دوفصلنامه محاسبات و سامانه‌های توزیع شده
www.jdcs.ir

توسعه‌ی بازاریابی با توصیه‌گر گروه‌های پیام‌رسان‌ها

سید علی هاشمی[1]، محمدعلی زارع چاهوکی[2]
[1] دانشجوی دکتری، گروه مهندسی کامپیوتر، پردیس فنی، دانشگاه یزد، یزد
[2] استادیار، گروه مهندسی کامپیوتر، پردیس فنی، دانشگاه یزد، یزد*

چکیده

امروزه شبکه‌های اجتماعی و پیام‌رسان‌ها عضو جداناپذیری از زندگی روزمره هستند. روزانه داده‌های باارزش زیادی در این محیط منتشر می‌شود. بازاریاب‌ها با استفاده از این داده‌ها می‌توانند بهتر از قبل جامعه‌ی هدف را پیدا کنند. از دیرباز کاربران در پیام‌رسان‌ها، گروه‌های اجتماعی تشکیل می‌دادند و بحث‌های خود را در آنها دنبال می‌کردند. پس از گسترش گوشی‌های هوشمند، پیام‌رسان‌های نوظهور مانند واتساپ و وایبر استفاده وسیع تری پیدا کردند. در ایران به دلیل پارامترهای اجتماعی، پیام‌رسان‌ها جایگاهی در سطح شبکه‌های اجتماعی پیدا کرده‌اند. خلاء عمده‌ای که در پیام‌رسانی مثل تلگرام در ارتباط با تبلیغات و بازاریابی وجود دارد، عدم امکان جستجوی گروه‌های اجتماعی و رتبه‌بندی مناسب آنها می‌باشد. در این مقاله هدف ما ارائه‌ی توصیه‌گری برای گروه‌های پیام‌رسان‌ها است به نحوی که بازاریاب پس انتخاب یک یا چند گروه به عنوان جامعه‌ی هدف، گروه‌های زیادی را به عنوان توصیه دریافت کند. توصیه‌گر پیشنهادی در سامانه ایده‌کاو که ابزاری برای بازاریابی مبتنی بر اطلاعات منتشرشده در پیام‌رسان‌هاست پیاده‌سازی شده‌است.

کلمات کلیدی: شبکه‌های اجتماعی، پیام‌رسان، سامانه‌های توصیه‌گر، بازیابی اطلاعات، بازاریابی مبتنی بر پیام‌رسان‌ها

Market Expansion via Instant Messengers' Groups Recommender System

Ali Hashemi, Mohammad Ali Zare Chahooki
Yazd University, Yazd, Iran

Abstract

Nowadays, social networks and instant messengers are inseparable parts of day-to-day life. A huge amount of valuable information is shared there. By means of this information, marketers can better target their potential audiences. Users used to create social groups and follow discussions. With the advent of smart phones, next generation of instant messengers like Whatsapp and Viber became popular. Because of social parameters in Iran, an instant messenger is utilized same as social networks. An instant messenger like Telegram provides neither a marketing platform nor the possibility to search among groups. In this paper we propose a method to search and rank instant messenger groups by query. Then we extend the retrieved group by means of our novel recommender system. Our method has been implemented in IdeKav which is a tool for instant messenger marketing.

تاریخچه مقاله:

تاریخ ارسال:۹۷/۳/۱۵
تاریخ اصلاحات:۹۷/۴/۳۱
تاریخ پذیرش:۹۷/۵/۲
تاریخ انتشار:**۹۷/۵/۱۵**

Keywords:

Online social network
Information retrieval
Recommender system

روش ارجاع به مقاله:
س.ع.هاشمی، م.ع.زارع چاهوکی، توسعه‌ی بازاریابی با توصیه‌گر گروه‌های پیام‌رسان‌ها، دوفصلنامه محاسبات و سامانه های توزیع شده، سال اول، شماره اول، ص ۲۱-۳۰، سال انتشار ۱۳۹۷.

* محمد علی زارع چاهوکی : chahooki@yazd.ac.ir

۱ – مقدمه

تبلیغات پیام‌محور به دلیل سادگی و سرعت وصول بالا همواره مورد توجه بازاریاب‌ها بوده‌است. تبلیغات پیامکی، مبتنی بر نرم‌افزارهای پیام‌رسان و مبتنی بر نوتیفیکیشن نمونه‌هایی از تبلیغات پیام‌محور هستند. برخلاف بسیاری از تکنولوژی‌های دیگر که پس از مدتی طرفداران خود را از دست می‌دهند یا جایگزین می‌شوند، تبلیغات پیام‌محور همچنان توسط مردم سطح جهان استفاده می‌شود.

