردی هستند که قبلاً توسط او انتخاب و پسندیده شده‌اند. دوم این که کاربر اقلامی را می‌پسندد که کاربران مشابه او به آنها امتیاز بالایی داده‌اند. سیستم‌های پیشنهاددهنده مبتنی بر CF براساس رفتار قبلی کاربر هدف و سایر کاربران، یک مدل می‌سازند و سپس از این مدل برای پیش‌بینی علایق کاربر استفاده می‌کنند.

با وجود مفیدبودن و محبوبیت رویکرد فیلترینگ مشارکتی، CF پایه با چالش‌هایی روبرو است که مهم‌ترین آنها دو مورد زیر است[۱, ۲]:

• خلوتی داده‌ها: فیلترینگ مشارکتی براساس ماتریس ارزشیابی‌های کاربر-آیتم عمل می‌کند. اما بسیاری از کاربران به ندرت اقلام را ارزشیابی می‌کنند. بنابراین بسیاری از سیستم‌های پیشنهاددهنده به ویژه در اوایل کار خود با ماتریس ارزشیابی خلوت مواجه هستند که این مشکل منجر به دقت و پوشش پایین توصیه‌ها می‌انجامد.

• شروع سرد: حتی اگر ماتریس ارزشیابی در مجموع خلوت نباشد، ممکن است سطر یا ستون مورد نیاز خالی یا خلوت باشد. به عبارت دیگر مشکل شروع سرد در دو حالت پیش می‌آید: حالت اول زمانی است که یک کاربر جدید اقلام بسیار کمی را ارزشیابی کرده است. حالت دوم وقتی است که یک آیتم جدید توسط کاربران بسیار کمی ارزشیابی شده است. در هر دو حالت، امکان محاسبه شباهت به صورت دقیق وجود ندارد و سیستم نمی‌تواند با دقت مطلوب، اقلام مناسب را شناسایی و پیشنهاد کند.

برای حل این مشکلات، راه‌حل‌های مختلفی توسط محققان ارائه شده است که ازجمله می‌توان به یادگیری فعال[۳]، آنتولوژی و روابط معنایی[۴]، قوانین وابستگی [۵] اشاره کرد.

از طرف دیگر، پیشرفت‌های اخیر در فناوری‌های موبایل و ردیابی موقعیت مکانی امکان توسعه سیستم‌های پیشنهاددهنده مبتنی بر مکان را فراهم آورده است. به عنوان نمونه در کارهای تحقیقاتی [۶-۹] مؤلفان راهکارهایی مبنی بر درنظرگرفتن مکان کاربر، اطلاعات جغرافیایی و حرکت کاربر پس از دریافت پیشنهادها به منظور بهبود صحت توصیه‌های سیستم‌های پیشنهاددهنده و حل مشکل شروع سرد ارائه کرده‌اند.

یک زمینه تحقیقاتی دیگر در حوزه سیستم‌های پیشنهاددهنده، درنظرگرفتن پارامتر زمان برای شناسایی علایق کاربر است. در [۱۰] پویایی زمانی ترجیحات کاربران براساس یک چارچوب تجزیه تنسور مدل‌سازی شده است. در این مدل، وزن ترجیحات و ارزشیابی‌های کاربر در گذشته برمبنای یک ضریب فرسایش مبتنی بر زمان کاهش می‌یابد. همچنین در [۱۱]، مؤلفان دو مدل برای پیشنهاد POI ارائه کرده‌اند و هر دو فاکتور مکان و زمان را در نظر گرفته‌اند.

همچنین برخی از کارهای تحقیقاتی اخیر از اعمال مفهوم زمینه و دسته (رده) اقلام و رابطه و شباهت بین زمینه‌ها به عنوان یک رویکرد کارآمد برای دستیابی به توصیه‌های دقیق‌تر استفاده می‌کنند. به عنوان نمونه در [۱۲] یک معیار شباهت معنایی برای یافتن همسایه‌های کاربر هدف و اقلام مورد علاقه وی پیشنهاد شده است.

یکی از موفق‌ترین تکنیک‌ها برای غلبه بر مشکلات خلوتی داده‌ها و شروع سرد اضافه‌کردن ارزیابی اعتماد به فیلترینگ مشارکتی سنتی است[۱۳–۱۵]. همچنین برخی از مدل‌های اعتماد محاسباتی ازجمله [۱۶–۱۸] پارامتر اطمینان را در محاسبه اعتماد و ارائه پیشنهاد برمبنای آن در نظر می‌گیرند. در حوزه مدیریت اعتماد، اطمینان سطح مطمئن‌بودن در مورد صحت محاسبه اعتماد را نشان می‌دهد[۱۹]. بهره‌گیری از یک مدل اعتماد آگاه از اطمینان می‌تواند صحت و کارآمدی سیستم را ارتقا ببخشد. بسیاری از مدل‌های اعتماد آگاه از اطمینان بر نظریه منطق ذهنی مبتنی هستند که توسط آودان جوسانگ ارائه شده است[۲۰–۲۲].

ازآنجا که مقدار اعتماد مستقیم همواره دردسترس نیست، تخمین سطح اعتماد به صورت غیرمستقیم می‌تواند در کاربردهای مختلف از جمله در سیستم‌های پیشنهاددهنده مبتنی بر اعتماد مفید باشد. در ادبیات اعتماد محاسباتی، این تخمین غیرمستقیم اعتماد اصطلاحاً انتشار اعتماد نامیده می‌شود[۲۳]. درصورتی‌که پارامترهای مختلف در انتشار اعتماد در نظر گرفته شود، کارآمدی و صحت سیستم‌های پیشنهاددهنده مبتنی بر اعتماد افزایش خواهد یافت. به طور مشخص درنظرگرفتن اثر فاصله مکانی، زمانی و زمینه‌ای در انتشار اعتماد در این راستا مؤثر خواهد بود. با وجود این در حد اطلاع ما اعمال هرسه این پارامترها به صورت همزمان در کارهای تحقیقاتی قبلی گزارش نشده است.

در این مقاله یک مدل انتشار اعتماد چهاربعدی برای استفاده در سیستم‌های پیشنهاددهنده ارائه شده است. در این مدل، فاصله اجتماعی (یعنی فاصله بین کاربران در شبکه اعتماد)، مکان، زمان و زمینه همگی در نظر گرفته شده است. همچنین تأثیر پارامتر اطمینان در انتشار اعتماد در نظر گرفته می‌شود. اطمینان به عنوان مبنای وزن‌دهی به زنجیره‌های مختلف در تجمیع اعتماد اعمال می‌شود. علاوه بر این، ارائه سطح اطمینان در کنار هر توصیه به کاربر، خروجی سیستم را برای وی غنی‌تر و مفیدتر می‌سازد.

مهم‌ترین نوآوری‌های این مقاله به شرح زیر است:

- در مدل پیشنهادی، بعدهای مکان و زمینه در انتشار اعتماد اعمال شده است.
- در این مقاله یک مدل انتشار چهاربعدی برمبنای اعمال همزمان روابط اجتماعی، زمان، مکان و زمینه برای استفاده در سیستم‌های پیشنهاددهنده معرفی شده است.
- مدل انتشار اعتماد پیشنهادی تأثیر پارامتر اطمینان را برای بهبود صحت در نظر می‌گیرد.
- مدل پیشنهادی از یک معیار جدید مبتنی بر درخت زمینه براساس وزن‌دهی به یال‌ها برای محاسبه شباهت معنایی بین زمینه‌ها استفاده می‌کند.

ساختار این مقاله در ادامه به این صورت است که در بخش ۲ کارهای مرتبط در حوزه سیستم‌های پیشنهاددهنده و انتشار اعتماد مرور و بررسی شده است. در بخش ۳ بیان مسأله ارائه می‌شود. در بخش ۴ مدل انتشار اعتماد چهاربعدی آگاه از اطمینان پیشنهادی معرفی و توصیف شده است. در بخش ۵ نتایج آزمایش‌های انجام‌شده ارائه و تحلیل می‌گردد. بخش ۶ به نتیجه‌گیری اختصاص دارد.

در سال‌های گذشته پیشرفت تکنولوژی ارتباطات باعث بوجود آمدن محیط محاسبات ابری شده‌است. محیط محاسبات ابری از دانش‌هایی مانند گرید، توزیع‌شده، وب۲ و مجازی‌سازی تشکیل شده‌است [۱, ۲]. یکی از چالش‌های مهم در محاسبات ابری، مسئله مصرف انرژی می‌باشد [۳]؛ محققان همواره به دنبال روش‌هایی هستند که انرژی مصرفی را کاهش دهند. افزایش مصرف انرژی باعث ایجاد گازهای کربنی شده و بر محیط زیست اثر مخرب دارد [۴, ۵]. با کاهش انرژی مصرفی در محاسبات ابری می‌توان به حفظ محیط زیست کمک نمود و کاهش هزینه‌های مرتبط با مصرف انرژی را رقم زد [۶].

