

پیشنهاد کاربران تلگرام بر اساس گراف عضویت و سنجش گروه‌ها با داشتن کاربرانی از جامعه هدف

داود کریم‌پور

دانشجوی کارشناسی ارشد، مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران

محمدعلی زارع چاهوکی^۱

استادیار، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران

مدیریت
اطلاعات

دوره ۷، شماره ۱

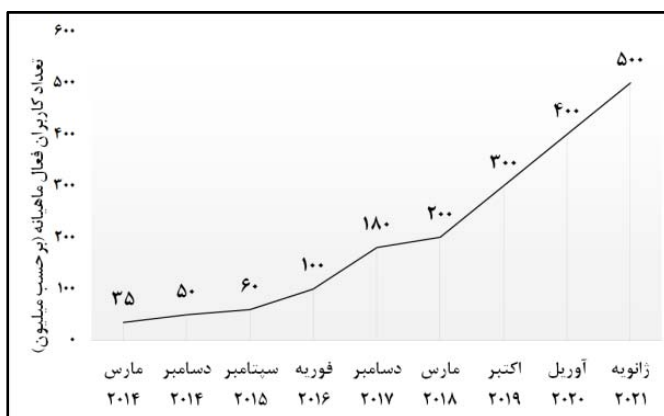
بهار و تابستان ۱۴۰۰

چکیده: تلگرام پیام‌رسانی مبتنی بر ابر است که بیش از ۵۰۰ میلیون کاربر فعال ماهیانه دارد. از ویژگی‌های ارائه‌شده در این پیام‌رسان می‌توان به ایجاد گروه، کانال، ربات و چت مخفی اشاره کرد که قابلیت‌های بیشتری از یک پیام‌رسان، در اختیار کاربران قرار داده است. تلگرام در ایران به‌عنوان شبکه اجتماعی استفاده می‌شود، اما تمامی ویژگی‌های یک شبکه اجتماعی همانند یافتن مجموعه‌ای از کاربران را ارائه نمی‌دهد. یافتن کاربران، برای بازاریابان به‌منظور پیدا کردن جامعه هدف بسیار مفید است. هدف این پژوهش، ارائه روشی برای یافتن کاربران در پیام‌رسان تلگرام است. در این مقاله، برای پیشنهاد کاربر، از گراف ارتباطی کاربران استفاده شده است. گراف ارتباطی، کاربران را از طریق گروه‌های مشترک به یکدیگر متصل می‌کند. رویکرد پیشنهادی با استفاده از گراف و متغیرهای عددی استخراج‌شده از سطح گروه‌های تلگرام، در دو بخش کمی و کیفی پنج پیشنهاد ارائه داده است. بخش کمی، گروه‌ها را بر اساس تعداد و درصد اشتراک با کاربران ورودی مدل‌سازی می‌کند. بخش کیفی نیز در ادامه بخش کمی ارائه شده است که گروه‌ها را بر اساس ویژگی تعداد مدیر در هر گروه رتبه‌بندی می‌کند. هر یک از پیشنهادها بخش کمی و کیفی برای تکمیل و رفع نواقص سایر پیشنهادها ارائه شده‌اند و هر یک را می‌توان روشی جداگانه برای پیشنهاد کاربر در نظر گرفت. داده‌های این پژوهش شامل اطلاعاتی در حدود ۷۰ میلیون کاربر و ۷۰۰ هزار گروه در تلگرام هستند. اطلاعات مد نظر، به‌وسیله سامانه ایده‌کاو و به‌صورت کاملاً قانونی توسط ابزاری که تلگرام در اختیار عموم قرار داده است، به دست آمده‌اند. در این مقاله، ارزیابی توسط گروه‌های محبوب تلگرام، با دو دسته اعتبارسنجی و آزمون، انجام شده است و نتایج آزمایش‌های تجربی تکرار نتایج و یکپارچگی روش پیشنهادی را نشان داده‌اند. روش پیشنهادی با هدف پوشش کل گراف در دو بخش کمی و کیفی، مؤثر بودن پیشنهادها و کاهش خطای پیش‌بینی را نسبت به سه پژوهش پیشین نشان داده است.

کلیدواژه‌ها: یافتن کاربران، شبکه‌های اجتماعی، بازاریابی مبتنی بر تلگرام، یافتن جامعه هدف بازاریابان

مقدمه

در یک دهه گذشته، پیام‌رسان‌ها و شبکه‌های اجتماعی بیش از سایر نرم‌افزارهای موبایلی محبوبیت پیدا کرده و به بستر ارتباطی مهمی تبدیل شده‌اند. تلگرام یک برنامه پیام‌رسان مبتنی بر رایانش ابری است که در آگوست ۲۰۱۳ به وجود آمده است. این پیام‌رسان در ۱۲ ژانویه ۲۰۲۱ اعلام کرده که از ۵۰۰ میلیون کاربر فعال ماهیانه عبور کرده است (Iqbal, 2021). با توجه به شکل ۱ که روند رشد تلگرام را به‌لحاظ تعداد کاربر ماهیانه نشان می‌دهد، این پیام‌رسان در بازه زمانی بین سال‌های ۲۰۱۸ تا ۲۰۲۱ در هر سال، ۱۰۰ میلیون کاربر فعال جدید دریافت کرده است. تلگرام امکانات مختلفی از جمله ایجاد گروه، ایجاد کانال، ساخت ربات و چت مخفی ارائه می‌دهد. ویژگی‌های گروه و کانال در این پیام‌رسان، بیشترین کاربرد را دارند، به‌طوری که بسیاری از شرکت‌های تجاری، یک کانال برای تبلیغ و یک گروه برای پشتیبانی و شرح محصولات دارند. با توجه به حجم بسیار بالای فعالیت کاربران در این پیام‌رسان، بازایی اطلاعات به‌منظور دستیابی به اطلاعات کاربران و بازایی اهمیت بسیاری پیدا کرده است.



شکل ۱. تعداد کاربر فعال ماهیانه تلگرام از مارس ۲۰۱۴ تا ژانویه ۲۰۲۱

منبع: Iqbal (2021)

بسیاری از پژوهشگران در ایران به محبوبیت تلگرام توجه نشان داده‌اند. برخی پژوهش‌های انجام‌شده عبارت‌اند از: پزشکی (Ghaffari et. al, 2017)، سیاسی (بیگدلو و هادیان، ۱۳۹۷) و زبان‌شناسی (Al Momani, 2020). هاشمی و زارع چاهوکی^۱ (۲۰۱۹) کیفیت گروه‌های تلگرام را بر اساس رفتار کاربران سنجیده‌اند. در پژوهشی دیگر، هاشمی و زارع چاهوکی (۲۰۲۱) گروه‌ها را بر اساس دریافت پرس‌وجو از کاربر، رتبه‌بندی کرده‌اند. درگهی‌نوبری، رشادتمند و نشاطی^۲ (۲۰۱۷) پیام‌های هرنامه را

1. Hashemi & Zare Chahooki

2. Nobari, Reshadatmand & Neshati

دسته‌بندی کرده‌اند. اسنفی و همکاران^۱ (۲۰۱۷) اثرگذاری کانال کتابخانه‌های ایرانی را روی کاربران سنجیده‌اند. قربان‌زاده و سعیدنیا^۲ (۲۰۱۸) نیز اثر تلگرام را بر نگرش و گفتار کاربران ایرانی بررسی کرده‌اند.

با وجود اینکه تلگرام یک پیام‌رسان است، اما در برخی کشورها از جمله ایران، به‌عنوان شبکه اجتماعی استفاده می‌شود. استفاده از تلگرام به‌عنوان شبکه اجتماعی مستلزم ارائه تمامی خدمات یک شبکه اجتماعی از جمله پیشنهاد دوست یا کاربر توسط این پیام‌رسان است. جست‌وجوی کاربران در تلگرام فقط با داشتن نام کاربری^۳ امکان‌پذیر است که آن هم بسیار محدود و در عمل دشوار است. زیرا نام کاربری هر کاربر، هیچ اطلاعاتی از علاقه وی بیان نمی‌کند. تلگرام برای جست‌وجوی کاربران، سرویس جامعی ندارد. به‌طور مثال، یک بازاریاب برای ارائه تبلیغات مربوط به محصولات خود، به‌دنبال کاربرانی مشابه با مشتریان محبوب و قدیمی است. از این رو، این مشکل، بازاریاب را بر آن داشته تا با تک‌تک کاربران گروه‌ها ارتباط برقرار کند. این کار بسیار طاقت‌فرسا و دشوار است، به‌خصوص اینکه یافتن گروه‌های مد نظر نیز بسیار دشوار است. بازاریابان به اطلاعات عضویت مشتریان خود در گروه‌ها دسترسی ندارند، بنابراین برای آنان مقدور نیست که تبلیغات خود را به تمامی کاربران گروه‌ها ارسال کنند. برخی از بازاریابان گروهی تشکیل داده و تبلیغات خود را در آن گروه قرار می‌دهند. آنان در این روش، از کاربران کمی شروع می‌کنند و از آنان می‌خواهند تا با عضو کردن کاربران جدید، هزینه‌ای دریافت کنند. این کار باعث می‌شود تا کاربران نامرتبب زیادی عضو شوند و گروه را سریع ترک کنند، در حالی که بازاریاب برای کاربران نامرتبب، هزینه زیادی پرداخت کرده است. برخی از بازاریابان نیز به‌وسیله یک ربات، پیام‌هایی را به‌طور مستقیم به کاربران مد نظر ارسال می‌کنند. این روش‌ها برای بازاریابان بسیار دشوار است، به‌خصوص اینکه ارائه آمار جعلی از اعضای گروه‌ها توسط سودجویان بسیار بالاست. بنابراین همان‌طور که بیان شد، مشکل اصلی در تلگرام این است که کاربران این پیام‌رسان روزبه‌روز در حال گسترش هستند، اما بازاریابان از کاربران برای یافتن جامعه هدف خود، هیچ اطلاعاتی ندارند. از این رو، با توجه به اینکه در این پیام‌رسان برای جست‌وجوی کاربر ابزاری جامع وجود ندارد، این مسئله مطرح می‌شود که چگونه کاربرانی مشابه با کاربران قدیمی یک بازاریاب را بیابیم؟