پیام‌رسان‌هایی مانند واتساپ و تلگرام جهت حفظ کیفیت شبکه‌ی خود، تلاش زیادی برای جلوگیری از ارسال پیام‌های تبلیغاتی به کاربران می‌کنند. ارسال پیام تبلیغاتی به کاربران دیگر باعث محدودشدن یا حذف حساب کاربری فرد متخلف خواهد شد. اگر تبلیغات ارسال‌شده هدفمند و متناسب با نیاز کاربر دریافت‌کننده باشند، پیام تبلیغاتی کاربر را آزرده نمی‌کند بلکه برای او مفید خواهد بود.

هدف از این تحقیق کمک به ایجاد سامانه‌ای جهت تبلیغات هدفمند در شبکه‌های پیام‌رسان است. در این مقاله سامانه‌ی توصیه‌گر[1] ایده‌کاو معرفی شده‌است که سعی دارد گروه‌های مرتبط با گروه‌های مورد نظر تبلیغ‌دهنده را به او پیشنهاد دهد.

سامانه‌های توصیه‌گر زیر مجموعه‌ای از سامانه‌های فیلتر اطلاعات هستند. هدف آن‌ها فیلترکردن اطلاعات به‌گونه‌ای است که فقط اطلاعات مطابق به سلیقه‌ی کاربر باقی بماند. توصیه‌گرها در انواع سامانه‌ها قابل استفاده هستند. توصیه‌ی فیلم، اخبار، صفحات وب، دوستان در شبکه‌ی اجتماعی و فرصت سرمایه‌گذاری از نمونه‌های کاربردی هستند. به طور معمول توصیه‌گر لیستی از پیشنهادات را به کاربر ارائه می‌دهد. این پیشنهادات با توجه به علایق کاربر که در گذشته توسط سامانه فراگرفته‌شده‌اند توصیه می‌شود.

امروزه اکثر پیام‌رسان‌ها امکان ایجاد گروه را فراهم کرده‌اند. کاربران با عضویت در گروه‌های مختلف به بحث و تبادل نظر در رابطه با موضوعات مورد علاقه‌ی خود می‌پردازند. با توجه به تعدد این گروه‌ها و عدم وجود امکان جستجوی گروه در اکثر پیام‌رسان‌ها، کاربران برای یافتن گروه‌های جدید مورد علاقه‌ی خود دچار مشکل هستند. در این پژوهش راهکاری برای توصیه‌ی گروه‌های جدید مرتبط به کاربران پیام‌رسان‌ها ارائه شده‌است.

چالش‌های متعددی در طراحی و پیاده‌سازی سامانه‌های توصیه‌گر وجود دارد. برخی از این چالش‌ها تا حد قابل‌قبولی در پژوهش‌های گذشته مرتفع شده‌اند. برخی دیگر از این چالش‌ها در حال حاضر موضوع پژوهشی هستند. چالش‌هایی نظیر حجم بالای داده‌ها، نوسان داده‌ها[2] و گوناگونی داده‌ها[3] در حال حاضر بسیار مورد توجه هستند [۱]. از طرف دیگر نحوه‌ی ارزیابی پیشنهادات توصیه‌گرها نیز چالش بزرگی است. پیشنهادهای ارائه‌شده توسط توصیه‌گرها از منظرهای گوناگون قابل بررسی و ارزیابی هستند. به عنوان مثال غیرمترقبه‌بودن[4] پیشنهادات از زوایای مختلفی قابل بررسی بوده و یکی از چالش‌های ارزیابی محسوب می‌شود [۲].

[1] Recommender System

[2] Data Volatility
[3] Data Variety
[4] Serendipity

در این پژوهش توصیه‌گری مبتنی بر گراف عضویت کاربران برای اولین بار ارائه‌شده است. این توصیه‌گر مقیاس‌پذیر[5] بوده و قابلیت تولید پیشنهادات را به صورت برخط دارد. در فصل ۲ مروری بر سامانه‌های توصیه‌گر با تمرکز بر شبکه‌های اجتماعی خواهیم‌داشت. در فصل ۳ روش پیشنهادی‌مان را توضیح خواهیم‌داد. فصل ۴ مربوط به نتایج آزمایش‌های تجربی است. در فصل ۵ نتیجه‌گیری می‌کنیم و اشاره‌ای به پژوهش‌های آتی خواهیم‌داشت.

۲ – مروری بر سامانه‌های توصیه‌گر

الگوریتم‌های مختلفی برای توصیه‌گرها ارائه شده‌است. برخی از مدل‌های ارائه‌شده اساس کار مدل‌های دیگر هستند. این مدل‌های پایه به چهار دسته‌ی مبتنی بر محتوا، مشارکتی[6]، مبتنی بر دانش و ترکیبی تقسیم می‌شوند. مدل‌های پیشرفته بر پایه‌ی یک یا چندتا از این مدل‌ها استوار هستند.