در ارتباط با کاهش میزان انرژی مصرفی الگوریتم‌های متفاوتی ارائه شده‌است که از لحاظ پارامترهای مختلف مورد بررسی قرار گرفته‌اند. اما راه‌حلی موردنظر است که کاهش انرژی مصرفی با برقراری توافق‌نامه‌ی سطح خدمات همراه باشد [۷–۱۰]. مسئله‌ی

مهاجرت در محاسبات ابری رویکردی است که موجب برقراری توافق‌نامه‌ی سطح خدمات با هدف ذخیره مصرف انرژی می‌گردد. مسئله‌ی تجمیع در محاسبات ابری دارای مراحل الف) پیدا کردن ماشین فیزیکی پربار، ب) پیدا کردن ماشین مجازی مناسب برای مهاجرت از ماشین فیزیکی پربار، ج) پیدا کردن مقصد مناسب برای مهاجرت ماشین مجازی انتخاب شده در مرحله قبل، و ه) پیدا کردن ماشین فیزیکی کم‌بار برای خاموش کردن می‌باشد. پس از تخصیص ماشین‌های مجازی به ماشین‌های فیزیکی، ممکن است تعدادی از ماشین‌های فیزیکی دارای ظرفیت بار پردازشی بیش از بهره پردازشی خود شوند و پربار گردند. و اگر دارای ظرفیت بار پردازشی کمتر از میزان حداقل بهره پردازش خود باشند؛ کم‌بار می‌گردند. در این حالت بایستی یک و یا چند ماشین مجازی از ماشین فیزیکی پربار به ماشینی فیزیکی که ظرفیت پردازشی کافی داشته باشد مهاجرت داده شود. زیرا در حالت پرباری ماشین های فیزیکی، مصرف پردازنده‌ی ماشین فیزیکی افزایش می یابد که این امر باعث افزایش انرژی مصرفی می‌گردد؛ و در حالت کم‌باری ماشین های فیزیکی، می توان با هدف بهبود مصرف انرژی تمام ماشین‌های مجازی واقع برروی آن‌ها را به ماشین فیزیکی دیگری که دارای ظرفیت پردازش کافی محاسبات می‌باشد، مهاجرت داد و ماشین های فیزیکی مذکور را خاموش کرد. تحقیقات نشان می دهد یک ماشین فیزیکی بیکار در حالت روشن بودن ۷۰٪ از انرژی مورد نیاز خود را مصرف می‌کند [۱۱, ۱۲].

در مرجع [۱۳] ملاحظاتی برای مهاجرت ماشین‌های مجازی و شناسایی مبدا و مقصد مهاجرت آمده‌است. به‌طوری که شناسایی ماشین‌های فیزیکی پربار براساس ظرفیت پردازشی هر ماشین فیزیکی، بار پردازشی که در حال حاضر برروی آن قرار گرفته و میزان درخواست‌های ماشین‌های مجازی انجام می‌گردد. همچنین اشاره شده‌است که پربار بودن یک ماشین فیزیکی به معنای لزوم مهاجرت نیست بلکه باید نقض قرارداد نیز اتفاق افتد؛ بنابراین ماشین‌های فیزیکی با توجه به بار پردازشی خود و درخواست‌های ماشین‌های مجازی واقع برروی آن‌ها به سه دسته تقسیم می‌شوند: ۱) ماشین‌های فیزیکی که پربار هستند و نقض قرارداد را دارند، ۲) ماشین‌های فیزیکی که پربار هستند اما توافق‌نامه‌ی سطح خدمات را نقض نمی‌کنند و ۳) ماشین‌های فیزیکی که پربار نیستند و بنابراین هیچگاه توافق‌نامه‌ی سطح خدمات را نقض نمی‌کنند. ماشین‌های فیزیکی دسته‌ی اول به عنوان مبدا مهاجرت تعیین می‌گردند. در این مقاله مقصد مهاجرت با توجه به معیار نقض قرارداد تعیین می‌گردد، بنابراین یک ماشین فیزیکی غیرپربار از دسته‌ی سوم انتخاب می‌شود که هیچ‌گاه توافق‌نامه‌ی سطح خدمات را نقض نمی‌کند. اگر ماشینی فیزیکی در دسته‌ی سوم به عنوان مقصد مهاجرت وجود نداشت؛ آنگاه ماشینی فیزیکی از دسته‌ی دوم به عنوان مقصد مهاجرت انتخاب می‌گردد.

در بخش‌های بعدی به بررسی مرور ادبیات تحقیق، شرح روش پیشنهادی، ارزیابی و تحلیل‌های انجام شده و در انتها به نتیجه‌گیری پرداخته می‌شود.

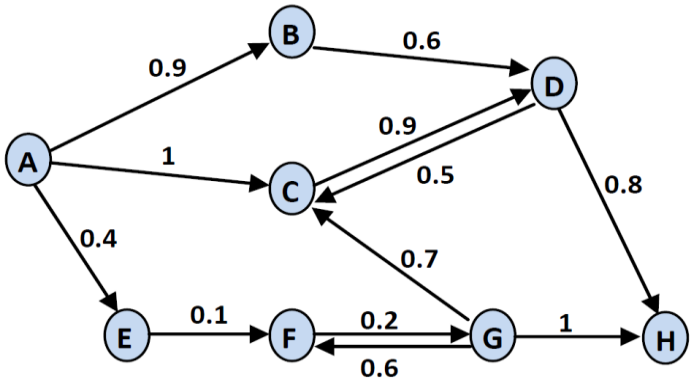
۲- کارهای مرتبط

در این بخش، مهم‌ترین کارهای تحقیقاتی مرتبط مرور و تحلیل می‌شود که در دو قسمت ارائه می‌شود: انتشار اعتماد و سیستم‌های پیشنهاددهنده مبتنی بر اعتماد، زمان، مکان و زمینه.

۲-۱ انتشار اعتماد

در حوزه انتشار اعتماد، اولین تکنیک برای محاسبه اعتماد غیرمستقیم براساس خاصیت تعدی اعتماد ارائه شد. تعدی یک خاصیت شهودی در ارتباط با اعتماد است که می‌توان آن را با یک مثال ساده بیان کرد: اگر Alice به Bob اعتماد داشته باشد و Bob به Claire اعتماد داشته باشد (و این اعتماد را به Alice گزارش کند)، در این صورت Alice تا حدی نسبت به Claire اعتماد پیدا خواهد کرد[24].

برای محاسبه مقدار اعتماد غیرمستقیم، عملگر «انتشار اعتماد» معمولاً با یک عملگر دیگر به نام «تجمیع اعتماد» ترکیب می‌شود. عملگر تجمیع برای محاسبه برآیند مقادیر اعتماد انتشاری در زنجیره‌های اعتماد مختلف استفاده می‌شود. روابط اعتماد بین افراد یک جامعه معمولاً به صورت یک «شبکه اعتماد»۱ نشان داده می‌شود که یک گراف جهت‌دار است که رأس‌های آن افراد جامعه و یال‌ها(لبه‌ها)ی آن روابط اعتماد بین افراد را نشان می‌دهد. یال‌ها معمولاً با مقادیر اعتماد برچسب‌گذاری می‌شوند (اغلب با مقادیر حقیقی متعلق به بازه [0, 1]). عملگرهای انتشار و تجمیع اعتماد به مسیرهای بین رأس مبدأ و رأس مقصد که «زنجیره‌های اعتماد» هم نامیده می‌شوند اعمال می‌شوند. شکل ۱ یک شبکه اعتماد نمونه را نشان می‌دهد.



شکل ۱. یک شبکه اعتماد نمونه

روش‌های مختلفی برای محاسبه مقدار اعتماد غیرمستقیم براساس خاصیت تعدی ارائه شده است. راهبرد موسوم به «ضرب مکرر»۲ یکی از معروف‌ترین و شهودی‌ترین روش‌ها برای این منظور است. این راهبرد بیان می‌کند که برچسب یال‌های کلیه یال‌های واقع بر زنجیره اعتماد (یعنی مسیر بین رأس‌های مبدأ و مقصد) در شبکه اعتماد به سادگی درهم ضرب شوند[25]. رابطه (۱) به صورت دقیق‌تر نحوه محاسبه اعتماد غیرمستقیم از رأس v1 به رأس vn در زنجیره v1...vn را براساس راهبرد ضرب مکرر بیان می‌کند.

$$IT_{1,n} = DT_{1,2}DT_{2,3} \dots DT_{n-1,n} = \prod_{i=1}^{n-1} DT_{i,i+1} \qquad (۱)$$

که در این رابطه، IT و DT به ترتیب مقادیر اعتماد غیرمستقیم و مستقیم را نشان می‌دهند.

یک گونه بهبودیافته از راهبرد ضرب مکرر در [26] ارائه شده است. در این رویکرد پیشنهاد شده است که برای محاسبه اعتماد غیرمستقیم، مقادیر «اعتماد ارجاعی»۳ همه یال‌ها بجز یال آخر و مقدار «اعتماد عملکردی»۴ یال آخر مسیر در هم ضرب شوند. رابطه (۲) نحوه اعمال این راهبرد اصلاح‌شده را نشان می‌دهد.