امروزه پیشنهاد کاربران در شبکه‌های اجتماعی و پیام‌رسان‌ها به مبحثی مهم تبدیل شده است. مسئله چالش‌برانگیز در این حوزه، این است که چگونه کاربرانی مشابه با کاربران موجود را بیابیم. برای پیشنهاد کاربران، الگوریتم‌ها و روش‌های بسیاری وجود دارند که در قالب سامانه‌های پیشنهاددهنده مطرح می‌شوند. این سامانه‌ها با توجه به میزان اطلاعاتی که از کاربران در اختیار است، به دو دسته توپولوژیکی و غیرتوپولوژیکی تقسیم می‌شوند (Kumar & Reddy, 2018). دو سامانه توپولوژیکی و غیرتوپولوژیکی را می‌توان با یکدیگر ترکیب کرد تا علاوه بر اطلاعات کاربران، از گراف آنان نیز بهره برد.

1. Asnafi, Moradi, Dokhtesmati & Naeini

2. Ghorbanzadeh & Saeednia

3. Username

پیشنهاد در سامانه‌های توپولوژیکی فقط به‌وسیله گراف انجام می‌شود. در بیشتر پیام‌رسان‌ها و شبکه‌های اجتماعی با یک الگوی کلی روبه‌رو هستیم. این الگو، گرافی از ارتباط کاربران را بیان می‌کند که هر گره در گراف، بیانگر کاربر و هر یال ارتباط مختلف کاربران را نشان می‌دهد. پیشنهاد به‌وسیله سامانه‌های غیرتوپولوژیکی با توجه به میزان اطلاعاتی که از کاربران استخراج می‌شود، به دسته‌هایی جداگانه تقسیم می‌شود. هر دسته برای هر کاربر به‌وسیله شباهت‌سنجی با کاربران دیگر و بدون استفاده از گراف، کاربرانی ارائه می‌دهد. پیشنهاد در سامانه‌هایی که روش توپولوژیکی و غیرتوپولوژیکی را با یکدیگر ترکیب می‌کنند، بدین صورت است که گره‌های گراف به‌وسیله اطلاعات کاربران برچسب زده می‌شود، سپس با پیمایش گراف و روش‌های شباهت‌سنجی، کاربرانی مشابه پیشنهاد می‌شوند.

روش ارائه‌شده در این مقاله، مبتنی بر سامانه‌های پیشنهاددهنده توپولوژیکی است. کلیت روش پیشنهادی در این مقاله بدین شرح است که ابتدا مجموعه‌ای کاربر به روش پیشنهادی داده می‌شود. هدف، دستیابی به مجموعه‌ای کاربر جدید، مشابه با کاربران ورودی است. در این مقاله برای پیشنهاد کاربران، از گرافی که ارتباط کاربران را به‌وسیله گروه‌های مشترک بیان می‌کند، استفاده شده است. با استفاده از گراف، ویژگی‌های عددی را از سطح گروه‌ها، استخراج کرده‌ایم، سپس با ارائه پنج پیشنهاد که هر یک دیگری را تکمیل می‌کند، به بیان روش پیشنهادی پرداخته‌ایم. روش پیشنهادی، دارای پنج پیشنهاد بوده و در دو بخش کمی و کیفی ارائه شده است. بخش کمی شامل دو پیشنهاد است که گروه‌ها را بر اساس تعداد و درصد کاربران مشترک آنها با مجموعه کاربر ورودی رتبه‌بندی می‌کند. بخش کیفی دارای سه پیشنهاد در ادامه پیشنهادهای بخش کمی ارائه شده است که گروه‌ها را بر اساس ویژگی تعداد مدیر در هر گروه رتبه‌بندی می‌کند. بخش کیفی نیز کاربران ورودی را از جهتی دیگر و بر اساس مدیر بودن در گروه‌ها بررسی می‌کند. در هر پیشنهاد، فهرستی از گروه‌های رتبه‌بندی‌شده به دست می‌آید که با روی هم ریختن اعضای گروه‌ها (به‌ترتیب فهرست)، مجموعه کاربران نهایی حاصل می‌شود.

داده‌های این پژوهش، اطلاعات کاربران و گروه‌های عمومی تلگرام است که به‌وسیله سامانه ایده‌کاو به دست آمده‌اند. این اطلاعات حدود ۷۰ میلیون کاربر و ۷۰۰/۰۰۰ گروه را در بر دارد. در این مقاله، برای ارزیابی روش پیشنهادی از گروه‌های محبوب تلگرام استفاده شده است. گروه‌های محبوب توسط خبره و به‌وسیله مطابقت پیام‌های هر گروه با موضوع آن، به دست آمده‌اند. روش پیشنهادی روی دو دسته اعتبارسنجی به‌منظور ایجاد مدل و آزمون به‌منظور بررسی خطای مدل، ارزیابی شده است. هر یک از دسته‌ها شامل ۳۰ گروه محبوب است. نتایج حاصل از ارزیابی، یکپارچگی روش پیشنهادی را در میزان پیش‌بینی کاربران نشان داده است. دلیل انتخاب و ارزیابی دو دسته از گروه‌ها، این است که نشان داده شود مدل روی گروه‌هایی خاص، بایاس نشده است و نتیجه برای هر دسته از گروه‌ها تغییر چندانی نخواهد کرد. پیشنهادهای ارائه‌شده در این مقاله، با سه پژوهش مرتبط پیشین، مقایسه شده است. نتایج مقایسه، مؤثر بودن پیشنهادها و کاهش خطای پیش‌بینی، نسبت به پژوهش‌های پیشین را نشان می‌دهد.

ادامه مقاله بدین ترتیب ساختاربندی شده است. ابتدا پیشینه پژوهش مرور می‌شود. در بخش بعد، به معرفی روش پیشنهادی و نحوه جمع‌آوری داده‌ها پرداخته می‌شود. در بخش انتهایی نیز نتایج و آزمایش‌های تجربی و همچنین، نتیجه‌گیری و موضوعاتی برای پژوهش‌های بعدی ارائه خواهد شد.

پیشینه پژوهش

سامانه‌های پیشنهاددهنده در بستر وب و شبکه‌های اجتماعی، به‌منظور کاهش ایجاد و پردازش پرس‌وجو به وجود آمده‌اند. این سامانه‌ها اغلب برای بازاریابی، پیشنهادهایی به کاربران از جمله فیلم، کتاب (Ricci, Rokach & Shapira, 2011)، کاربر یا دوست (Silva et. al, 2010; Jiang, Leung, & Pazdor, 2016) و جوامعی از کاربران (Han et. al, 2016; AlSuwaidan & Ykhlef, 2018) ارائه می‌دهند. در این مقاله، پیشینه پژوهش از دو جنبه بررسی شده است. ابتدا، در خصوص پژوهش‌های مرتبط با سامانه‌های توپولوژیکی مبتنی بر گراف بحث می‌شود، سپس پژوهش‌هایی که پیشنهاد کاربران را در قالب توصیه جوامع^۱ بیان کرده‌اند، مطرح می‌شوند.

سامانه‌های توپولوژیکی مبتنی بر گراف

سامانه‌های پیشنهاددهنده مبتنی بر ساختار توپولوژیکی، اغلب بر اساس گراف ارتباطی کاربران و خصوصیات ساختار شبکه ایجاد می‌شوند. سیلوا و همکاران^۲ (۲۰۱۰) روشی برای پیشنهاد دوست مبتنی بر گراف به‌وسیله الگوریتم ژنتیک در شبکه اجتماعی ارو - آرو^۳ ارائه دادند. آنان در مرحله نخست، مجموعه‌ای از کاربران کاندیدا که ارتباط بیشتری با کاربر اصلی دارند را بر اساس سه درجه جدایی انتخاب کرده‌اند. در مرحله دوم، شاخص بین هر کاندیدا و کاربر اصلی ارزیابی می‌شود، سپس از الگوریتم ژنتیک برای انتخاب بهترین ترکیب وزن استفاده می‌شود. جیانگ، لیانگ و پازدر^۴ (۲۰۱۶) مدلی بر اساس گراف کاربران به‌وسیله روش مپ‌ردیوس^۵ برای کشف گروه‌هایی از کاربران در توپیتر^۶ ارائه دادند. منطق کار آنها به این صورت است که برای مثال پس از یافتن گروهی متشکل از چهار کاربر که بیست دوست مشترک دارند، دو کاربر از چهار کاربر به کاربرانی که با دو کاربر دیگر دوست هستند، پیشنهاد می‌شود. هاشمی و زارع چاهوکی (۱۳۹۷) به‌منظور توسعه بازاریابی از طریق پیشنهاد گروه‌ها در تلگرام روشی ارائه داده‌اند. در این روش بررسی می‌شود که از هر گروه چه تعداد کاربر در بین کاربران ورودی قرار دارند و به‌صورت نزولی، گروه‌ها را فهرست می‌کند. ایراد این روش این است که گروه‌های با تعداد عضو زیاد همیشه در ابتدای فهرست قرار می‌گیرند. این کار باعث می‌شود تا گروه‌های با تعداد عضو کم، هیچ‌وقت نتوانند به ابتدای فهرست بیایند. در واقع مشکل این روش، همانند مشکل مشهور در بازاریابی اطلاعات به نام ثروتمند