مدل‌های مبتنی بر محتوا با توجه به فعالیت‌های یک کاربر و محتوایی که فعالیت‌ها روی آن صورت گرفته‌است تصمیم‌گیری می‌کنند. در این مدل‌ها فعالیت کاربران دیگر اثری در تصمیم‌گیری ندارد.

منبع اصلی تصمیم‌گیری در مدل‌های مشارکتی فعالیت‌های عده‌ای از کاربران (غالباً مشابه) است. چالش اصلی این مدل‌ها تنک‌بودن ماتریس فعالیت‌های کاربران است. تنک‌بودن[7] ماتریس فعالیت کاربران باعث می‌شود که مقدار عددی فاصله‌های محاسبه‌شده در ماتریس همگی نزدیک‌به‌هم باشند. به همین دلیل، کاندیداهای زیادی برای توصیه‌شدن خواهیم‌داشت که نمی‌توانیم با اطمینان بهترین‌های آن‌ها را انتخاب کنیم. برای حل این مشکل، روش‌های انتسابی[8] ایجاد شده‌اند.

روش‌های متنوعی برای توصیه‌ی مشارکتی مبتنی بر گراف وجود دارد. سیم‌رنک[9] روشی عمومی برای یافتن گره‌های مشابه در گراف است [۳]. یافتن صفحات وب مشابه، دانش‌آموزان هم‌سطح کلاس و محصولات مرتبط در فروشگاه مثال‌هایی از کاربرد این روش هستند. سیم‌رنک به صورت بازگشتی[10] تشابه گره‌های گراف را محاسبه می‌کند. برای یافتن تشابه دو گره، تشابه گره‌هایی که به آن‌ها متصل هستند محاسبه می‌شود. برای محاسبه‌ی تشابه گره‌های متصل نیز، تشابه گره‌های متصل به آن‌ها محاسبه می‌شود. این کار تا زمانی ادامه می‌یابد تا به گرهی برسیم که در هر دو مسیر مشترک باشد.

مدل‌های مشارکتی نیاز به فعالیت کاربران برای پرشدن ماتریس فعالیت دارند. هر چه فعالیت کاربران بیشتر باشد، مدل‌های مشارکتی کیفیت بهتری خواهندداشت. علاوه بر این، در مدل‌های مشارکتی مشکل شروع سرد[11] وجود دارد. از طرف دیگر در برخی از کاربردها انتخاب کاربر ظرافت خاصی دارد. مثلاً در خرید خانه یا ماشین ظرافت‌های خاصی وجود دارد که مدل‌های مشارکتی یا مبتنی بر محتوا قابلیت در نظر گرفتن آن را ندارند. در این نوع خریدها خصوصیاتی وجود دارد که از نظر این مدل‌ها پنهان می‌ماند. مدل‌های مبتنی بر دانش این ظرافت‌ها را با

[5] Scalable
[6] Collaborative
[7] Sparsity
[8] Imputative
[9] SimRank
[10] Recursive
[11] Cold Start

دانشی که از قبل وارد سامانه شده‌است برای کاربر مهیا می‌کنند. بنابراین برای توصیه‌کردن خانه یا ماشین، تمامی جزئیات مربوطه در پایگاه داده‌ی متصل به سامانه وجود دارد. این پایگاه داده‌ی غنی برای مدل‌های مبتنی بر دانش ضروری است.

هر یک از سه مدل فوق‌الذکر معایب و مزایایی داشته و در شرایط خاصی بهتر عمل می‌کنند. مدل‌های ترکیبی، با ادغام این مدل‌ها، در هر شرایطی از مدل خاصی استفاده می‌کنند یا ترکیبی از توصیه‌های مدل‌های مختلف را به کاربر نشان می‌دهند.

مدل‌های پایه، اساس کار مدل‌های پیشرفته هستند. مدل‌های پیشرفته با تغییر، ترکیب و یا تقویت الگوریتم‌های پایه، سعی در بهبود پیشنهادها دارند. دینگ و لی مدلی با دیدگاه مبتنی بر تازگی برای سامانه‌های توصیه‌گر ارائه کرده‌اند [۴]. در این مدل برای هر فعالیت با توجه به زمان آن وزنی در نظر گرفته شده‌است. هر چه فعالیتی تازه‌تر باشد، وزن آن بیشتر است. در نهایت وزن هر فعالیت در مقدار آن ضرب شده و به عنوان مقدار نهایی در نظر گرفته‌می‌شود.