۱ web of trust

۲ Iterative Multiplication Strategy

۳ referal trust

۴ functional trust

$$IT_{1,n} = RT_{1,2}RT_{2,3} \dots RT_{n-2,n-1}FT_{n-1,n} = (\prod_{i=1}^{n-2} RT_{i,i+1}).FT_{n-1,n} \quad (۲)$$

که در رابطه فوق، RT و FT به ترتیب نشان‌دهنده اعتماد ارجاعی و عملکردی هستند.

از طرف دیگر برخی از مدل‌های اعتماد محاسباتی پارامتر عدم قطعیت را در بازنمایی نظرات اعتماد و عملگر انتشار اعتماد در نظر می‌گیرند. یکی از معروف‌ترین مدل‌ها در این زمینه، مدل منطق ذهنی است که در [27] ارائه شده و در کارهای تحقیقاتی بعدی بسط داده شده است. براساس این مدل، یک نظر اعتماد با یک سه‌تایی (b, d, u) نشان داده می‌شود به طوری که:

$$b + d + u = 1, \qquad \{b, d, u\} \in [0,1]^3 \quad (۳)$$

که b، d و u به ترتیب باور[۱]، بی‌باوری[۲] و عدم قطعیت[۳] را در یک نظر اعتماد نشان می‌دهند.

مدل منطق ذهنی همچنین عملگرهایی برای انتشار اعتماد با درنظرگرفتن پارامتر عدم قطعیت ارائه می‌کند[20].

برخی از کارهای تحقیقاتی مفهوم عدم قطعیت در نظرات اعتماد را از دیدگاه دیگری مورد توجه قرار داده‌اند و به جای عدم قطعیت، پارامتر اطمینان را در تخمین اعتماد اعمال می‌کنند. اطمینان را می‌توان نقطه مقابل عدم قطعیت در نظر گرفت. در [19] یک بازنمایی موسوم به بازه اعتماد برای نمایش همزمان اعتماد و اطمینان ارائه شده است. یک بازه اعتماد، بازه‌ای به صورت [L, U] است که در آن L و U به ترتیب کران پایین و بالای اعتماد را نشان می‌دهند و عرض بازه یعنی W = U – L مقدار عدم قطعیت نظر اعتماد را نشان می‌دهد. در این مدل، روابط زیر برای محاسبه اعتماد غیرمستقیم رأس مبدأ S به رأس مقصد D براساس نظر و توصیه رأس واسطه X ارائه شده است:

$$[LSX, USX] \times [LXD, UXD] = [LSD, USD] \quad (۴)$$

به طوری که:

$$LSD = LSX.LXD \quad (۵)$$

$$USD = 1 - LSX.(1-UXD) \quad (۶)$$

براساس روابط فوق، الگوریتم انتشار اعتماد در [19] در شکل ۲ نشان داده شده است.

برخی از کارهای تحقیقاتی مدل‌های اعتماد پویا ارائه کرده‌اند. یکی از این مدل‌ها مدل FJ نام دارد و در [28] ارائه شده است. در این مدل، رابطه (۷) برای تخمین نظرات اعتماد در یک لحظه زمانی خاص پیشنهاد شده است.

$$y(t+1) = (l_n - \delta)Wy(t) + \delta y(0), t = 0, 1, 2, \dots \quad (۷)$$

که در این رابطه y(t) بردار نشان‌دهنده نظرات افراد در لحظه زمانی t است و δ قابلیت کنشگرها برای نفوذهای بین فردی و ln ماتریس همانی n×n است. W=(wij)m×m اهمیت نسبی کنشگرها بر روی ترجیحات هر کاربر را نشان می‌دهد.

۱ belief

۲ disbelief

۳ uncertainty

Function PropagateTrust

Inputs:

Trust Network (TN),

Source node (n_0),

Sink node (n_∞)

Output:

The propagated trust which describes the amount of trust that the source has for the sink (TI_{prop})

Begin

$TI_{total} = [0, 1]$

For all paths p from n_0 to n_∞ do

Begin

$TI_{path} = [1, 1]$

For all edges n_in_j in the path p do

$TI_{path} = TI_{path} \times TI_{ij}$ (using Equation (9) and (10))

End

$TI_{total} = TI_{total} + TI_{path}$

Return TI_{total}

شکل ۲. الگوریتم انتشار اعتماد ارائه‌شده در [19]

۲-۲ تأثیر رابطه اجتماعی، زمان، مکان و زمینه در سیستم‌های پیشنهاددهنده

کارهای تحقیقاتی مختلفی با هدف بررسی تأثیر روابط اجتماعی، زمان، مکان و زمینه بر کارایی سیستم‌های پیشنهاددهنده انجام شده است. به عنوان نمونه در مرجع [29] یک الگوریتم توصیه مبتنی بر مکان برای ارتقاء صحت اقلام پیشنهادی براساس یادگیری با توجه به تحلیل پروفایل کاربر در شبکه‌های اجتماعی و موقعیت مکانی وی ارائه شده است. اگرچه این الگوریتم، نوعی زمینه را در نظر می‌گیرد ولی تلقی آن از مفهوم «زمینه» با آنچه در مدل پیشنهادی ما مورد نظر است، متفاوت است. در واقع در [29]، اصطلاح «زمینه» به شرایط و حس‌وحال[۱] کاربر اطلاق می‌شود. درحالی‌که در مدل پیشنهادی ما منظور از زمینه، حوزه و دامنه اقلام پیشنهادی به کاربر است.

در [30] یک الگوریتم پیشنهاد POI ارائه شده است که دو خاصیت زیر را برای پیشنهادهای POI بعدی در نظر می‌گیرد:

۱) احتمال‌ها پیشرو و پسرو بین یک جفت POI که به صورت متوالی توسط کاربر مورد بازدید قرار می‌گیرند، نامتقارن هستند.

۱ feeling

۲) کاربران معمولاً POIهای متوالی مختلفی را در زمان‌های مختلف ترجیح می‌دهند.

در [8] یک مدل گروه‌بندی کاربران ارائه شده است تا اطلاعاتی در مورد کاربران از طریق یک شبکه اجتماعی مبتنی بر مکان (LBSN) به دست آید. ترجیحات این کاربران، نزدیکی موقعیت مکان‌های بازدیدشده توسط کاربران، روزهای فراغت کاربران و روابط اجتماعی بین آنها به صورت خودکار از سابقه مکان کاربران و پروفایل‌های آنها در شبکه اجتماعی استخراج می‌شود. با ترکیب این فاکتورها شباهت بین کاربران تعیین می‌شود و کاربران به گروه‌هایی دسته‌بندی می‌شوند.

در مرجع [10] رویکردی برای مدل‌سازی پویایی ترجیحات کاربر در سیستم‌های پیشنهاددهنده براساس چارچوب تجزیه تنسور زوجی ارائه شده است. ترجیحات گذشته کاربر وزن‌دهی می‌شوند و اهمیت آنها به تدریج براساس یک فاکتور موسوم به ضریب فرسایش زمانی فردی کاهش می‌یابد. مقدار این فاکتور به نرخ پویایی ترجیحات کاربر براساس [31]وابسته است که به صورت زیر محاسبه می‌شود:

$$UPD_i = 1 - \frac{\left|I^i_{prev} \cap I^i_{cur}\right|}{\left|I^i_{prev} \cup I^i_{cur}\right|} \quad (۸)$$

که در رابطه فوق، I^i_{prev} و I^i_{cur} به ترتیب مجموعه‌های ژانرهای فیلم‌هایی است که کاربر در دوره زمانی گذشته و حال با آنها تعامل داشته است. همچنین مدل مذکور از اطلاعات دموگرافیک کاربران و شباهت‌های بین آنها به عنوان دانش پیشین درباره پویایی ترجیحات کاربر استفاده می‌کند.

در [11]، مؤلفان دو مدل برای ارائه توصیه‌های نظرات و مکان‌ها پیشنهاد کرده‌اند که فاکتورهای مکانی، متنی و زمانی مرتبط با نظرات کاربران و نیز تأثیر سابقه بازدید کاربران از مکان‌ها و نفوذ اجتماعی نظرات کاربران در آنها در نظر گرفته شده است. هر دو مدل برای انتخاب مرتبط‌ترین کاربران از پیش‌پالایش اطلاعات [32] استفاده می‌کنند. براساس [33, 34]، چهار دلیل اصلی تغییر ترجیحات کاربران یعنی کشف مکان‌های جدید، تجربه کاربر، محبوبیت و نفوذ اجتماعی در نظر گرفته شده‌اند.

مدل‌های ارائه‌شده در [11] مدل پالایش مبتنی بر آیتم را به دو روش توسعه می‌دهند: (۱) اعمال فاکتور تقریب در محاسبه شباهت بین دو کاربر (۲) درنظرگرفتن تاریخچه نظرات فقط کاربران نزدیک به جای کلیه نظرات.

کار تحقیقاتی ارائه‌شده در [35] یک مرور مفهومی از رویکردهای سیستم‌های پیشنهاددهنده مبتنی بر اعتماد و مطالعات اخیر بر روی این سیستم‌ها و همچنین تحلیلی از ملاک‌های اعتماد مورد استفاده در این حوزه ارائه می‌کند. مؤلفان نتیجه می‌گیرند که در ساختن ملاک اعتماد باید زمان ارزشیابی همراه با اطلاعات ارزشیابی آیتم در نظر گرفته شود تا ملاک‌های پویایی اعتماد، با کیفیت‌تر شود. آنها همچنین پیشنهاد می‌کنند که اطلاعات زمینه‌ای و رفتاری کاربران و اقلام در ملاک‌های اعتماد در نظر گرفته شوند تا معیارهای وابسته به زمینه تعیین شوند.