1. Community Recommendation

2. Silva et. al

3. Oro-Aro

4. Jiang, Leung & Pazdor

5. MapReduce

6. Twitter

ثروتمندتر^۱ می‌شود، است که بیانگر رتبه‌بندی بهتر صفحات وب پربازدید است. کریم‌پور، زارع چاهوکی و هاشمی^۲ (۲۰۲۱) روشی به نام مهاجرت برای پیشنهاد گروه‌های تلگرامی ارائه داده‌اند. این روش کاربران را در گروه‌ها با سه وضعیت (عضو قدیمی، عضو کنونی و عضو مدیر) در نظر می‌گیرد. سپس با بررسی اینکه اعضای قدیمی هر گروه، هم‌اکنون در چه گروه‌هایی عضو کنونی و مدیر هستند، مهاجرت کاربران سنجیده می‌شود. کریم‌پور و همکاران (۲۰۲۱) در پژوهشی دیگر، برای پیشنهاد کاربران تلگرامی روشی ترکیبی ارائه داده‌اند. روش آنان دارای دو مرحله است. مرحله نخست برای هر گروه، کیسه‌ای از کلمات می‌سازد. مرحله دوم به صورت برخط است که کاربران ورودی دریافت کرده و سپس فهرستی از گروه‌ها را بر اساس روش هاشمی و زارع چاهوکی (۱۳۹۷) تولید می‌کنند. در نهایت، به ترتیب فهرست، کیسه کلمات حاصل از مرحله نخست با یکدیگر ترکیب می‌شوند تا به این پرسش پاسخ داده شود که ترکیب کیسه کلمات چه تعداد گروه در دستیابی به کاربران هدف مفید است. آنان پس از ترکیب کیسه کلمات گروه‌ها و بررسی میزان خطا در هر ترکیب به این نتیجه دست یافتند که ترکیب کیسه کلمات چهار گروه ابتدایی در فهرست، پیش‌بینی بیشتری از مجموعه کاربران هدف دارد.

پیش‌بینی لینک

در برخی پژوهش‌ها، پیشنهاد کاربر فقط از طریق گراف انجام می‌شود که این مبحث با موضوعی با عنوان پیش‌بینی لینک^۳ مطرح می‌شود. پیش‌بینی لینک مبحثی برای بررسی گروه‌ها (کاربران) در گراف شبکه اجتماعی است که ارتباط آینده بین کاربران را پیش‌بینی می‌کند. تاکنون در زمینه پیش‌بینی لینک تلاش‌های زیادی شده است. الگوریتم‌های زیادی از جمله شاخص آدامیک آدار^۴، شاخص کاتز^۵ و شاخص جاکارد^۶ در این زمینه هستند که به عنوان روش‌هایی پایه شناخته می‌شوند (برای بررسی و مقایسه این روش‌ها به مرجع لیو و ژاو^۷ (۲۰۱۱) مراجعه شود). در زمینه پیش‌بینی لینک و حل مسائل مربوط به گراف در خصوص یافتن لینک جدید بین کاربران، پژوهش‌هایی انجام شده است. جیبونی و همکاران^۸ (۲۰۱۸) روشی مبتنی بر شباهت برای پیشنهاد کاربر در فیس‌بوک^۹ ارائه داده‌اند و ارتباط آینده بین کاربران را بر اساس عمق مسیر و درجه گره بررسی کرده‌اند. در این روش، بر اساس ویژگی‌های توپولوژیکی به منظور بهبود عملکرد روش پیش‌بینی لینک، معیاری جدید ارائه شده است.

برخی پژوهش‌ها، همانند پژوهش پیش رو بازاریابی را مد نظر قرار داده‌اند. لی و همکاران^{۱۰} (۲۰۲۰) مدلی برای شناسایی حلقه‌ای از دوستان کاربران در شبکه اجتماعی توئیتر و گوگل پلاس^{۱۱} ارائه داده‌اند.

1. Rich Get Richer
2. Karimpour, Zare Chahooki & Hashemi
3. Link Prediction
4. Adamic/Adar Index
5. Katz Index
6. Jaccard Index
7. Lu & Zhou
8. Jibouni, Lotfi, El Marraki & Hammouch
9. Facebook
10. Li et al.
11. Google-plus

آنان برای این کار، به منظور شناسایی دوستی احتمالی بین کاربران در حلقه‌های مختلف و شاخص تخصیص توجه^۱، از پیش‌بینی لینک استفاده کرده‌اند. پژوهشگران علاوه بر همسایگان مستقیم هر کاربر، از همسایگان غیرمستقیم کاربران نیز بهره برده‌اند. بازاریابی در این پژوهش با پیشنهاد حلقه‌ای از دوستان به هر کاربر انجام می‌شود. اکثر پژوهش‌های گذشته روی شباهت بین دو کاربر تمرکز دارند، در حالی که پژوهش پیش رو شباهت بین مجموعه‌ای از کاربران را به دست می‌آورد. برخی پژوهش‌ها روی شباهت مجموعه‌ای از کاربران تمرکز دارند، اما جامعه‌ای از کاربران را به هر کاربر از مجموعه کاربران، پیشنهاد می‌دهند.

توصیه جوامع

برخی پژوهش‌ها روی شباهت بین جوامعی از کاربران تمرکز دارند. این پژوهش‌ها بیشتر با عنوان توصیه جوامع بیان می‌شوند و به جوامعی از کاربران پیشنهاد می‌دهند. هان و همکاران^۲ (۲۰۱۶) برای بررسی شباهت علاقه بین چند کاربر در شبکه اجتماعی فیس‌بوک، روشی ارائه داده‌اند. آنان در این روش، بیش از یک میلیون جامعه مختلف را در کمتر از یک ساعت شناسایی و شباهت بین آنها را محاسبه کردند. این پژوهش علاوه بر سنجش شباهت، روشی برای انتخاب جوامع نیز ارائه داده است. السویدان و یخلف^۳ (۲۰۱۸) به منظور خوشه‌بندی کاربران مبتنی بر علاقه، برای تشخیص جوامع روشی ارائه کردند. در این روش، کاربران در گراف بر اساس دو پارامتر دوستان و علاقه آنها در نظر گرفته شده‌اند.

روش پیشنهادی

این پژوهش از نظر هدف کاربردی است. هدف اصلی در این پژوهش، ارائه روشی است که از طریق آن بتوان کاربرانی مشابه با کاربران هدف را در پیام‌رسان تلگرام پیشنهاد داد. جامعه هدف، مجموعه کاربرانی هستند که توسط سامانه پیشنهادی ارائه می‌شوند. برای این منظور، اطلاعات گروه‌های عمومی و کاربران پیام‌رسان تلگرام به وسیله سامانه ایده‌کاو به دست آمده است. همچنین از سطح گروه‌های تلگرام، ویژگی‌هایی از جمله تعداد اعضا و تعداد مدیر در هر گروه استخراج شده است. به‌طور کلی روش پیشنهادی شامل پنج پیشنهاد و در دو بخش کمی و کیفی به‌منظور توسعه روش هاشمی و زارع چاهوکی (۱۳۹۷)، ایجاد شده است. برای سهولت در بیان مطالب، روش هاشمی و زارع چاهوکی (۱۳۹۷) با نام روش گروه‌های مشابه ادامه می‌یابد. هر یک از پنج پیشنهاد، پیشنهاد قبل از خود را تکمیل می‌کنند و هر یک را می‌توان روشی جداگانه برای پیشنهاد کاربر در نظر گرفت. برای بیان بهتر روش پیشنهادی، این بخش به سه زیربخش تقسیم شده است. در زیربخش نخست، مسئله مد نظر بیان و بررسی می‌شود. در

1. Attention Allocation Index

2. Han et al.

3. AlSuwaidan & Ykhlef

زیربخش دوم، سناریوی کلی نوآوری بررسی شده و در زیربخش سوم، هر یک از روش‌های پیشنهاد کاربر تشریح می‌شوند.

تعریف مسئله

تلگرام در ایران به‌عنوان یک شبکه اجتماعی استفاده می‌شود، اما تمامی خدمات یک شبکه اجتماعی از جمله پیشنهاد کاربر و بازاریابی مستقیم را ارائه نمی‌دهد. با توجه به اینکه تاکنون افزودن قابلیت توصیه‌گری در تلگرام جزء اهداف و استراتژی مالکان این پیام‌رسان قرار نگرفته است و ضمن اینکه نبود این قابلیت در این پیام‌رسان بسیار مشاهده می‌شود، این مشکل وجود دارد که چگونه کاربران تلگرام را برای گسترش بازاریابی، جست‌وجو کرده و پیشنهاد دهیم.