با افزایش استفاده‌ی کاربران از اینترنت و تولیدشدن داده‌های حجیم با سرعت زیاد، تمرکز پژوهشگران بر روی مدل‌های مبتنی بر بستر داده‌های حجیم[12] بیشتر می‌شود. اغلب مدل‌های پایه در بستر داده‌های حجیم پیاده‌سازی شده‌اند. مثلاً آقای ژائو و همکاران مدل مشارکتی را در این بستر ایجاد کرده‌اند [۵]. آقای چین و همکاران نیز کتابخانه‌ای در این بستر پیاده‌سازی کرده‌اند که از ماتریس فعالیت کاربران فاکتور می‌گیرد [۶].

توصیه‌گر یوتیوب نیز در بستر داده‌های حجیم ایجاد شده‌است [۷]. محاسبات توصیه‌گر یوتیوب به صورت آفلاین انجام می‌شود. تمامی توصیه‌ها به صورت دوره‌ای برای کاربران ایجاد می‌شوند و زمانی که کاربر وارد سامانه می‌شود، این توصیه‌ها فقط بازیابی می‌شوند. الگوریتم توصیه‌گر یوتیوب در دو مرحله کار می‌کند. مرحله‌ی اول ویدئوهای کاندیدا را پیدا کرده و مرحله‌ی دوم آن‌ها را رتبه‌بندی می‌کند. این الگوریتم، مدلی مبتنی بر محتوا است.

آقای ژو [۸] سامانه‌ی توصیه‌گری مبتنی بر شبکه‌های اجتماعی پیشنهاد کرده‌است که در آن ترجیحات کاربران، ارتباطات اجتماعی و روابط بین کاربران و دیگر اجزای شبکه در محاسبه‌ی تشابه در نظر گرفته می‌شوند. برای کاهش تنک‌بودن داده‌ها و حل مشکل شروع سرد از فاکتورگیری ماتریس استفاده شده‌است. همچنین از یک روش خوشه‌بندی مبتنی بر KHM[13] و PSO[14] برای کاهش حساسیت این روش بر حالت اولیه‌ی سیستم استفاده می‌شود.

آقای احمدیان و همکاران [۹] سامانه‌ی توصیه‌گری مبتنی بر گراف ارتباطات کاربران پیشنهاد داده‌اند. برای ایجاد گراف اولیه، نزدیک‌ترین همسایه‌های هر کاربر با روش خوشه‌بندی پیدا شده‌اند. با استفاده از روشی دوره‌ای، مراکز خوشه‌ها بهبود می‌یابند. در نهایت همسایه‌های بی‌کیفیت با مدل اعتمادمحور حذف می‌شوند.

[12] Big Data

[13] K-Harmonic Means
[14] Particle Swarm Optimization

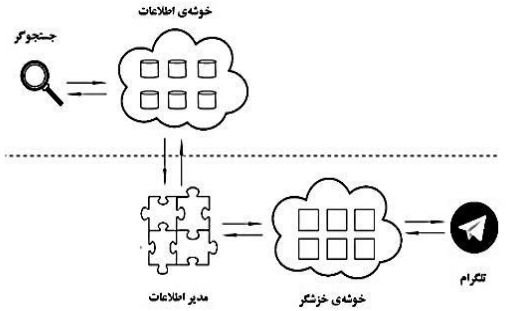
در روشی که توسط خانم زرین‌کلام و همکاران ارائه شده‌است [۱۰] گرافی مبتنی بر مدل فعالیت‌های کاربران ایجاد می‌شود. مدل فعالیت کاربران شامل فعالیت‌های عمومی وی در موضوعات مختلف، ارتباطاتش با دیگر کاربران و ارتباط موضوعات مورد علاقه‌اش است. هدف از تشکیل این گراف استخراج علایق پنهان کاربران در شبکه‌ی اجتماعی است. علایق پنهان کاربران با توجه به اصل هموفیلی[15] و ساختار ناهمگون گراف استخراج خواهدشد.

۳ – روش پیشنهادی

شبکه‌های اجتماعی و نرم‌افزارهای پیام‌رسان منبع عظیمی از اطلاعات هستند. کاربران در موضوعات متنوع اظهار نظر می‌کنند. از این اطلاعات می‌توان علایق هر فرد را مشخص کرد. ایده‌کاو[16] در حال حاضر از اطلاعات گروه‌های تلگرام برای این کار استفاده می‌کند. در تلگرام گروه‌های زیادی به زبان فارسی وجود دارند. موضوع فعالیت این گروه‌ها بسیار گسترده است. کسب‌وکارهای زیادی نیز در تلگرام مشغول به فعالیت هستند. کاربران با توجه به موضوعات مورد علاقه‌ی خود در این گروه‌ها عضو می‌شوند.