در [36] یک رویکرد بین دامنه‌ای برای سیستم‌های پیشنهاددهنده گروهی ارائه شده است. پیشنهادهای ارائه‌شده توسط کاربران قابل اعتماد و خوش‌نام در گروه، پذیرش توصیه‌ها را نسبت به پیشنهادهای سایر افراد گروه بهبود می‌بخشد. سیستم به گونه‌ای طراحی شده است که اطلاعات زیردامنه‌های مختلف دامنه گردشگری را در نظر می‌گیرد.

در [37]، یک سیستم پیشنهاددهنده بین دامنه‌ای[1] (CDRS) معرفی شده است که از داده‌های چندین دامنه برای کاهش مشکل خلوتی استفاده می‌کند. این سیستم اطلاعات اعتماد و بی‌اعتمادی را برای بهبود قابلیت اعتماد پیشنهادهای تولیدشده در نظر می‌گیرد.

در [38]، یک رویکرد پس‌پالایش مبتنی بر اعتماد آگاه از زمینه برای غلبه بر مشکلات خلوتی داده‌ها و شروع سرد در سیستم‌های پیشنهاددهنده پیشنهاد شده است. این رویکرد از میانگین اختلاف نسبی بین زمینه‌ها استفاده می‌کند. مؤلفان ابتدا میانگین امتیاز را برای هر شرایط زمینه‌ای محاسبه می‌کنند و همه ارزشیابی‌ها را براساس وضعیت زمینه‌ای چندگانه متوازن می‌کنند. آنها همچنین از مفهوم اطمینان برای حذف کاربران غیرقابل‌اطمینان شبکه اعتماد قبل از تولید پیش‌بینی‌ها استفاده می‌کنند.

۳ - بیان مسأله

این بخش یک تعریف فرمال از مسأله ارائه می‌کند و نمادهای مورد استفاده در ادامه مقاله را تعریف می‌کند.

به صورت فرمال، مسأله این تحقیق را می‌توان به صورت زیر بیان کرد:

ورودی:

- **u**: شناسه کاربر فعال، یعنی کاربری که قرار است امتیاز او به آیتم هدف تخمین زده شود.
- **i**: شناسه آیتم هدف، یعنی آیتمی که پیش‌بینی امتیاز برای آن قرار است انجام شود.
- **t**: زمان فعلی
- **l**: موقعیت مکانی کاربر فعال

خروجی:

- **r(u, i, t, l)**: مقدار تخمین‌زده‌شده برای امتیاز کاربر فعال **(u)** به آیتم هدف **(i)** در لحظه زمانی کنونی **(t)** و موقعیت مکانی فعلی **(l)** که به عنوان مقدار تابع **r** نشان داده می‌شود.

مفروضات:

- پایگاه داده ارزشیابی‌های آگاه از زمان و مکان کاربران در دسترس است و یا در زمان شروع به کار سیستم امکان ساختن آن از مجموعه ارزشیابی‌های خام کاربران وجود دارد. هر رکورد این پایگاه داده به شکل یک پنج‌تایی **<u, i, t, l, r>** است. این پنج فیلد به ترتیب شناسه کاربر، شناسه آیتم، زمان، مکان و مقدار ارزشیابی را نشان می‌دهند. البته مقدار **r** در بسیاری از رکوردهای پایگاه داده تهی (نامعلوم) است.
- پروفایل کاربران سیستم پیشنهاددهنده در شبکه‌های اجتماعی یا حداقل برخی از فیلدهای پروفایل آنها ممکن است در دسترس باشد. همچنین ممکن است شبکه مقادیر اعتماد صریح بین کاربران موجود باشد. در این صورت دقت پیش‌بینی‌ها و پیشنهادها افزایش خواهد یافت.

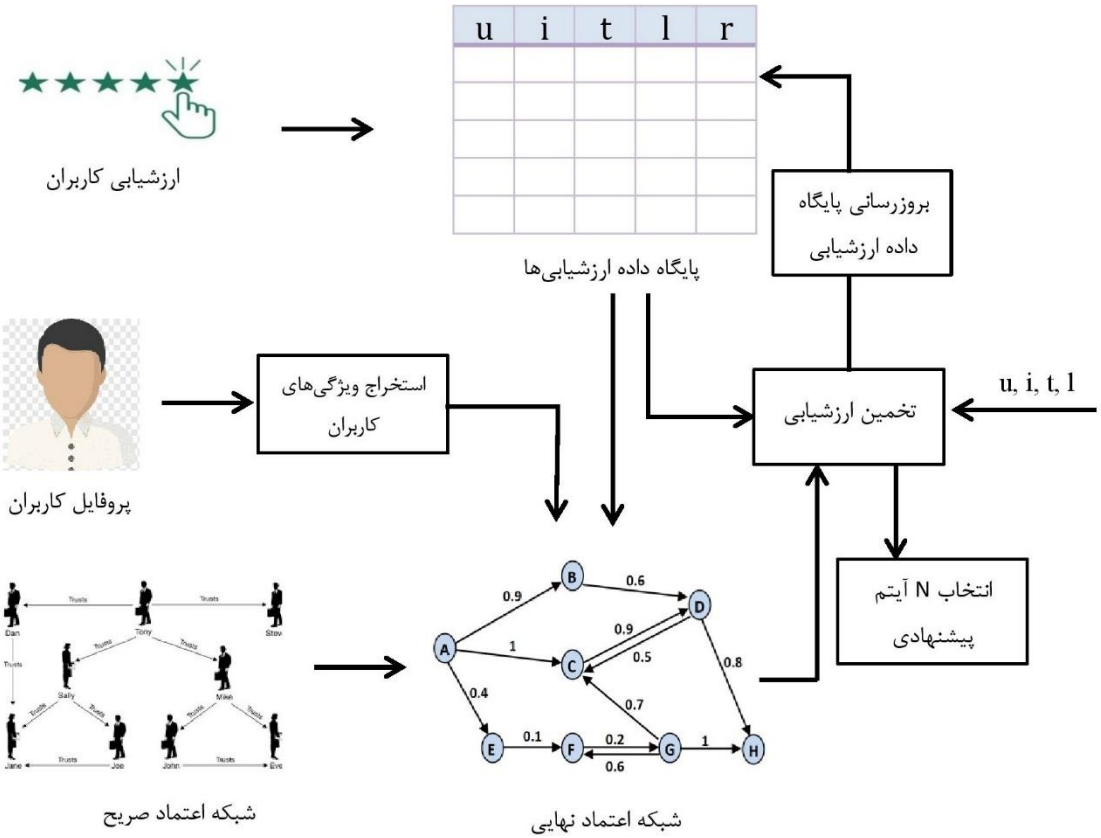
[1] Cross-Domain Recommender System

۴ – مدل پیشنهادی

در این بخش، مدل پیشنهادی توصیف می‌شود. ابتدا در زیربخش ۴–۱ توصیف کلی مدل و فلوچارت و شبه‌کد رویکرد پیشنهادی برای تخمین ارزشیابی و انتخاب N آیتم برتر ارائه شده است سپس در زیربخش ۴–۲ ماژول تخمین ارزشیابی با جزئیات بیشتر تشریح می‌گردد.

۴–۱ توصیف کلی مدل پیشنهادی

طرح کلی مدل پیشنهادی در شکل ۳ نشان داده شده است.



شکل ۳. طرح کلی روش پیشنهادی

همان‌طور که در شکل ۳ دیده می‌شود، در ابتدای کار مقدار ارزشیابی کاربران به صورت خام در دسترس است. چند گام پیش‌پردازش بر روی این داده‌های خام لازم است تا یک پایگاه داده ارزشیابی ساخت‌یافته[۱] ساخته شود. به عنوان نمونه، ارزشیابی‌های خام ممکن است در قالب متن باشند یا حتی ممکن است ارزشیابی‌ها صریحاً در دسترس نباشند بلکه لازم باشد تا ارزشیابی‌های کاربران به صورت ضمنی براساس رفتار آنها با آیتم‌ها (مثلاً تعداد دفعات تماشای یک فیلم یا مدت زمان ماندن بر روی یک صفحه وب) تخمین زده شود.

۱ structured

از طرف دیگر اگر پروفایل کاربر در دسترس باشد، این امکان وجود دارد که ویژگی‌ها و اطلاعات دموگرافیک کاربر از طریق پردازش داده‌های پروفایل، استخراج شود. این اطلاعات بعداً به عنوان یکی از مهم‌ترین پارامترها در محاسبه شباهت بین کاربران و ارتقاء صحت پیشنهادها استفاده خواهد شد.

اگر روابط اعتماد بین کاربران به صورت صریح و براساس نظرات اعتماد مستقیم در دسترس باشد، این روابط به شکل یک گراف جهت‌دار یعنی یک شبکه اعتماد مستقیم به عنوان یک ورودی سیستم دریافت می‌شود. بعداً این مقادیر اعتماد صریح با مقادیر اعتماد ضمنی که از پایگاه داده ارزشیابی‌ها و شباهت‌های پروفایل کاربران استخراج شده است ترکیب خواهد شد تا شبکه اعتماد نهایی شکل بگیرد.