برای بیان بهتر مسئله، برای مثال فرد بازاریابی را در نظر می‌گیریم که در زمینه زعفران فعالیت دارد. این بازاریاب، به دنبال مجموعه‌ای کاربر است که خریدار زعفران در قالب خواروبارفروشی‌ها، عطاری‌ها و رستوران‌داران هستند، زیرا بازاریاب می‌خواهد محصولات خود را در بین این جامعه هدف، تبلیغ کرده و ارائه دهد. این بازاریاب برای تبلیغ محصولات خود، فقط از طریق یافتن گروه‌هایی در حوزه زعفران، می‌تواند با پرداخت هزینه به مدیران آن گروه‌ها، پیام تبلیغاتی خود را برای مدت زمانی محدود در گروه آنان قرار دهد. البته ارسال پیام تبلیغاتی در برخی از گروه‌ها رایگان است. همچنین این نوع بازاریابی برای کانال نیز صدق می‌کند، با این اختلاف که مدیر کانال پس از دریافت هزینه از بازاریاب، پیام تبلیغ را برای مدتی در کانال قرار می‌دهد. در برخی از گروه‌ها، بازاریابان علاوه بر اینکه در هر گروه پیام خود را ارسال می‌کنند، تبلیغ محصول خود را به هر کاربر عضو در آن گروه به‌صورت جداگانه نیز ارسال می‌کنند. این گونه از تبلیغات برای بازاریاب مبهم است، زیرا بدون هیچ‌گونه پیش‌زمینه اولیه‌ای، به ارسال تبلیغ اقدام کرده است و درباره مؤثر بودن تبلیغات خود هیچ‌گونه اطمینانی ندارد. در تبلیغات غیرهدفمند ضمن اینکه بازاریاب هزینه زیادی برای ارسال تبلیغ به مدیران گروه‌ها پرداخت کرده است، به‌طور دقیق نمی‌تواند برای تبلیغ محصولات خود جامعه هدف مناسبی بیابد. در این نوع از تبلیغات، احتمال اینکه بازاریاب توسط کاربران هدف به‌عنوان مزاحم و پیام او به‌عنوان هرزنامه و اسپم گزارش شود، بسیار زیاد است. همان‌طور که بازاریاب به‌دنبال جامعه هدفی برای تبلیغ و فروش محصولات خود است، کاربران خریدار نیز به‌دنبال بازاریابانی برای خرید محصولات حوزه کاری خود هستند. از این رو، این دو نوع کاربر ممکن است هیچ‌وقت یکدیگر را پیدا نکنند، زیرا تلگرام هیچ‌گونه بازاریابی مستقیمی را در اختیار بازاریابان قرار نداده است. بنابراین، مشکل اصلی در تلگرام این است که میلیون‌ها کاربر در این پیام‌رسان فعالیت دارند، اما فرد بازاریاب نمی‌داند کدام کاربران جزء جامعه هدف بازاریابی او هستند.

در این پژوهش، مشکل بیان‌شده را تا حد امکان رفع کرده‌ایم، به‌طوری که برای مثال در بازاریابی زعفران، فرد بازاریابی که مشتریان خوبی دارد و به‌دنبال یافتن مشتریان جدیدی مشابه آنان است، می‌تواند با دادن شماره تلفن یک یا چند کاربر مد نظر خود (مثل رستوران‌داران و عطاری‌ها) به سامانه پیشنهادی، مجموعه‌ای از افراد مشابه را دریافت کند. در صورتی که بازاریاب، مشتریانی را به‌عنوان شروع

برای این کار ندارد، می‌تواند در ابتدا مدیران کانال‌ها و گروه‌های زعفران را به‌عنوان یک دانه^۱ اولیه انتخاب کرده، سپس جامعه هدف خود را به‌وسیله سامانه ارائه‌شده در این پژوهش، دقیق‌تر کند.

فرض و هدف مسئله

فرض مسئله در این پژوهش این است که در ابتدا تعدادی کاربر به‌عنوان کاربر هدف و همچنین گراف ارتباطی کاربران به‌عنوان ورودی به روش پیشنهادی داده می‌شود. گراف ارتباطی، کاربران را به‌وسیله عضویت خود در گروه‌های تلگرام به یکدیگر متصل می‌کند. هدف مسئله، یافتن کاربرانی مشابه در تلگرام است که ویژگی‌های آنان با کاربران هدف، شباهت داشته باشد. در واقع، می‌خواهیم به این پرسش پاسخ دهیم که چگونه و تا چه اندازه می‌توان پیشنهاد فهرستی از کاربران مشابه با کاربران ورودی را به‌وسیله گراف ارتباطی در پیام‌رسان تلگرام بهبود بخشید؟

روش رسیدن به جواب مسئله

با توجه به مسئله مطرح‌شده در زیربخش قبل، روش رسیدن به جواب مسئله در بخش پیشینه پژوهش با این پرسش که در پژوهش‌های دیگر چگونه با این مشکل برخورد شده، بررسی شده است. از آنجا که در تلگرام تاکنون فقط یک پژوهش با هدف مطرح‌شده در این مقاله (پیشنهاد کاربر) انجام شده است، بررسی کردیم که در شبکه‌های اجتماعی دیگر مانند توییتر، فیس‌بوک و گوگل‌پلاس، با این مسئله چگونه برخورد شده است. در توییتر و فیس‌بوک مسئله با سبک دیگری با عنوان توصیه دوست و توصیه جوامع بیان شده که در مرور پژوهش‌ها در بخش پیشینه پژوهش به آن پرداخته شده است. اما مجدد نیز تأکید می‌شود که راه‌حل آنان در تلگرام به‌صورت دقیق مصداق ندارد. زیرا در شبکه‌های اجتماعی مختلف، کاربران در مقایسه با تلگرام، پروفایل کامل‌تری دارند. به‌طور مثال، در شبکه اجتماعی توییتر، کاربران دارای مشخصاتی از جمله رنگ و غذای مورد علاقه، موضوعات مطالعه‌ای، دوستان مورد علاقه و بسیاری اطلاعات دیگر هستند. اما در تلگرام تنها نقطه عطف کاربران، گروه‌ها هستند. از این رو، در این مقاله روش پیشنهادی جدیدی برای پیشنهاد کاربران با بهره‌گیری از گراف ارتباطی کاربران و در نظر گرفتن وضعیت‌های مختلف برای هر کاربر در گروه‌ها ارائه شده است. این پژوهش با تعریف ویژگی‌هایی عددی در گروه‌ها روشی طی پنج پیشنهاد ارائه می‌دهد.

هر یک از سامانه‌های پیشنهاددهنده کاربر در شبکه‌های اجتماعی با در نظر گرفتن میزان اطلاعاتی که از کاربران در آن شبکه استخراج می‌شود، به‌طور خاص کاربرانی مشابه با کاربران هدف را بر اساس قابلیت‌های آن شبکه اجتماعی، پیشنهاد می‌دهد. به همین دلیل در این پژوهش روشی ارائه دادیم که علاوه بر اجرا شدن روی داده‌های تلگرام می‌توان آن را روی سایر شبکه‌های اجتماعی و پیام‌رسان‌ها که ویژگی ایجاد گروه را دارا هستند، بررسی و اجرا کرد. برای پیشنهاد کاربران، روش‌های بسیاری وجود دارند که در قالب سامانه‌های پیشنهاددهنده مطرح می‌شوند. همان‌طور که در پیشینه پژوهش مطرح شد،

سامانه‌های پیشنهاددهنده با توجه به میزان اطلاعاتی که از کاربران در اختیار است، به دو دسته توپولوژیکی و غیرتوپولوژیکی تقسیم می‌شوند. روش این پژوهش با توجه به میزان اطلاعات به‌دست‌آمده از کاربران تلگرام، جزء روش‌های توپولوژیکی است. روش‌های توپولوژیکی برای ارائه پیشنهادها، فقط از گراف ارتباطی کاربران استفاده می‌کنند.

سناریوی کلی از کارکرد روش پیشنهادی

به‌طور کلی، روش پیشنهادی طی پنج پیشنهاد تکامل یافته است. سناریوی کلی روش پیشنهادی به این شرح است که ورودی هر پیشنهاد مجموعه‌ای کاربر بوده و خروجی هر یک، فهرستی از گروه‌های رتبه‌بندی شده است. در فهرست گروه به‌دست‌آمده از هر پیشنهاد، کاربران گروه‌ها به‌ترتیب فهرست تا رسیدن به مجموعه هدف با یکدیگر ادغام می‌شوند. مجموعه هدف، تعیین حداکثر تعداد کاربر به‌دست‌آمده نهایی، توسط هر یک از پنج پیشنهاد است. به‌طور مثال، هدف را می‌توان دستیابی به مجموعه‌ای ۵، ۱۰ یا ۲۰ برابر مجموعه کاربر ورودی در نظر گرفت. هر یک از پیشنهادها سعی در رسیدن به مجموعه‌ای کمتر مساوی مجموعه هدف دارند. در صورتی که تعداد کاربران نخستین گروه در فهرست به‌دست‌آمده توسط هر پیشنهاد، بیشتر از تعداد کاربر هدف باشد، عملیات ادغام کاربران گروه‌ها در همان مرحله متوقف می‌شود.

پیشنهاد‌های بخش کمی و کیفی

به‌طور کلی، پیشنهاد‌های ارائه‌شده به دو بخش کمی و کیفی تقسیم شده‌اند. بخش کمی به بیان رتبه‌بندی گروه‌ها بر اساس گراف ارتباطی کاربران می‌پردازد و بخش کیفی علاوه بر گراف، گروه‌ها را بر اساس کیفیت تعداد مدیر کم در هر گروه، رتبه‌بندی می‌کند. برای بیان بهتر پیشنهادها، ابتدا ویژگی‌هایی عددی تعریف شده‌اند، سپس هر یک از دو بخش کمی و کیفی بیان می‌شوند. در هر یک از بخش‌های کمی و کیفی، نتایج هر پیشنهاد به‌صورت متغیرهایی بیان می‌شوند که در پیشنهادها دیگر استفاده می‌شوند.