۳-۱- سامانه‌ی ایده‌کاو

در نگاه کلی، ایده‌کاو از سه بخش خزش‌گر[17]، مدیر اطلاعات و جستجوگر تشکیل شده‌است. خزش‌گر با شنود گروه‌های تلگرامی، داده‌های آن‌ها را به صورت لحظه‌ای ذخیره می‌کند. داده‌های خزش‌گر برای پردازش به بخش مدیر اطلاعات ارسال می‌شود. این داده‌ها به صورت آنلاین پردازش شده و اطلاعات مستخرج از آن به شکل مناسبی ذخیره‌سازی می‌شود. جستجوگر ایده‌کاو این اطلاعات را جستجو و رتبه‌بندی می‌کند. شکل ۱ شمای کلی معماری ایده‌کاو را نمایش می‌دهد.



(شکل-۱): شمای کلی معماری ایده‌کاو

(Figure-1): IdeKav overall architecture

خزش‌گر ایده‌کاو بر پایه‌ی پروتکل MTPROTO به زبان پایتون پیاده‌سازی شده‌است. هر شماره‌ی سیم‌کارت یک نمونه[18] از خزش‌گر را اجرا می‌کند. نمونه‌های خزش‌گر در کنار هم استخر[19] خزش‌گر را ایجاد می‌کنند. استخر خزش‌گر با برنامه‌ریز اجرای نمونه‌های خزش‌گر را مدیریت می‌کند. هر نمونه‌ی خزشگر با تخصیص حافظه‌ی اندکی قابل اجراست. استخر خزشگر مقاوم در برابر خطا[21] است. خطای نمونه‌ها باعث از کار افتادن استخر نمی‌شود.

داده‌های به‌دست‌آمده از خزش‌گر نظیر پیام‌ها و اطلاعات عضویت افراد در گروه‌ها به بخش مدیر اطلاعات ارسال می‌شود. مدیر اطلاعات از این داده‌ها، اطلاعات مورد نیاز را استخراج کرده و با ساختمان

15 Homophily
16 http://idekav.ir
17 Crawler

18 Instance
19 Pool
20 Scheduler
21 FaultTolerant

داده‌ای مناسب ذخیره‌سازی می‌کند. لینک‌های جدید در این بخش پیدا شده و برای عضویت به خزش‌گر ارسال می‌شود. محتوای گروه‌ها نیز همین‌جا بررسی می‌شود. گروه‌هایی که محتوای اسپم یا کم‌ارزش دارند شناسایی شده و برای ترک‌کردن به خزش‌گر فرستاده می‌شود. اطلاعات مورد نیاز دیگر نیز در این بخش نمایه‌سازی می‌شوند. ذخیره‌سازی اطلاعات بلادرنگ بوده و به صورت توزیع‌شده بین سرورهای خوشه[۲۲] انجام می‌شود. ذخیره‌سازی در تکه‌های[۲۳] کوچک به تعداد زیاد انجام می‌شود.

جستجوگر ایده‌کاو بر پایه‌ی نمایه‌ساز[۲۴] و جستجوگر لوسین[۲۵] پیاده‌سازی شده‌است. جستجو بر روی سه نمایه‌ی گروه‌ها، کاربران و پیام‌ها قابل انجام است. معماری ذخیره‌سازی اطلاعات، نمایه‌ی معکوس[۲۶]است. نمایه‌ی معکوس سرعت بالایی در بازیابی اطلاعات دارد. علاوه بر این، امکان اعمال فیلترهای مختلف بر روی نمایه‌ی معکوس وجود دارد. معماری جستجوی اطلاعات، TFIDF است. در این معماری کلمات مهم متن به سرعت پیدا شده و از کلمات عمومی متمایز می‌شوند.

جستجو با ارسال پرس‌وجو[۲۷] می‌تواند نتایج مرتبط با موضوع مورد نظر کاربر را بازیابی کند. عموماً در این نوع پرس‌وجو، نتایجی که در رتبه‌ی بالا قرار دارند بسیار مرتبط هستند ولی هر چه پایین‌تر برویم ارتباط پرس‌وجو با اسناد بازیابی‌شده کمتر می‌شود. معماری اصلی مورد استفاده برای جستجوی با پرس‌وجو، معماری نمایه‌ای است. در این معماری بازیابی نتایج مرتبط سریع است ولی هرچه پایین‌تر برویم بازیابی زمان بیشتری می‌گیرد. این مشکلات باعث می‌شود جستجوی مبتنی بر پرس‌وجو به تنهایی نتواند جوامع هدف بزرگ را بازیابی کند.