برای مشخص‌کردن اقلام مورد علاقه کاربر فعال، شناسه کاربر و شناسه اقلام کاندید، زمان جاری و موقعیت مکانی کاربر به عنوان چهار پارامتر u، i، t و l به سیستم در نظر گرفته می‌شوند. در ماژول پیش‌بینی ارزشیابی (Rating Prediction) که مهم‌ترین بخش مدل است، مقدار r(u, i, t, l) براساس شبکه اعتماد نهایی آگاه از اطمینان و پایگاه داده ارزشیابی‌های موجود محاسبه می‌شود. خروجی این ماژول برای بروزرسانی پایگاه داده ارزشیابی مورد استفاده قرار می‌گیرد و همچنین جهت انتخاب اقلام برتر نهایی از دید کاربر فعال به ماژول انتخاب N آیتم برتر (Top-N Selection) تحویل داده می‌شود.

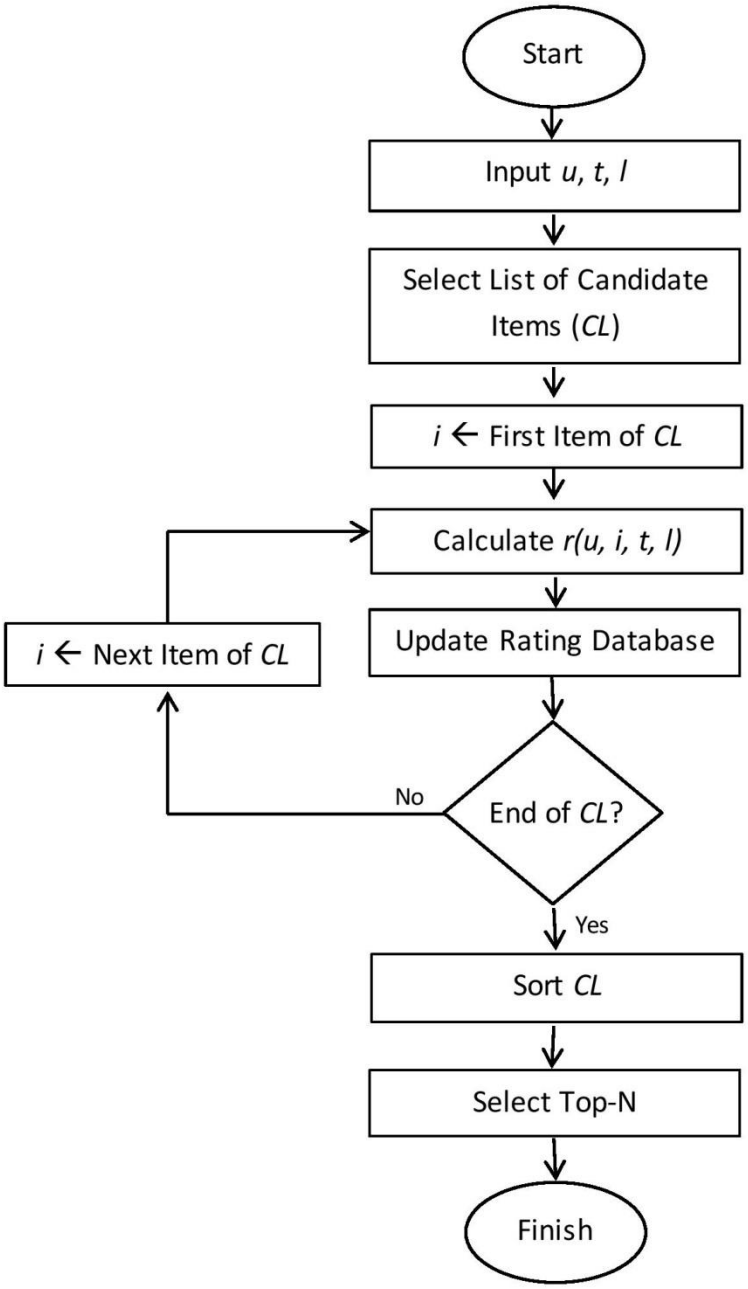
فلوچارت رویکرد پیشنهادی برای تخمین اقلام مورد علاقه کاربر فعال در شکل ۴ نشان داده شده است.

همان‌طور که در فلوچارت نشان داده شده است، شناسه کاربر فعال (u)، زمان جاری (t)، و موقعیت مکانی کاربر (l) به هنگام ورود او به سیستم دریافت می‌شود. سپس باید اقلام کاندید برای ارائه به کاربر انتخاب شوند. این اقلام آنهایی هستند که محتمل است که مورد علاقه کاربر باشند. انتخاب اقلام کاندید در کارایی سیستم اهمیت زیادی دارد زیرا اگر قرار باشد ارزشیابی کاربر به کل اقلام تخمین زده شود، این کار بسیار زمان‌بر خواهد بود. برای انتخاب اقلام کاندید، پیشنهاد می‌شود که اقلام مشابه اقلام انتخاب‌شده توسط کاربر فعال و همچنین اقلامی که توسط کاربران مشابه کاربر فعال انتخاب شده‌اند در نظر گرفته شوند. به عبارت دیگر برای تعیین لیست اقلام پیشنهادی (CL)، ترکیب دو رویکرد کاربر-آیتم-آیتم و کاربر-کاربر-آیتم اعمال می‌شود.

در گام بعدی، در ماژول تخمین ارزشیابی، مقدار r(u, i, t, l) برای هر یک از اقلام لیست کاندید محاسبه می‌شود. جزئیات بیشتر در مورد این ماژول بعداً در این بخش ارائه می‌شود. بعد از محاسبه ارزشیابی تخمینی کاربر به اقلام کاندید، این ارزشیابی‌ها به عنوان رکوردهای جدید در پایگاه داده ارزشیابی‌ها درج می‌شوند.

در مرحله بعد، CL براساس ترتیب نزولی مقادیر ارزشیابی (r) مرتب می‌شود و بالاخره در آخرین گام، نخستین N سطر CL انتخاب و به عنوان N آیتم برتر به کاربر پیشنهاد می‌شود.

شبه کد الگوریتم پیشنهادی در شکل ۵ نشان داده شده است.



شکل ۴. فلوچارت رویکرد پیشنهادی برای استخراج N آیتم برتر

۴-۲ ماژول تخمین ارزشیابی

همان‌گونه که قبلاً اشاره شد، مهم‌ترین بخش از مدل پیشنهادی، ماژول تخمین ارزشیابی است که هدف آن تخمین مقدار r(u, i, t, l) است. این ماژول براساس انتشار اعتماد در فضای چهاربعدی شامل ابعاد روابط اجتماعی، زمان، مکان و زمینه عمل می‌کند. رویکرد پیشنهادی برای انتشار اعتماد و تخمین ارزشیابی مبتنی بر آن در ادامه توضیح داده شده است.

```
Input u, t, l
UP = Set of Items close to u's preferences
UN = Set of Items selected by u's neighbors
CL = UP ∪ UN
For each item i∈CL
Begin
Calculate r(u, i, t, l)
Insert tuple <u, i, t, l, r(u, i, t, l)> into Rating DB
End
Sort CL in descending order based on the values of r
Recommend top N items in CL to u
```

شکل ۵. شبه کد الگوریتم پیشنهادی

قبل از اجرای ماژول تخمین ارزشیابی، شبکه اعتماد نهایی براساس ترکیب وزنی مقادیر اعتماد صریح، شباهت پروفایل کاربران، و شباهت ارزشیابی‌های کاربران ساخته می‌شود. برای محاسبه شباهت، معیارهای شباهت مختلفی شامل شاخص کسینوسی، جاکارد و ضریب همبستگی پیرسون امتحان شد. نتایج هر معیار و انتخاب نهایی در بخش ۵ گزارش شده است. علاوه بر این برای رابطه اعتماد (هر یال در شبکه اعتماد)، سطح اطمینان نیز محاسبه می‌شود و بنابراین گراف نهایی یک شبکه اعتماد آگاه از اطمینان است. مقدار اطمینان به صورت زیر محاسبه می‌شود:

$$C = \frac{n}{n+1} \tag{۹}$$

که n تعداد ویژگی‌ها یا ارزشیابی‌هایی است که مقدار اعتماد براساس آنها محاسبه شده است. مشخصاً در محاسبه اعتماد مبتنی بر شباهت پروفایلی، n تعداد ویژگی‌ها در پروفایل است و در محاسبه اعتماد برمبنای شباهت ارزشیابی‌ها، n تعداد اقلامی که توسط هر دو کاربر مورد ارزشیابی قرار گرفته است.