U: کاربران ورودی

S_A : تعداد کاربران با وضعیت مدیر در هر گروه

S: تعداد کاربران یا تعداد اعضا در هر گروه

C: مقدار عددی (رتبه) هر گروه در فهرست گروه‌های حاصل از روش گروه‌های مشابه

بخش کمی

بخش کمی به بیان پیشنهاد‌هایی مبتنی بر گراف، متناسب با گروه‌های کاربران مجموعه U می‌پردازد. در ابتدای شروع این بخش، روش گروه‌های مشابه اجرا می‌شود و فهرستی از گروه‌های رتبه‌بندی‌شده به دست می‌آید. سپس، این فهرست به پیشنهاد نخست داده می‌شود. در این بخش، ویژگی‌های عددی

استخراج شده از گروه‌ها با یکدیگر ترکیب می‌شوند و تا رسیدن به مرحله نرمال^۱ پیش می‌روند. مرحله نرمال مرحله‌ای است که فقط گروه‌های با تعداد عضو کم یا با تعداد عضو زیاد در ابتدای فهرست قرار ندارند، بلکه رتبه‌بندی گروه‌ها به صورت ترکیبی از گروه‌های با تعداد عضو کم و زیاد انجام می‌شود. در این بخش هیچ گونه تضمین کیفیتی وجود ندارد و فقط کاربران ورودی بر اساس ویژگی‌های عددی استخراج شده از گروه‌ها، رتبه‌بندی می‌شوند. منظور از کیفیت، رتبه‌بندی گروه‌ها بر اساس برخی ویژگی‌های عددی استخراج شده از گروه‌ها است که با در نظر گرفتن فرضیه‌هایی، کیفیت گروه را تضمین می‌کنند.

• پیشنهاد نخست

ورودی این پیشنهاد، فهرست گروه‌های حاصل از روش گروه‌های مشابه است. این پیشنهاد در رابطه ۱ نشان داده شده که C مقدار عددی (رتبه) هر گروه در فهرست گروه‌های حاصل از روش گروه‌های مشابه است. به بیان دیگر، C تعداد کاربر مشترک هر گروه با مجموعه U است. در این پیشنهاد، برای رفع مشکل روش گروه‌های مشابه، C بر S تقسیم می‌شود تا به جای تعداد کاربران، درصد عضویت کاربران از هر گروه در مجموعه U در نظر گرفته شود. به طور مثال، گروهی با $S=20$ و $C=10$ نسبت به گروهی با $S=200$ و $C=80$ اولویت بیشتری دارد، زیرا با در نظر گرفتن درصد عضویت، گروه اول با مقدار $\frac{C}{S}=0.5$ نسبت به گروه دوم با مقدار $\frac{C}{S}=0.4$ اولویت بیشتری دارد.

این پیشنهاد در حالی که مشکل روش گروه‌های مشابه را رفع کرده، بیانگر مشکل دیگری است. ایراد این پیشنهاد برخلاف روش گروه‌های مشابه، این است که گروه‌های با S کم را به ابتدای فهرست می‌آورد. دلیل ایجاد این مشکل این است که در رابطه ۱، اگر S از حدی بیشتر شود، مقدار C دیگر نمی‌تواند به همان نسبت افزایش یابد که موجب می‌شود تا گروه‌هایی که S کمتری دارند، به ابتدای فهرست بیایند.

$$R_1 = \frac{C}{S} \quad \text{رابطه ۱}$$

• پیشنهاد دوم

در این پیشنهاد برای رفع مشکل پیشنهاد نخست، کسر رابطه ۱، نرمال شده است. منظور از نرمال این است که اثر خطی و مستقیم یک کسر خنثی شود تا مقیاس‌های بزرگ آن کاهش یابند. هدف از نرمال کردن کسر $\frac{C}{S}$ ، این است که رتبه‌بندی حداکثری گروه‌های با S کم کاهش یافته و رتبه‌بندی به صورت ترکیبی از گروه‌های با S زیاد و S کم انجام شود. به منظور نرمال کردن کسر $\frac{C}{S}$ و اینکه با بیشتر شدن S مقدار C افت پیدا نکند، این کسر در لگاریتم S ضرب شده و در رابطه ۲ نشان داده شده است. با انجام این ضرب، اثر خطی و افزایش زیاد S در مخرج کسر $\frac{C}{S}$ ، خنثی شده است. برای ضرب کسر $\frac{C}{S}$ در لگاریتم S ، هر پایه‌ای از لگاریتم در نظر گرفته شود تفاوتی ندارد، زیرا فقط حدود اعداد تغییر می‌کند.

در این پیشنهاد و پیشنهاد نخست، هیچ‌گونه کیفیتی برای رتبه‌بندی گروه‌ها در نظر گرفته نشده است. مشکل این دو پیشنهاد این است که امکان دارد، گروه‌هایی بی‌کیفیت در اولویت فهرست گروه‌ها قرار گرفته باشند که در نتیجه، پیشنهاد کاربران از گروه‌های بی‌کیفیت انجام می‌شود. منظور از گروه‌های بی‌کیفیت، گروه‌هایی هستند که کاربران نامرتبط و مدیران زیادی دارند و بسیاری از پیام‌ها در آنان، هرزنامه است.

$$R_2 = \frac{C}{S} * \log_{10} S \quad \text{رابطه ۲}$$

بخش کیفی

در این بخش، پیشنهادهایی مبتنی بر کیفیت‌سنجی گروه‌ها به پیشنهادهای بخش کمی افزوده می‌شوند.

• پیشنهاد سوم

به‌منظور سنجش کیفیت گروه‌های تلگرام روش‌های بسیاری در پژوهشی که هاشمی و زارع چاهوکی (۲۰۱۹) ارائه داده‌اند، بررسی شده است. در این پژوهش، به‌منظور بررسی کیفیت گروه‌ها، فقط با در نظر گرفتن این فرض که هرچقدر S_A در هر گروه کم باشد، آن گروه کیفیت بیشتری دارد، کیفیت را در گروه‌ها بررسی کرده‌ایم. گروه‌ها دارای مدیر یا مدیرانی هستند که به گروه نظم می‌دهند. آنان می‌توانند فعالیت کاربران مخرب را در گروه، محدود کرده یا کاربری را از گروه اخراج کنند. هنگام افزایش S_A ، کیفیت گروه به این دلیل کاهش می‌یابد که در داخل گروه، هرج‌ومرج می‌شود و با وجود اینکه مدیریتی چندگانه به وجود می‌آید، امکان اینکه در این گروه‌ها، کارهای غیرمعارفی انجام شود بسیار زیاد است. به‌طور مثال، کیفیت گروهی با $S_A=1$ بسیار بهتر از گروهی با $S_A=5$ است. طبق این بررسی می‌توان نتیجه گرفت که در هر گروه، با افزایش S هرچقدر مقدار S_A کاهش یابد آن گروه دارای کیفیت بالایی است. به‌طور کلی، برای هر گروه هرچقدر مقدار حاصل از نسبت $\frac{S_A}{S}$ کمتر باشد، آن گروه کیفیت بهتری دارد. به‌دلیل اینکه هرچقدر نتیجه حاصل از پیشنهاد دوم بیشتر باشد، بهتر است، برای افزودن نسبت $\frac{S_A}{S}$ به پیشنهاد دوم، این نسبت معکوس شده و در رابطه ۳، نشان داده شده است.

مشکل این پیشنهاد همانند پیشنهاد نخست بیانگر این است که در رابطه ۳، با افزایش S مقدار S_A نمی‌تواند افزایش یابد. پس از بررسی مثال‌هایی عددی به این نتیجه دست یافتیم که این پیشنهاد در برخی موارد از نظر منطقی نتیجه مد نظر را ارائه نمی‌دهد. برای مثال، با در نظر گرفتن مقادیر زیر، در حالی که گروه دوم نسبت به گروه نخست کیفیت بالاتری دارد، هر دو گروه در یک رتبه قرار گرفته‌اند.

$$\text{گروه نخست: } S=200/000 \text{ و } S_A=200$$

$$\text{گروه دوم: } S=200/000 \text{ و } S_A=2$$

$$\frac{1}{\frac{S_A}{S}} [\text{گروه نخست}] = \frac{1}{\frac{S_A}{S}} [\text{گروه دوم}] = 10/000$$

$$R_2 = \frac{C}{S} * \log_{10} S \quad \text{رابطه ۳}$$

• پیشنهاد چهارم

در این پیشنهاد برای رفع مشکل پیشنهاد سوم، کسر $\frac{1}{S_A}$ در رابطه ۳، نرمال شده است. منظور از نرمال این است که اثر خطی و مستقیم کسر $\frac{1}{S_A}$ خنثی شود. هدف از نرمال کردن نسبت $\frac{1}{S_A}$ ، این است که اثرگذاری زیاد S کاهش یابد. برای نرمال کردن کسر $\frac{1}{S_A}$ ، این نسبت در $\frac{1}{\sqrt{S}}$ ضرب شده و در رابطه ۴ نشان داده شده است. در این پیشنهاد، همانند پیشنهاد دوم، تفاوتی ندارد که برای ضرب کردن $\frac{1}{S_A}$ در $\frac{1}{\sqrt{S}}$ ، چه مبنایی از جذر در نظر گرفته می‌شود، زیرا فقط حدود اعداد تغییر می‌کند. دلیل استفاده از جذر به جای لگاریتم در این پیشنهاد این است که با بررسی مثال‌هایی عددی به این نتیجه دست یافتیم که جذر S در مقایسه با لگاریتم S نتیجه درست‌تری دارد. به‌طور مثال، با در نظر گرفتن مقادیر زیر برای دو گروه، نتیجه استفاده از جذر نسبت به لگاریتم در رتبه‌بندی گروه با S_A کمتر، برتری دارد.