۴– توصیه‌گر ایده‌کاو

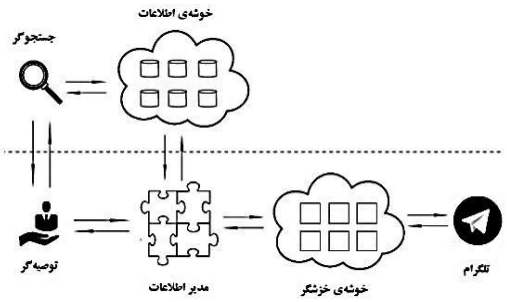
جستجوی مبتنی بر پرس‌وجو می‌تواند شروع خوبی برای یافتن جامعه‌ی هدف باشد. پس از بازیابی تعدادی نتیجه، مرتبط‌ترین اسناد با پرس‌وجو از لحاظ محتوایی پیدا می‌شوند. در صورتی که بتوانیم با روشی اسناد مشابه از منظرهای دیگر را نیز پیدا کنیم، به نتیجه‌ی بهتری خواهیم رسید. علاوه بر این در صورتی که تبلیغ‌دهنده نیاز به جامعه‌ی هدف بزرگی داشته‌باشد، جستجو با پرس‌وجو بسیار زمان‌بر خواهدبود. روشی که بتواند بدون دخالت انسانی با دقت مناسبی مرتبط‌ترین افراد را به تبلیغ‌دهنده معرفی کند این‌جا مورد نیاز است. این روش باید علاوه بر ارتباط محتوایی، روابط دیگر در شبکه‌های اجتماعی را نیز در نظر بگیرد.

در صورتی که تبلیغ‌دهنده به ادبیات و کلمات کلیدی جامعه‌ی هدفش آشنایی کامل نداشته‌باشد، تشابه محتوایی شاید نتواند جامعه‌ی هدف او را به طور کامل شناسایی کند. در بسیاری از موارد جامعه‌ی هدف را نمی‌توان با کلمات کلیدی مشخص کرد. توصیه‌گر ایده‌کاو راه‌حلی پیشنهادی برای حل این مشکل است. هدف این توصیه‌گر یافتن جامعه‌ی هدف تبلیغ‌دهنده به سادگی است. ورودی این توصیه‌گر یک یا چند مورد بسیار مطلوب برای تبلیغ‌دهنده است.

[22] Cluster
[23] Shard
[24] Indexer
[25] Lucene
[26] Inverted Index
[27] Query

توصیه‌گر با توجه به گروه‌های ورودی، گروه‌های مشابه را به عنوان خروجی برمی‌گرداند.

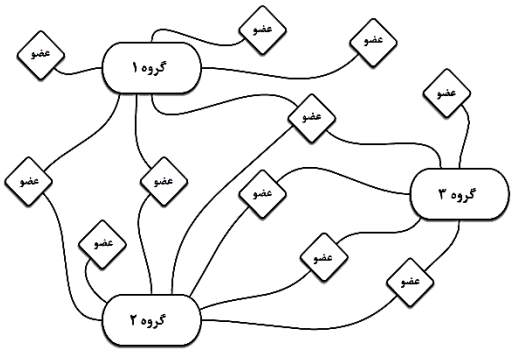
توصیه‌گر ایده‌کاو به جستجوگر و مدیر اطلاعات متصل خواهدبود. توصیه‌گر از خوشه‌ی اطلاعات استفاده خواهدکرد تا اطلاعات مورد نیاز برای توصیه را به‌دست آورد. جستجوگر نیز امکان جستجوهای خاص‌منظوره را به توصیه‌گر می‌دهد. شکل ۲ محل قرارگیری توصیه‌گر در معماری ایده‌کاو را نشان می‌دهد.



(شکل-۲): توصیه‌گر در معماری ایده‌کاو

(Figure-2): Recommender module

توصیه‌گر پیشنهادی بر پایه گراف ناهمگون عضویت کاربران کار می‌کند. ابتدا اعضای گروه ورودی در گراف بازیابی می‌شوند. هر یک از این اعضا در گروه‌های دیگری نیز ممکن است عضو باشند. تمامی این گروه‌های جدید به همراه تعداد تکرار آن‌ها (اعضای مشترک) از گراف بازیابی می‌شود. تعداد کل اعضای هر یک از گروه‌های بازیابی‌شده نیز مشخص است. هر چه نسبت تعداد تکرار یک گروه به تعداد اعضایش بیشتر باشد، آن گروه رتبه‌ی بالاتری در توصیه‌گر خواهدگرفت. به‌این‌ترتیب توصیه‌گر با توجه به عضویت کاربران در گروه‌ها پیشنهاد گروه‌های جدید ارائه می‌کند.



(شکل-۳): نمونه‌ای از گراف ناهمگون عضویت برای ۳ گروه

(Figure-3): Sample of heterogeneous membership graph for 3 groups

شکل ۳ نمونه‌ای از گراف ناهمگون عضویت را برای ۳ گروه نشان می‌دهد. اعضایی که فقط در یک گروه عضو هستند برگ این گراف محسوب شده و در محاسبات صرف‌نظر می‌شوند.