هدف از انتشار اعتماد بر اساس روابط اجتماعی این است که نظر اعتماد (T, C) با استفاده از ترکیب دو نظر اعتماد (T1, C1) و (T2, C2) محاسبه شود. به عبارت دیگر هدف این است که رابطه مناسبی برای محاسبه اعتماد غیرمستقیم زیر ارائه شود:

$$(T, C) = (T_1, C_1) \otimes (T_2, C_2) \tag{۱۰}$$

که (T1, C1) و (T2, C2) یال‌های متوالی در شبکه اعتماد هست و نماد $\otimes$ عملگر انتشار اعتماد را نشان می‌دهد.

برای اعمال انتشار اعتماد مبتنی بر زمان، مقادیر اعتماد از کاربر u1 به کاربر u2 در طول زمان به عنوان یک تابع در نظر گرفته می‌شود و براساس تغییرات این تابع، مقدار اعتماد در لحظه زمانی جاری t تخمین زده می‌شود.

برای اعمال انتشار اعتماد مبتنی بر مکان، با فرض این که اعتماد از کاربر u1 به کاربر u2 در موقعیت مکانی l0 معلوم است، برای تخمین مقدار اعتماد در موقعیت مکانی فعلی یعنی l، فاصله جغرافیایی بین l0 و l اندازه‌گیری می‌شود و مقدار اعتماد در مکان l براساس میزان نزدیک‌بودن دو مکان تخمین زده می‌شود. هر یک از مکان‌های l0 و l به صورت یک

نقطه (موقعیت) در نظر گرفته می‌شود که با یک زوج مرتب طول و عرض جغرافیایی مشخص می‌شود و فاصله این دو نقطه از یکدیگر برمبنای معیار فاصله اقلیدسی مطابق رابطه ۱۱ تعیین می‌گردد:

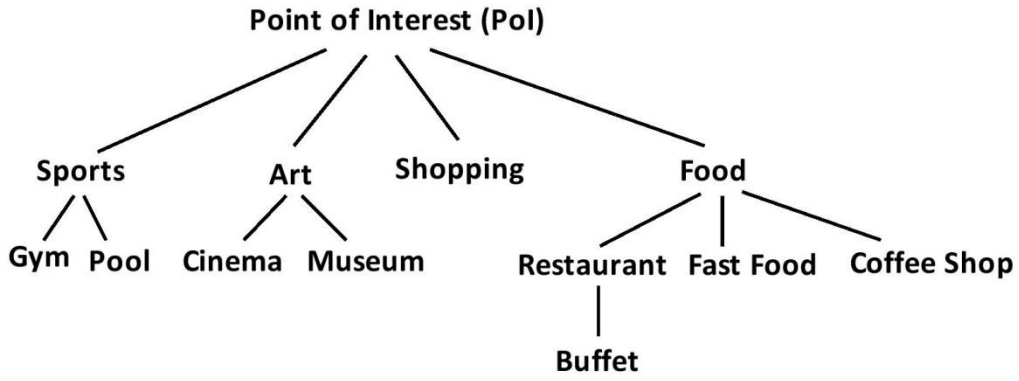
$$d = \sqrt{(x - x_0)^2 + (y - y_0)^2} \qquad (۱۱)$$

که در این رابطه، d فاصله بین دو نقطه l0 و l را نشان می‌دهد و مختصات این دو نقطه به ترتیب به صورت (x0,y0) و (x, y) در نظر گرفته شده است. پس از محاسبه d، ضریب انتشار مکانی اعتماد طبق رابطه ۱۲ محاسبه می‌شود:

$$\frac{d}{dmax}.\alpha - c = 1 \qquad (۱۲)$$

که dmax حداکثر فاصله جغرافیایی بین دو موقعیت را نشان می‌دهد و $0<\alpha<1$ پارامتر تنظیم انتشار مکانی است به‌طوری که ضریب انتشار اعتماد بین دو نقطه با بیشترین فاصله جغرافیایی برابر است با 1-α.

برای انتشار اعتماد مبتنی بر زمینه، پیشنهاد شده است که روابط بین زمینه‌ها در یک ساختار سلسله‌مراتبی به نام درخت زمینه بازنمایی شود. به عنوان نمونه، برای سیستم‌های پیشنهاددهنده POI، درخت زمینه می‌تواند به صورت درخت شکل ۶ یا شبیه آن باشد.



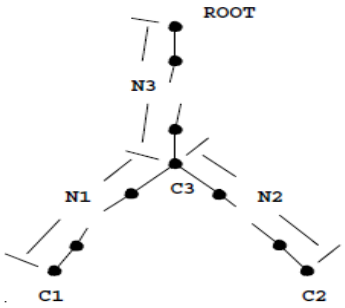
شکل ۶. یک درخت زمینه نمونه برای انتشار اعتماد مبتنی بر زمینه در سیستم‌های پیشنهاددهنده POI

با استفاده از درخت زمینه شکل ۷، اگر اعتماد از یک کاربر u1 به کاربر u2 در یک زمینه مشخص c0 داده شده باشد، برای تخمین مقدار اعتماد در زمینه جاری c ابتدا زوج گره متناظر با زمینه‌های c0 و c در درخت با توجه به نوع POIها (یا در حالت کلی آیتم‌ها) پیدا می‌شوند. سپس مقدار نزدیک‌بودن بین دو زمینه با استفاده از معیار فاصله (یا شباهت) محاسبه می‌شود. هرچه زمینه‌های c0 و c نزدیک‌تر باشند، انتشار اعتماد از c0 به c قوی‌تر خواهد بود. برای محاسبه شباهت بین دو زمینه، می‌توان از معیار شباهت Wu & Palmer (WUP) [39] استفاده کرد. براساس این معیار، برای تعیین مقدار شباهت بین گره‌های C1 و C2 در درخت، ابتدا پایین‌ترین اَبَرمفهوم مشترک۱ (LCS) گره‌های C1 و C2 به عنوان C3 تعیین می‌شود. سپس شباهت زمینه‌ای بین C1 و C2 به صورت زیر محاسبه می‌شود:

$$ConSim(C_1, C_2) = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3} \qquad (۱۳)$$

همان‌طور که در شکل ۷ دیده می‌شود، N1 تعداد گره‌های واقع در مسیر از C1 به C3، N2 تعداد گره‌های واقع در مسیر از C2 به C3، و N3 تعداد گره‌های واقع در مسیر از ریشه به C3 است.

۱ Least Common Superconcept



شکل ۷. نمادگذاری مورد استفاده در معیار شباهت زمینه‌ای WUP [39]

اما در روش پیشنهادی معیاری غیر از WUP برای تخمین شباهت بین زمینه‌ها پیشنهاد و استفاده شده است. این معیار طول مسیر بین گره‌های C1 و C2 را به عنوان معیار شباهت در نظر می‌گیرد که برای محاسبه طول مسیر به هر یال یک برچسب وزن، اختصاص داده شده و مجموع برچسب‌های واقع در مسیر بین دو گروه به صورت نرمال‌شده فاصله را نشان می‌دهد. سپس شباهت براساس فاصله تعیین می‌شود. رابطه زیر نحوه محاسبه فاصله در معیار پیشنهادی را نشان می‌دهد:

$$ContextDistance(C_1, C_2) = \frac{\sum_{ij \in path(C1,C2)} w_{ij}}{\sum_{ij \in MaxPath} w_{ij}} \quad (۱۴)$$

که در رابطه فوق path(C1,C2) مسیر واقع بین دو گره C1 و C2 است (بدیهی است که با توجه به تعریف درخت، بین هر دو گره فقط یک مسیر یکتا وجود دارد)، MaxPath مسیر بین دو گره درخت را نشان می‌دهد که بیشترین مقدار فاصله را از یکدیگر دارند و wij برچسب وزنی یال ij را نشان می‌دهد.

اکنون شباهت بین دو زمینه براساس فاصله تعیین می‌شود:

$$ContextSimilarity(C_1, C_2) = 1 - ContextDistance(C_1, C_2) \quad (۱۵)$$

برای اختصاص برچسب وزن به یال‌های درخت، دو سناریوی مختلف مورد بررسی قرار گرفت: در سناریوی اول به هر یال به صورت جداگانه و مستقل از سطحی که در آن واقع است، وزن تخصیص یافت. در سناریوی دوم به کلیه یال‌های واقع در هر سطح از درخت، یک وزن یکسان اختصاص داده شد. با توجه به این که براساس آزمایش‌های انجام‌شده، سناریوی اول نتایج با دقت بالاتری ارائه می‌کند، در روش پیشنهادی، سناریوی اول مورد استفاده قرار گرفته است.

۵- ارزیابی مدل پیشنهادی

در این بخش نتایج ارزیابی مدل پیشنهادی براساس آزمایش‌های مختلف ارائه و تحلیل شده است. ابتدا در ۵-۱ مجموعه‌داده‌های مورد استفاده معرفی شده است. سپس در ۵-۲ معیارهای ارزیابی و الگوریتم‌های مورد مقایسه توصیف می‌شوند و بالاخره در ۵-۳ نتایج ارزیابی رویکرد پیشنهادی ارائه و با نتایج مربوط به چند روش موجود مقایسه شده است.

۵-۱ مجموعه‌داده‌ها

آزمایش‌ها برروی سه مجموعه‌داده عمومی واقعی یعنی Yelp [40]، Foursquare [41] و Gowalla [42] انجام شده است. توصیف این سه مجموعه‌داده به شرح زیر است:

Yelp: این مجموعه‌داده اطلاعاتی در مورد ارزشیابی‌ها و نظرات کاربران برای کسب‌وکارهایی از ۱۰ شهر را دربردارد.