گروه نخست: $S=10/000$ و $S_A=1$

گروه دوم: $S=100/000$ و $S_A=8$

$$\frac{1}{S_A * \log_{10} S} [\text{گروه نخست}] = \frac{1}{S_A * \log_{10} S} [\text{گروه دوم}] = 2500$$

$$\frac{1}{S_A * \sqrt{S}} [\text{گروه نخست}] = 100 \quad \text{و} \quad \frac{1}{S_A * \sqrt{S}} [\text{گروه دوم}] = 39/5$$

بر اساس این مثال، در صورت استفاده از لگاریتم، نتیجه برای هر دو گروه، برابر ۲۵۰۰ می‌شود و هر دو در فهرست گروه‌ها، رتبه‌های برابری خواهند داشت. اما در صورت استفاده از جذر، گروه نخست برابر با ۱۰۰ و گروه دوم برابر ۳۹/۵ می‌شود که گروه نخست در مقایسه با گروه دوم، در رتبه‌بندی اولویت خواهد داشت.

$$R_4 = \frac{C}{S} * \log_{10} S * \frac{1}{S_A * \sqrt{S}} \quad (\text{رابطه ۴})$$

• پیشنهاد پنجم

این پیشنهاد به‌عنوان روشی احتمالی روی نتایج حاصل از پیشنهاد چهارم اعمال شده است. در این پیشنهاد، کاربران U از جهتی دیگر و بر اساس مدیر بودن بررسی می‌شوند. دلیل احتمالی بودن این روش، این است که به‌صورت قطعی در میان کاربران ورودی، کاربری از نوع مدیر وجود ندارد. ابتدا در این پیشنهاد، گروه‌هایی که کاربران U در آنها مدیر هستند، استخراج می‌شوند. مجموعه گروه‌های به‌دست‌آمده P نامیده شده‌اند. سپس گروه‌های مجموعه P در ابتدای فهرست نتایج به‌دست‌آمده از پیشنهاد چهارم قرار می‌گیرند. پس از بررسی نتایج بیان‌شده در بخش نتایج تجربی، مشخص شد که اکثر گروه‌های مجموعه P در فهرست گروه‌های به‌دست‌آمده از پیشنهاد چهارم وجود دارند، از این رو این گروه‌ها با همان رتبه به ابتدای فهرست آمده‌اند. گروه‌هایی از مجموعه P که در فهرست حاصل از پیشنهاد

چهارم وجود ندارند، در انتهای گروه‌های موجود قرار می‌گیرند. دلیل بررسی این پیشنهاد این است که احتمال دارد کاربران مدیر، متناسب با گروه‌های خود در گروه‌های دیگر عضو باشند.

توصیف داده‌های جمع‌آوری‌شده

داده‌های این پژوهش، اطلاعات کاربران و گروه‌های پیام‌رسان تلگرام است که توسط سامانه ایده‌کاو خزش شده‌اند. سامانه ایده‌کاو به‌عنوان ابزاری برای جست‌وجو و بازاریابی در بستر تلگرام به وجود آمده است. این سامانه یک سرویس شخص ثالث است که به‌طور مداوم بر گروه‌ها و کانال‌های تلگرام نظارت دارد تا تمام اطلاعات ممکن استخراج شوند. ایده‌کاو خوشه‌ای از خزنده‌ها به‌منظور جمع‌آوری و بروز اطلاعات تلگرام دارد. هر یک از خزنده‌ها به‌طور مستقل عمل می‌کنند و همه به یک پایگاه داده مرکزی متصل هستند (Hashemi & Chahooki, 2019).

نحوه ایجاد گراف به این صورت است که ابتدا اعضای گروه‌ها واکشی می‌شوند، سپس کاربرانی که در یک گروه هستند، به یکدیگر متصل می‌شوند. شایان ذکر است که گراف، هم‌زمان با خزش اطلاعات توسط خزش‌گر ایده‌کاو، ایجاد و به‌روزرسانی می‌شود. هنگامی که خزش‌گر ایده‌کاو، اطلاعات را از سطح گروه‌ها به دست می‌آورد، وضعیت هر کاربر را در هر گروه به‌روزرسانی می‌کند. داده‌های این پژوهش به‌طور دقیق شامل ۶۹/۹۷۴/۱۷۱ کاربر و ۶۸۲/۲۱۸ ابرگروه^۱ است که در بازه زمانی دی ۱۳۹۸ تا فروردین ۱۳۹۹ توسط خزش‌گر ایده‌کاو به دست آمده‌اند. تمامی گروه‌های استفاده‌شده در این پژوهش از نوع ابرگروه هستند، به‌طوری که هر یک حداقل دارای ۲۰۰ عضو هستند.

نتایج تجربی

در این بخش، ابتدا داده‌های ارزیابی‌شده و روش ارزیابی بیان می‌شود، سپس آزمایش‌های فرضیه‌ای انجام شده و نتایج، تجزیه و تحلیل شده‌اند.

در این مقاله، روش پیشنهادی به‌وسیله گروه‌های محبوب در تلگرام ارزیابی شده است. گروه‌های محبوب، گروه‌هایی هستند که اعضا در آنان پیام‌هایی مرتبط با موضوع همان گروه ارسال می‌کنند و پیام‌های تبلیغاتی در آنان ارسال نمی‌شود. در این مقاله، گروه‌های محبوب را توسط دو فرد خبره شناسایی کرده‌ایم. دو فرد مد نظر با عضویت در گروه‌های تلگرام، فقط گروه‌هایی را استخراج کرده‌اند که پیام‌های آنان در خصوص موضوع گروه باشد و در آن گروه هیچ‌گونه پیام تبلیغاتی نامرتبلی ارسال نشده باشد. در این مقاله، دو دسته از گروه‌ها انتخاب شده‌اند و هر دسته شامل ۳۰ گروه محبوب در تلگرام است. گروه‌های محبوب انتخاب‌شده در هر دو دسته، تعداد اعضای در بازه ۲/۰۰۰ تا ۱۰/۰۰۰ دارند. برای انتخاب دو دسته، سعی بر این بوده است تا گروه‌ها بر اساس تعداد اعضا به‌طور مساوی بین دو دسته تقسیم شوند. اطلاعات گروه‌های محبوب انتخاب‌شده، برای دسته اول و دسته دوم در جدول ۱ نشان داده شده است. دلیل انتخاب دو دسته از گروه‌های محبوب این است که نشان داده شود روش پیشنهادی روی

مجموعه‌ای خاص از گروه‌های اعتبارسنجی شده ارزیابی نشده است و نتایج تکرارشدنی هستند. دسته نخست، فقط برای تنظیم پارامترها و اعتبارسنجی هر یک از پیشنهادها روش پیشنهادی ارائه شده است. دسته دوم نیز فقط هر یک از پیشنهادها روش پیشنهادی را آزمون می‌کند. گفتنی است که این دسته در ایجاد و تعیین پارامترها هیچ گونه نقشی نداشته است. با استفاده از مقایسه نتایج حاصل از ارزیابی هر یک از دو دسته، روایی و پایایی نتایج به دست آمده را بررسی کرده‌ایم. اعتبارسنجی در این پژوهش به منظور تنظیم پارامترها و به دست آوردن کمترین خطا توسط هر یک از گروه‌های محبوب در دسته نخست به صورت جداگانه با تقسیم بندی اعضای هر گروه به دو دسته ۸۰ و ۲۰ درصد انجام شده است. همچنین، در انتها پس از انجام اعتبارسنجی، گروه‌های محبوب در دسته آزمون به صورت جداگانه توسط همان تقسیم بندی ۸۰ و ۲۰ درصد ارزیابی شده‌اند. شایان ذکر است که برای ارزیابی هر گروه محبوب، اطلاعات آن گروه (عضویت کاربران) را از گراف ارتباطی کاربران به منظور بررسی هر یک از پیشنهادها و پژوهش‌های پیشین حذف کرده‌ایم.

ارزیابی در دو دسته اعتبارسنجی و آزمون به این صورت است که برای هر گروه محبوب، ۸۰ درصد کاربران آن گروه در اختیار پیشنهادها موجود در روش پیشنهادی قرار داده می‌شود و ۲۰ درصد کاربران برای ارزیابی نهایی، نگهداری می‌شوند. هر یک از پیشنهادها اول تا پنجم در روش پیشنهادی، برای رسیدن حداکثری به مجموعه هدف سعی دارند. مجموعه هدف یا مجموعه کاربران که می‌خواهند در نهایت پیشنهاد داده شوند را می‌توان مقادیر دلخواهی در نظر گرفت. به طور مثال، می‌توان مجموعه هدف را ۱۰، ۲۰ یا ۳۰ برابر مجموعه کاربران ورودی (۸۰ درصد) در نظر گرفت. برای انجام آزمایش‌ها روی پیشنهادها مطرح شده در این مقاله و سه پژوهش پیشین، هدف تمامی روش‌ها، دستیابی به مجموعه‌ای کاربر، ۱۰ برابر مجموعه کاربران ورودی در نظر گرفته شده است.