۵ – نتایج تجربی

قبل از بررسی نتایج تجربی روش پیشنهادی آمار استخراج‌شده از تلگرام توسط ایده‌کاو را بررسی می‌کنیم. خزش‌گر ایده‌کاو فقط در گروه‌های عمومی که لینک آن‌ها در گروه‌های دیگر منتشر شده‌است عضو می‌شود. جدول ۱ آمار به‌دست‌آمده از خزش گروه‌های تلگرامی را نشان می‌دهد.

بررسی آمار ارائه‌شده در جدول ۱ نشان می‌دهد که کاربران در تلگرام به خروج[۲۸] از گروه‌های فعلی و عضویت در گروه‌های جدید تمایل دارند. میانگین تعداد گروه‌هایی که هر کاربر در آن عضو است ۲/۴۸ می‌باشد. هر کاربر نیز از ۲/۳ گروه‌هایش به طور میانگین خارج می‌شود. از این آمار می‌توان نتیجه گرفت که تقریباً هر کاربر از همه‌ی گروه‌هایش به مرور

[28] Leave

زمان خارج شده و به گروه‌های جدید می‌پیوندد. این نوع رفتار کاربران می‌تواند به دلیل عدم وجود امکان جستجوی جامع در تلگرام باشد. کاربران برای رفع نیازهای مقطعی وارد گروهی می‌شوند و سپس از آن خارج می‌شوند. این موضوع اهمیت وجود موتور جستجویی برای گروه‌های تلگرامی را به وضوح نشان می‌دهد.

ما پیام‌های شامل کلمات و موضوعات خاص یا لینک‌های متعدد را اسپم تشخیص می‌دهیم. نکته‌ی دیگری که در مورد این آمار قابل توجه است، حجم بالای پیام‌های اسپم می‌باشد. روزانه از بین ۲۵ میلیون پیام دریافتی، ۱۹ میلیون از آن‌ها اسپم تشخیص داده می‌شوند. این در حالی‌است که خزش‌گر ما از گروه‌هایی که محتوای کاملاً اسپم دارند خارج می‌شود و پیام‌های این گروه‌ها در این آمار وجود ندارند. بسیاری از گروه‌ها برای مدیریت اسپم، ربات[29] حذف اسپم دارند. باید به این موضوع نیز اشاره کنیم که خزش‌گر ما پیام‌هایی که توسط ربات‌ها پاک می‌شوند را نیز در این آمار در نظر گرفته‌است.

(جدول-۱): آمار ایده‌کاو از تلگرام

(Table-1): IdeKav's Telegram statistics

میانگین تعداد پیام‌های خزش‌شده روزانه (تقریبی)	۲۵ میلیون
سطر بالا پس از حذف پیام‌های spam (تقریبی)	۶ میلیون
میانگین تعداد پیام‌های هر گروه در روز	۱۵۹
سطر بالا پس از حذف spam	۳۷
میانگین طول پیام (تعداد کاراکتر)	۱۷۳
تعداد گروه‌های تحت پوشش	۱۵۳۶۱۰
تعداد کاربران تحت پوشش	۴۰۴۲۵۰۸۳
میانگین تعداد اعضای هر گروه	۶۵۳
میانگین تعداد گروه‌هایی که هر کاربر عضو است	۲/۴۸
تعداد کاربرانی که حداقل از یک گروه خارج شده‌اند	۲۵۹۴۹۹۱۳
میانگین تعداد گروهی که هر کاربر از آن خارج شده	۲/۳
میانگین تعداد گروه‌های هر مدیر گروه	۱/۲۶

روش پیشنهادی این پژوهش در ایده‌کاو پیاده‌سازی شده و برای عموم قابل دسترس است. تعدادی از پیشنهادهای توصیه‌گر ایده‌کاو را برای نمونه در اینجا آورده‌ایم. جدول ۲ گروه‌های پیشنهادشده برای گروه «تالار گفتمان شطرنج» را نشان می‌دهد. جدول ۳ گروه‌های پیشنهادی با ورودی گروه «پیمانکاران» است. جدول ۴ نیز شامل گروه‌هایی است که به واسطه‌ی کلیک بر روی گروه «پرورش ماکیان ایران» پیشنهاد شده‌اند.