Foursquare: این مجموعه‌داده داده‌های حضور کاربران از ۱۲ آپریل ۲۰۱۲ تا ۱۶ فوریه ۲۰۱۳ در مکان‌های مورد علاقه در شهر نیویورک را دربردارد. در این تحقیق مانند [30] برای دستیابی به داده‌های کاربردی‌تر POIهایی انتخاب شده است که حداقل توسط ۵ کاربر بازدید شده است و کاربرانی که حداقل از ۱۰ مکان بازدید کرده‌اند.

Gowalla: این مجموعه‌داده اطلاعات حضور کاربران در شهرهای مختلف جهان در مکان‌های مورد علاقه در بازه زمانی فوریه ۲۰۰۹ تا اکتبر ۲۰۱۰ را ارائه می‌کند. مشابه [30] رکوردهای مربوط به POIهای کالیفرنیا انتخاب شده است و مکان‌های غیرمحبوب که توسط کمتر از ۱۵ کاربر بازدید شده‌اند و همچنین کاربرانی که که کمتر از ۲۰ مکان را بازدید کرده‌اند فیلتر شده است.

اطلاعات آماری پایه در مورد سه مجموعه‌داده مورد پردازش در جدول ۱ ارائه شده است.

جدول ۱. اطلاعات آماری در مورد مجموعه‌داده‌های مورد استفاده در ارزیابی

Dataset	#Users	#POIs	#Check-ins	#Avg. check-ins per user
Yelp	552338	77078	2225204	4.03
Foursquare	1078	2944	71622	66.44
Gowalla	2166	4047	100986	46.62

۵-۲ معیارهای ارزیابی و الگوریتم‌های مورد مقایسه

در این تحقیق سه معیار متداول برای ارزیابی رویکرد پیشنهادی و مقایسه صحت و کارامدی آن با چند روش موجود مورد استفاده قرار گرفته است:

Precision@N (دقت): نسبت تعداد آیتم‌های مرتبط در لیست N آیتم برتر به N. به عبارت دیگر این معیار به صورت زیر تعریف می‌شود:

$$Precision = \frac{Relevant \cap Retrieved}{Retrieved} \quad (۱۶)$$

Recall@N (فراخوانی/بازخوانی/یادآور/یادآوری): نسبت تعداد آیتم‌های مرتبط در لیست N آیتم پیشنهادی به تعداد کل آیتم‌های مرتبط. به عبارت دیگر recall از رابطه زیر تعیین می‌شود:

$$Recall = \frac{Relevant \cap Retrieved}{Relevant} \quad (۱۷)$$

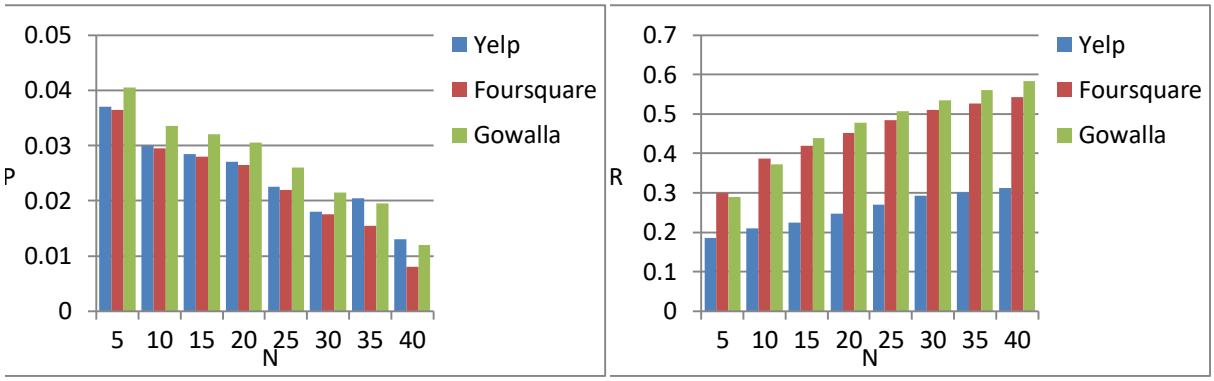
F1@N: یک معیار برقراری توازن بین دقت و فراخوانی است که هر دو ملاک را در نظر می‌گیرد. این معیار به درک مصالحه بین صحت و پوشش کمک می‌کند و به صورت رابطه زیر تعریف می‌شود:

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (۱۸)$$

نتایج مدل پیشنهادی با سه کار تحقیقاتی مشابه موجود شامل USSTC [11]، MEAP-T [30] و+LOCABAL [43] مقایسه شده است.

۳-۵ نتایج و تحلیل

اولین آزمایش برای تحلیل حساسیت مدل پیشنهادی از نظر معیارهای دقت و فراخوانی نسبت به تغییرات مقدار N برای پیشنهاد N آیتم برتر انجام شد. شکل ۸ نتایج این آزمایش را برای مقادیر مختلف N شامل N= 5, 10, 15, 20, 25, 30, 35, 40 بر روی سه مجموعه‌داده مورد استفاده نشان می‌دهد. همان‌طورکه انتظار می‌رود، با زیادشدن تعداد POIهای پیشنهادشده، دقت کاهش می‌یابد. این موضوع قابل درک است زیرا با افزایش POIهای بیشتر به کاربر، تعداد حالت‌های منفی کاذب افزایش می‌یابد. از طرف دیگر با افزایش N، مقدار فراخوانی زیاد می‌شود. این موضوع به این دلیل است که با پیشنهاد تعداد بیشتر POI، موارد مرتبط بیشتری هم به کاربر پیشنهاد می‌شوند. این به معنی افزایش تعداد مثبت‌های درست است که به افزایش مقدار پارامتر دقت می‌انجامد.



شکل ۸. تحلیل حساسیت برحسب مقادیر مختلف N برحسب (a) دقت (b) فراخوانی

آزمایش‌های دیگری برای ارزیابی و مقایسه کارایی مدل پیشنهادی نسبت به سایر روش‌ها انجام شد. معیارهای perecison@10، recall@10، F1@10 یعنی مقدار دقت و فراخوانی و F1 به ازای N=10 اندازه‌گیری شد.

جدول ۲ خلاصه نتایج ارزیابی را برحسب perecison@10 نشان می‌دهد. همان‌گونه که در جدول مشاهده می‌شود، روش پیشنهادی برای مجموعه‌داده‌های Yelp و Foursquare بهترین نتایج را ارائه می‌کند. اما در مورد مجموعه‌داده Gowalla، روش LOCABAL+ بهترین مقادیر دقت را ارائه می‌کند.

جدول ۳ نتایج recall@10 روش‌های مورد مقایسه بر روی مجموعه‌داده‌های مختلف را نشان می‌دهد. با توجه به جدول، روش پیشنهادی بر روی مجموعه‌داده‌های Yelp و Gowalla بهتر از روش‌های دیگر عمل می‌کند.

در نهایت، مقادیر F1 براساس رابطه (۱۶) برای کلیه حالت‌ها محاسبه شده است. نتایج حاصل در جدول ۴ ارائه شده است. مطابق آنچه که در جدول مشاهده می‌شود، روش پیشنهادی بهترین عملکرد کلی را برحسب معیار F1 ارائه می‌کند.

جدول ۲. مقادیر Precision@10 بر روی مجموعه‌داده‌های Yelp، Foursquare و Gowalla

	Yelp	Foursquare	Gowalla

USSTC	0.0272	0.0301	0.0320
MEAP-T	0.0289	0.0333	0.0348
LOCABAL+	0.0314	0.0357	0.0373
Proposed	0.0325	0.0371	0.0372

جدول ۳. مقادیر Recall@10 بر روی مجموعه‌داده‌های Yelp، Foursquare و Gowalla

	Yelp	Foursquare	Gowalla
USSTC	0.190	0.296	0.301
MEAP-T	0.232	0.321	0.326
LOCABAL+	0.289	0.342	0.341
Proposed	0.305	0.343	0.355

جدول ۴. مقادیر F1@10 بر روی مجموعه‌داده‌های Yelp، Foursquare و Gowalla

	Yelp	Foursquare	Gowalla
USSTC	0.0476	0.0546	0.0578
MEAP-T	0.0514	0.0603	0.0629
LOCABAL+	0.0566	0.0647	0.0672
Proposed	0.0587	0.0670	0.673

۶- نتیجه‌گیری

در حوزه سیستم‌های پیشنهاددهنده، یک راهکار پیشنهادی برای بهبود رویکرد پالایش مشارکتی سنتی و حل مشکلات آن، اعمال پارامتر اعتماد در الگوریتم تخمین ارزشیابی کاربر است. این مقاله یک مدل انتشار اعتماد جدید برای افزایش صحت در سیستم‌های پیشنهاددهنده ارائه می‌کند. مدل پیشنهادی چهار پارامتر را در محاسبه مقدار اعتماد غیرمستقیم در نظر می‌گیرد که عبارتند از فاصله اجتماعی، موقعیت مکانی، زمان و زمینه. همچنین در این مدل، مقدار پارامتر اطمینان اعمال می‌شود که منجر به پیشنهادهای دقیق‌تر سیستم می‌شود. علاوه بر این، مدل پیشنهادی از یک معیار مبتنی بر درخت برای محاسبه شباهت معنایی بین زمینه‌ها استفاده می‌کند که براساس وزن‌دهی به یال‌ها در درخت زمینه برمبنای دو سناریوی مختلف عمل می‌کند. نتایج آزمایش‌های مختلف بر روی چند مجموعه‌داده شناخته‌شده نشان می‌دهد که مدل پیشنهادی باعث بهبود معنادار صحت توصیه‌ها نسبت به روش‌های مشابه موجود می‌گردد و به ویژه استفاده از این مدل می‌تواند در سیستم‌های پیشنهاد POI تأثیر قابل توجهی بر کارآمدی سیستم داشته باشد.