برای سهولت در مقایسه و بررسی پیشنهادها ارائه شده در این مقاله با سه پژوهش پیشین، هر یک از سه پژوهش را با نامی مشخص نام گذاری کرده‌ایم. روش هاشمی و زارع چاهوکی (۱۳۹۷) با نام گروه‌های مشابه، روش کریم پور و همکاران (۲۰۲۱) با نام مهاجرت و دومین روش کریم پور و همکاران (۲۰۲۱) با نام ترکیبی مشخص شده‌اند. با توجه به اینکه روش گروه‌های مشابه و روش مهاجرت، برای پیشنهاد گروه هستند، به منظور مقایسه، کارکرد این دو روش را به عنوان پیشنهاد کاربر در نظر گرفته‌ایم. زیرا هر یک از این دو روش، همانند پیشنهادها اول تا پنجم ارائه شده، فهرستی از گروه‌ها ارائه می‌دهند. در نهایت، پس از دستیابی به فهرست گروه‌ها، اعضای گروه‌ها به ترتیب فهرست تا رسیدن به مجموعه هدف با یکدیگر ادغام می‌شوند. روش ترکیبی، روشی برای پیشنهاد کاربر به وسیله ترکیب کلمات کلیدی گروه‌ها است. این روش، کلمات کلیدی گروه‌های حاصل از روش گروه‌های مشابه را به ترتیب فهرست با یکدیگر ترکیب می‌کند. پژوهشگران در این روش به این نتیجه دست یافتند که ترکیب کلمات کلیدی چهار گروه، پیش‌بینی بهتری از مجموعه کاربران هدف دارد. از این رو، برای مقایسه این روش با روش پیشنهادی، هر یک از ۳۰ گروه موجود در دسته آزمون را بر اساس ترکیب کیسه کلمات چهار گروه در نظر گرفته‌ایم.

جدول ۱. اطلاعات نام و تعداد اعضای گروه‌ها در دو دسته اعتبارسنجی و آزمون

دسته دوم (آزمون)			دسته نخست (اعتبارسنجی)		
تعداد اعضا	نام گروه	شناسه گروه	تعداد اعضا	نام گروه	شناسه گروه
۲۰۲۴	گروه تخصصی عیب‌یابی لیزر	گروه ۱	۲۱۴۸	املاک بندرهای شهرک غرب مشهد	گروه ۱
۲۰۶۹	گروه تخصصی نرم‌افزار	گروه ۲	۲۴۰۰	عیب‌یابی تخصصی الکترونیک	گروه ۲
۳۱۱۶	گروه تخصصی نرم‌افزار FLAC	گروه ۳	۲۸۲۴	جامعه مهندسان ژئوتکنیک	گروه ۳
۲۹۵۱	انجمن متخصصان طب ایرانی	گروه ۴	۲۸۲۸	IELTS Speaking Group	گروه ۴
۳۰۴۰	پسته، انار و انجیر نیریز	گروه ۵	۳۲۶۲	Raspberry-Python-Linux	گروه ۵
۳۲۵۵	گروه طراحی داروسازان	گروه ۶	۳۳۶۰	گروه تحلیل داروهای دندانپزشکی	گروه ۶
۳۹۴۵	تجارت پسته استان کرمان	گروه ۷	۳۷۲۵	گروه بزرگ داروخانه نسخ ناخونا	گروه ۷
۴۹۵۰	تبادلات دارو و نسخه خوانی	گروه ۸	۴۳۷۹	تولیدکنندگان تجهیزات پزشکی	گروه ۸
۴۹۹۷	اعلام بار نیسان. سه تن. خاور	گروه ۹	۴۶۲۷	داروهای کیمیا خاص کشوری	گروه ۹
۵۴۸۸	مشاوره وینیو ارشد و دکتری	گروه ۱۰	۵۲۷۵	هم‌بستگی مرغداران ایران	گروه ۱۰
۵۴۹۲	گروه تخصصی برق قدرت	گروه ۱۱	۵۳۱۸	گروه تخصصی کیلومتر	گروه ۱۱
۵۵۷۹	گروه اعلام بار و ترانزیت بندرعباس	گروه ۱۲	۵۳۵۰	مهاجرت کاری به انگلیس	گروه ۱۲
۵۶۷۵	بررسی تجهیزات هایما اس ۷	گروه ۱۳	۵۸۱۱	IELTS Prime™ Group	گروه ۱۳
۵۸۷۰	ارشد و دکتری وزارت بهداشت	گروه ۱۴	۵۹۲۴	گروه تخصصی ماهواره استور	گروه ۱۴
۶۷۱۳	نسخه‌خوانی حرفه‌ای	گروه ۱۵	۶۰۱۴	دنیای بیوشیمی (دکتر خلیلی)	گروه ۱۵
۶۷۴۱	نسخه‌خوانی و تبادل دارو	گروه ۱۶	۶۳۵۱	گروه بررسی تخصصی سلاح بادی	گروه ۱۶
۶۹۳۵	اعلام بار رسمی کل کشور	گروه ۱۷	۶۵۵۲	دنیای جزوات سلولی و مولکولی	گروه ۱۷
۷۰۷۶	تخصصی دستگاه HPLC - GC- AAS	گروه ۱۸	۶۷۹۱	هم‌فکری برق ساختمان	گروه ۱۸
۷۵۴۰	برنامه‌نویسی پایتون Q/A	گروه ۱۹	۷۰۳۱	اعلام بار از مبدأ یزد به تمام کشور	گروه ۱۹
۷۵۷۳	اپای دانشجویی آمریکا ۲۰۲۱	گروه ۲۰	۷۷۲۰	گروه تخصصی تعمیر موتورخانه	گروه ۲۰
۷۶۳۴	پرسش و پاسخ Matlab	گروه ۲۱	۷۸۷۱	گروه تخصصی انواع نژاد فنچ	گروه ۲۱
۸۳۶۲	انگورکاران حرفه‌ای ایران	گروه ۲۲	۸۰۲۳	تحلیل رنو ساندرو	گروه ۲۲
۸۵۹۱	بررسی انواع لوازم‌آرایی	گروه ۲۳	۸۰۳۱	محاسبات عمران سازه	گروه ۲۳
۹۱۹۱	آکادمی طراحان عکاسان	گروه ۲۴	۸۵۶۴	فارسی‌زبانان مقیم آلمان	گروه ۲۴
۹۶۷۹	مقاومت ترمز آلومینیومی و اینورتر	گروه ۲۵	۹۵۱۵	تخصصی اسپلیت و کولرگازی	گروه ۲۵
۹۹۹۹	طب سنتی - پرسش و درمان امراض	گروه ۲۶	۹۷۲۸	گروه خریدوفروش نهادهای کشاورزی	گروه ۲۶
۱۰۰۰۰	تبادل دارو و کمبودهای دارویی	گروه ۲۷	۹۹۱۹	Rotary Engineering Group	گروه ۲۷
۱۰۰۲۴	گروه تخصصی کود کشاورزی	گروه ۲۸	۱۰۱۵۴	تجهیزات تخصصی پاک‌سازی پوست	گروه ۲۸
۱۰۰۶۰	گلخانه داران کیمیا سبز	گروه ۲۹	۱۰۷۵۹	اتحادیه فعالان صنف میوه و تره‌بار	گروه ۲۹
۱۰۴۱۷	آموزش تخصصی شنا	گروه ۳۰	۱۰۸۸۲	DLeX: Python + Linux	گروه ۳۰
۶۴۶۶/۲		میانگین	۶۳۷۱/۲		میانگین

معیار ارزیابی

اکثر پژوهش‌ها و الگوریتم‌های ارائه‌شده در سامانه‌های پیشنهاددهنده، برای بررسی میزان پیش‌بینی مدل از معیار خطا بهره برده‌اند. در این مقاله نیز برای ارزیابی میزان پیش‌بینی روش پیشنهادی از خطای جذر میانگین مربعات^۱ که روشی پرکاربرد برای یافتن خطای میزان پیش‌بینی مدل است، استفاده شده است. این خطا، در رابطه ۵ نشان داده شده است که Predicted_i مجموعه‌ای از صفر و ۱ برای کاربران پیش‌بینی‌شده توسط روش پیشنهادی است. ۱ بیانگر عدم پیش‌بینی درست و صفر عکس آن است. Actual_i، بیان‌کننده وضعیت واقعی کاربران پیشنهادی است. N نیز مجموعه خطای پیشنهادها را بیان می‌کند. برای محاسبه خطا در هر یک از پیشنهادهای این مقاله و پژوهش‌های پیشین، توسط هر یک از گروه‌های محبوب، مجموعه Predicted_i و Actual_i را ایجاد کرده و سپس RMSE را محاسبه کرده‌ایم.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{Predicted}_i - \text{Actual}_i)^2}{N}} \quad \text{رابطه ۵}$$

انجام آزمایش‌ها و تحلیل نتایج

در این زیربخش، پیشنهادهای ارائه‌شده در دو بخش کمی و کیفی روی هر یک از گروه‌های محبوب آزمایش شده‌اند. هر گروه در دو دسته اعتبارسنجی و آزمون، به‌صورت جداگانه ارزیابی شده است. نتایج ارزیابی، برای گروه‌های دسته اعتبارسنجی در جدول ۲ و دسته آزمون در جدول ۳ نشان داده شده است. با توجه به جدول‌های ۲ و ۳ هر یک از پیشنهادهای ارائه‌شده در این مقاله با سه ستون نمایش داده شده‌اند. ستون اول با نام G، نشان می‌دهد که روش مد نظر با چه تعداد گروه توانسته به مجموعه‌ای کمتر مساوی ۱۰ برابر کاربران ورودی (هدف) برسد. ستون دوم با نام T، حداکثر تعداد کاربران به‌دست‌آمده توسط هر روش را نشان می‌دهد که از روی هم ریختن کاربران گروه‌ها با یکدیگر ایجاد می‌شود. ستون سوم با نام E، خطای RMSE را نشان می‌دهد. برای بیان بهتر سه ستون، شناسه گروه ۵ در جدول ۲ را در نظر بگیرید. با توجه به جدول ۱ در دسته اول (اعتبارسنجی)، هدف این گروه، ۱۰ برابر مجموعه کاربر ورودی و رسیدن به ۲۶۱۰۰ کاربر است. بر اساس جدول ۲ این گروه در پیشنهاد نخست، با روی هم ریختن کاربران ۱۲۶ گروه، توانسته است حداکثر به تعداد ۲۵۹۰۴ کاربر با خطای ۰/۷ برسد. به همین ترتیب، هر روش تا رسیدن به مجموعه‌ای کمتر مساوی مجموعه هدف پیش می‌رود.