(جدول-۲): پیشنهادات گروه «تالار گفتمان شطرنج»

(Table-2): Recommendation list for "Chess Forum"

گروه شطرنج استاد بزرگ ایران
Best Game
گروه انجمن شطرنج ایران
انتقادات و پیشنهادها در شطرنج
♚♚ شطرنج خراسان خراسان ♚♚
باشگاه شطرنج ذهن برتر
گروه شطرنج بازان ایران
Online Worldchess
شروع بازی شطرنج
✿✿جام دوستی‌شطرنج ایران✿✿

[29] Telegram Bot

(جدول-۳): پیشنهادات گروه «پیمانکاران»

(Table-3): Recommendation list for "Contractors"

جامعه پیمانکاران شهرداریهای کشور
عمران راهسازی
صنعت ساختمان تهران سازان
ماشین الات راه ساختمان معدن
آبر گروه تبلیغات تجاری سراسر کشور
رهبری خانگی کشوری
جامعه فعالان املاک و مستغلات (۲۳/۱۲/۱۳۹۴)
آبر گروه تبلیغات تجاری سراسری
هلدینگ ساختمانی کلکسیون مشهدمقدس
جامعه صنعت ساختمان

(جدول-۴): پیشنهادات گروه «پرورش ماکیان ایران»

(Table-4): Recommendation list for "Iran Poultry Farming"

خریدوفروش ماکیان زینتی و پرنده وسگ
جوجه کشی تخصصی
کرک کلی سید در شمال کشور
دستگاه جوجه کشی هدهد خراسان
خبرنامه دام ،طیور وآبزیان ایران
تلاشگران صنعت طیور ایران
بازار خرید و فروش طیور محلی
ماکیان طلایی
پرورش مرغ بومی(اصولی و اقتصادی)
فارم داران مرغ و بلدرچین ایران

۶- نتیجه‌گیری و پژوهش‌های آتی

در این مقاله روشی جدید برای توصیه‌ی گروه‌ها در شبکه‌های اجتماعی ارائه شده‌است. ورودی توصیه‌گر پیشنهادی یک یا چند گروه مطلوب است. خروجی توصیه‌گر، گروه‌هایی هستند که با توجه به ورودی توصیه می‌شوند. این روش بر مبنای گراف ناهمگون عضویت کاربران در گروه‌ها، عمل می‌کند. تکرار گروه‌های دیگر اعضای هر گروه محاسبه شده و به نسبت تعداد اعضای آن‌ها رتبه‌بندی می‌شوند. در واقع گروه‌هایی که بیشترین عضو مشترک را دارند، بیشترین شانس را برای پیشنهادشدن در توصیه‌گر خواهند داشت.

مزیت این روش سرعت، سادگی و قابلیت مقیاس‌پذیری آن است. روش پیشنهادی در سامانه‌ی ایده‌کاو پیاده‌سازی شده و هم‌اکنون در دسترس عموم قرار دارد. نتایج تجربی نشان می‌دهد که این روش بسیار خوب عمل می‌کند.

در حال حاضر این روش با همه‌ی اعضای گروه به طور یکسان برخورد می‌کند. در مطالعات آتی می‌توان بین مدیران گروه، اعضای فعال گروه و اعضای غیر فعال گروه تفاوت قائل شد. همچنین امکان دخیل‌کردن متن پیام‌های هر کاربر در این روش وجود دارد. می‌توان کاربرانی که در مورد موضوعی مرتبط با موضوع گروه صحبت کرده‌اند را با دیگران متمایز کرد.

۷ – مراجع

[1] M. Eirinaki, J. Gao, I. Varlamis and K. Tserpes, "Recommender Systems for Large-Scale Social Networks: A review of challenges and solutions", Future Generation Computer Systems, vol. 78, pp. 413-418, 2018.

[2] D. Kotkov and S. Wang, "Challenges of serendipity in recommender systems", in Proceedings of the 12th International conference on web information systems and technologies, 2016.

[3] G. Jeh and J. Widom, "SimRank: a measure of structural-context similarity", in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 538-543.

[4] Y. Ding and X. Li, "Time weight collaborative filtering", ACM International Conference on Information and Knowledge Management, 2005, pp. 485-492.

[5] Z. Zhao and M. Shang, "User-based collaborative-filtering recommendation algorithms on hadoop", in Third International Conference on Knowledge Discovery and Data Mining WKDD'10, 2010, pp. 478-481.

[6] W. Chin and B. Yuan, "LIBMF: a library for parallel matrix factorization in shared-memory systems", The Journal of Machine Learning Research, vol. 17, no. 1, pp. 2971-2975, 2016.

[7] J. Davidson, B. Liebald, J. Liu and P. Nandy, "The YouTube video recommendation system", in Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 293-296.

[8] C. Xu, "A novel recommendation method based on social network using matrix factorization technique ", Information Processing and Management, vol. 54.3, pp. 463-474, 2018.

[9] S. Ahmadian, M. Meghdadi, and M. Afsharchi, "A social recommendation method based on an adaptive neighbor selection mechanism", Information Processing and Management, vol. 54.4, pp. 707-725, 2018.

[10] F. Zarrinkalam, M. Kahani, and E. Bagheri, "Mining user interests over active topics on social networks", Information Processing and Management, vol. 54.2, pp. 339-357, 2018.