در راستای ادامه این تحقیق و ارتقاء آن ایده‌های مختلفی قابل طرح است. ازجمله این که بهینه‌سازی مناسب برای یافتن بهترین مقادیر پارامترهای مدل می‌تواند انجام شود. پارامترهای متعددی در مدل پیشنهادی وجود دارد که یافتن و اعمال مقدار بهینه آنها می‌تواند به بهبود نتایج سیستم منجر شود. از طرف دیگر این امکان وجود دارد که در مواردی از الگوریتم‌های فراابتکاری و تکاملی برای بهینه‌سازی پارامترها استفاده شود. پیشنهاد دیگر این است که در درخت زمینه برای یال‌های مختلف درخت، وزن متفاوتی در محاسبه شباهت زمینه‌ای بین گره‌ها در نظر گرفته شود. در این موارد هم مرحله بعدی

می‌تواند استفاده از تکنیک‌های بهینه‌سازی مناسب برای تعیین بهینه وزن‌ها باشد. همچنین می‌توان با استفاده از یک فیلترینگ زمانی- مکانی و بر اساس موقعیت مکانی و زمانی که کاربر در آن قرار دارد انتخاب‌های دقیق‌تری به کاربر پیشنهاد نمود. مثلاً در تعطیلات آخر هفته می‌توان مکان‌های در فواصل دورتر را نیز به کاربر هدف پیشنهاد کرد درصورتی‌که در بقیه روزهای هفته مکان‌های نزدیک‌تر به کاربر پیشنهاد داده می‌شود.

مراجع

[1] Elahi, M., et al., User Preference Elicitation, Rating Sparsity and Cold Start. 2018.

[2] Antolić, G. and L. Brkić. Recommender system based on the analysis of publicly available data. in 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2017. IEEE.

[3] Rubens, N., et al., Active learning in recommender systems, in Recommender systems handbook. 2015, Springer. p. 809-846.

[4] Nilashi, M., O. Ibrahim, and K. Bagherifard, A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. Expert Systems with Applications, 2018. 92: p. 507-520.

[5] Osadchiy, T., et al., Recommender system based on pairwise association rules. Expert Systems with Applications, 2019. 115: p. 535-542.

[6] AlBanna, B., et al., Interest aware location-based recommender system using geo-tagged social media. ISPRS International Journal of Geo-Information, 2016. 5(12): p. 245.

[7] Celdrán, A.H., et al., Design of a recommender system based on users' behavior and collaborative location and tracking. Journal of Computational Science, 2016. 12: p. 83-94.

[8] Khazaei, E. and A. Alimohammadi, An automatic user grouping model for a group recommender system in location-based social networks. ISPRS international journal of geo-information, 2018. 7(2): p. 67.

[9] Gao, H., et al. Content-aware point of interest recommendation on location-based social networks. in Twenty-ninth AAAI conference on Artificial Intelligence. 2015.

[10] Tahmasbi, H., M. Jalali, and H. Shakeri. Modeling temporal dynamics of user preferences in movie recommendation. in 2018 8th International Conference on Computer and Knowledge Engineering (ICCKE). 2018. IEEE.

[11] Kefalas, P. and Y. Manolopoulos, A time-aware spatio-textual recommender system. Expert Systems with Applications, 2017. 78: p. 396-406.

[12] Fard, K.B., et al., Recommender System Based on Semantic Similarity. International Journal of Electrical & Computer Engineering (2088-8708), 2013. 3(6).

[13] Zheng, X.-L., et al., A hybrid trust-based recommender system for online communities of practice. IEEE Transactions on Learning Technologies, 2015. 8(4): p. 345-356.

[14] Sani, N.S. and F.N. Tabriz, A new strategy in trust-based recommender system using k-means clustering. International Journal of Advanced Computer Science And Applications, 2017. 8(9): p. 152-156.

[15] Nobahari, V., M. Jalali, and S.J.S. Mahdavi, ISoTrustSeq: a social recommender system based on implicit interest, trust and sequential behaviors of users using matrix factorization. Journal of Intelligent Information Systems, 2019. 52(2): p. 239-268.

[16] Shakeri, H. and B.A. GHAEMI, A confidence-aware interval-based trust model. 2012.

[17] Filali, F.Z. and B. Yagoubi, Global trust: A trust model for cloud service selection. International Journal of Computer Network & Information Security, 2015. 7(5).

[18] Ceolin, D., et al., Trust Evaluation through User Reputation and Provenance Analysis. URSW, 2012. 900: p. 15-26.
[19] Shakeri, H. and A.G. Bafghi. Propagation of trust and confidence using intervals. in 2011 International Conference for Internet Technology and Secured Transactions. 2011. IEEE.
[20] Josang, A., R. Hayward, and S. Pope. Trust network analysis with subjective logic. in Conference Proceedings of the Twenty-Ninth Australasian Computer Science Conference (ACSW 2006). 2006. Australian Computer Society.
[21] Jøsang, A., Subjective logic. 2016: Springer.
[22] Jøsang, A. and T. Bhuiyan. Optimal trust network analysis with subjective logic. in 2008 Second International Conference on Emerging Security Information, Systems and Technologies. 2008. IEEE.
[23] Ghavipour, M. and M.R. Meybodi, Trust propagation algorithm based on learning automata for inferring local trust in online social networks. Knowledge-Based Systems, 2018. 143: p. 307-316.
[24] Jøsang, A., S. Marsh, and S. Pope. Exploring different types of trust propagation. in International Conference on Trust Management. 2006. Springer.
[25] Hasan, O., L. Brunie, and J.-M. Pierson. Evaluation of the iterative multiplication strategy for trust propagation in pervasive environments. in Proceedings of the 2009 international conference on Pervasive services. 2009.
[26] Shakeri, H. and A.G. Bafghi. RTBIMS: Accuracy enhancement in iterative multiplication strategy for computing propagated trust. in 2011 8th International ISC Conference on Information Security and Cryptology. 2011. IEEE.
[27] Jøsang, A. Artificial reasoning with subjective logic. in Proceedings of the second Australian workshop on commonsense reasoning. 1997. Citeseer.
[28] Urena, R., et al., A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. Information Sciences, 2019. 478: p. 461-475.
[29] Savage, N.S., et al., I'm feeling loco: A location based context aware recommendation system, in Advances in Location-Based Services. 2012, Springer. p. 37-54.
[30] Ying, H., et al., Time-aware metric embedding with asymmetric projection for successive POI recommendation. World Wide Web, 2019. 22(5): p. 2209-2224.
[31] Rafailidis, D. and A. Nanopoulos, Modeling users preference dynamics and side information in recommender systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2015. 46(6): p. 782-792.
[32] Adomavicius, G. and A. Tuzhilin, Context-aware recommender systems, in Recommender systems handbook. 2011, Springer. p. 217-253.
[33] Lathia, N., et al. Temporal diversity in recommender systems. in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. 2010.
[34] Xiong, L., et al. Temporal collaborative filtering with bayesian probabilistic tensor factorization. in Proceedings of the 2010 SIAM international conference on data mining. 2010. SIAM.
[35] Roy, F., A Comparative Analysis of Different Trust Metrics in User-User Trust-Based Recommender System. 2020.
[36] Bedi, P., Combining trust and reputation as user influence in cross domain group recommender system (CDGRS). Journal of Intelligent & Fuzzy Systems, 2020. 38(5): p. 6235-6246.
[37] Richa and P. Bedi, Trust and Distrust based Cross-domain Recommender System. Applied Artificial Intelligence, 2021. 35(4): p. 326-351.

[38] El Yebdri, Z., et al., Context-aware recommender system using trust network. Computing, 2021: p. 1-19.
[39] Wu, Z. and M. Palmer. V Verbs semantics and lexical selection. in Proceedings of the 32nd annual meeting on Association for Computational Linguistics. 1994.
[40] Yelp Challenge Dataset. 2018.
[41] Liu, Y., et al., An experimental evaluation of point-of-interest recommendation in location-based social networks. Proceedings of the VLDB Endowment, 2017. 10(10): p. 1010-1021.
[42] Cho, E., S.A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. 2011.
[43] Ardissono, L. and N. Mauro, A compositional model of multi-faceted trust for personalized item recommendation. Expert Systems with Applications, 2020. 140: p. 112880.