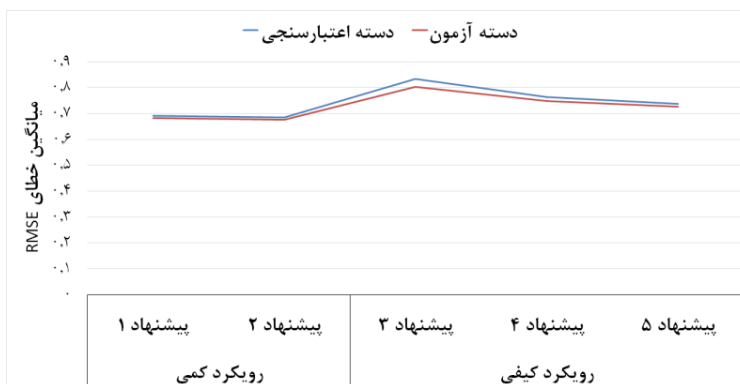
بر اساس نتایج حاصل از جدول‌های ۲ و ۳، پیشنهاد دوم در دو دسته اعتبارسنجی و آزمون در مقایسه با تمامی پیشنهادها، خطای کمتر و پیش‌بینی بیشتری به دست آورده است. این پیشنهاد، ارائه کاربران را از گروه‌های باکیفیت در نظر ندارد. به همین دلیل، پیشنهادی که کیفیت را در نظر داشته باشد و RMSE آن نسبت به سایر پیشنهادها کمتر باشد، پیشنهاد پنجم است. پیشنهاد پنجم نسبت به پیشنهاد دوم RMSE را در دسته اعتبارسنجی به ۰/۵۱ و در دسته آزمون به ۰/۵۲ افزایش داده است.

پیش‌بینی کلی												پیش‌بینی کمی												شماره گروه					
پیش‌بینی ۵						پیش‌بینی ۴						پیش‌بینی ۳						پیش‌بینی ۲						پیش‌بینی ۱					
E	T	G	E	T	G	E	T	G	E	T	G	E	T	G	E	T	G	E	T	G	E	T	G						
۰.۸۴۳	۱۴۹۶	۱۱	۰.۸۴۶	۱۳۱۶	۱۲	۰.۹۴۵	۱۶۰۰۴	۱	۰.۷۸۹	۱۶۵۶۲	۳۲	۰.۷۸۳	۱۱۵۴۱	۶۰	۱														
۰.۵۵۷	۱۸۱۹۳	۱۸	۰.۶۶۵	۱۸۵۶۲	۲۰	۰.۶۹۳	۱۴۵۳۱	۷	۰.۵۱۶	۱۸۵۴۹	۴۷	۰.۵۱۳	۱۸۵۸۳	۹۱	۲														
۰.۷۸۸	۲۳۴۱۴	۲۸	۰.۷۸۵	۱۲۱۴۶	۲۵	۰.۷۳۱	۲۲۰۴۷	۱۶	۰.۷۳۸	۲۲۲۱۷	۴۵	۰.۷۳۷	۲۱۹۳۴	۷۴	۳														
۰.۷۸۵	۱۴۴۱۶	۶	۰.۷۸۵	۱۴۴۱۶	۶	۰.۹۷۴	۱۸۷۸۰	۲	۰.۴۶۸	۲۰۷۸۰	۱۴	۰.۴۵۸	۲۱۵۴۴	۲۰															
۰.۷۸۶	۲۵۴۰۱	۲۱	۰.۸۵۷	۲۰۴۶۵	۱۳	۰.۹۵۴	۱۷۶۶۵	۳	۰.۶۸۱	۲۵۳۹۵	۵۰	۰.۷	۲۵۴۰۴	۱۶۶	۵														
۰.۸۰۲	۲۶۳۷۱	۲۹	۰.۷۷۱	۲۶۳۷۱	۲۳	۰.۸۰۱	۲۳۴۰۲	۹	۰.۷۵۵	۲۶۹۹۱	۴۴	۰.۷۶۶	۲۶۹۵۶	۱۶۱	۶														
۰.۸۵۶	۲۸۱۲۶	۱۶	۰.۷۷۸	۲۸۱۲۷	۱۸	۰.۹۲۴	۱۹۱۵۵	۴	۰.۷۴۵	۲۸۸۰۹	۴۵	۰.۷۶۸	۲۹۶۶۵	۱۰۸	۷														
۰.۶۰۶	۲۲۰۷۰	۲۵	۰.۸۵۵	۲۵۱۵۱	۲۳	۰.۹۴۳	۲۲۴۲	۲	۰.۷۸۴	۲۳۴۰۹	۶۷	۰.۷۹	۲۳۸۲۸	۱۰۰	۹														
۰.۷۸۶	۴۲۱۸۲	۲۴	۰.۸۴۳	۴۱۹۱۹	۲۲	۰.۸۵۳	۴۱۰۲۵	۸	۰.۷۱	۴۱۵۱۲	۵۶	۰.۷۱۲	۴۱۰۱۸	۱۱۳	۱۰														
۰.۷۸۷	۳۴۰۴۸	۶	۰.۷۸۳	۳۸۴۶۸	۱۰	۰.۸۰۱	۳۴۴۴۵	۱	۰.۷۷۷	۳۳۲۱۱	۱۸	۰.۷۷۲	۴۱۶۰۹	۱۶۲	۱۲														
۰.۴۴۱	۲۵۷۸۵	۲۷	۰.۸۴۸	۲۸۴۷۴	۱۳	۰.۹۴۴	۲۵۴۷۶	۲	۰.۷۷۲	۴۱۲۴۶	۷۵	۰.۷۵۲	۲۸۷۹۱	۱۹۱	۱۳														
۰.۴۴۱	۴۶۷۸۱	۲۰	۰.۴۴۵	۴۲۸۱۶	۲۵	۰.۵۳۳	۲۵۵۱۶	۶	۰.۷۷۷	۴۱۲۴۶	۴۴	۰.۷۷۴	۴۵۴۹۹	۶۷	۱۴														
۰.۶۲۴	۴۷۲۹۹	۴۷	۰.۶۲۷	۳۴۷۰۷	۲۵	۰.۷۴۴	۴۷۵۰۲	۵	۰.۶۲۱	۴۸۱۰۹	۹۰	۰.۶۱	۴۸۰۴۹	۱۹۰	۱۵														
۰.۵۵۵	۴۲۵۵۰	۱۱	۰.۶۵۱	۴۶۹۸۷	۷	۰.۷۵۶	۴۱۲۹۰	۲	۰.۵۴۳	۴۴۹۹۷	۲۰	۰.۵۳۱	۵۰۴۳۵	۵۶	۱۶														
۰.۷۸۳	۵۲۱۲۷	۱۱	۰.۷۸۳	۷۸۶۲۷	۱۴	۰.۷۰۶	۴۳۶۰۰	۶	۰.۶۵۲	۵۰۷۰۴	۱۰۷	۰.۵۳۵	۵۲۲۶۴	۲۲۱	۱۷														
۰.۵۷۱	۵۲۷۱۴	۲۳	۰.۶۴۴	۵۲۸۴۹	۲۹	۰.۷۰۹	۵۳۰۰۰	۱۱	۰.۵۳۳																				
۰.۷۸۶	۵۵۰۵۵	۱۷	۰.۸۵۵	۵۴۸۴۳	۵	۰.۸	۴۹۰۴۲	۱	۰.۷۶	۵۶۱۹۵	۴۹	۰.۷۶۹	۵۵۵۰۰	۲۰۸	۱۹														
۰.۶۶۶	۵۵۰۵۵	۲۰	۰.۶۵۳	۵۶۷۸۳	۲۱	۰.۷۱۲	۴۲۱۸۸	۵	۰.۵۵۳	۶۱۲۰۷	۵۱	۰.۵۵۴	۶۱۶۵۵	۱۱۵	۲۰														
۰.۸۴۳	۶۲۱۲۹	۲۶	۰.۹۴۴	۶۲۶۲۹	۲۳	۰.۹۷	۲۵۵۹۴	۳	۰.۸۴۶	۶۲۸۳۰	۱۵۳	۰.۸۶۵	۶۲۷۹۵	۷۸۹	۲۱														
۰.۶۸۱	۵۳۲۰۰	۲۴	۰.۵۳۲	۶۲۰۱۲	۲۱	۰.۷۷۹	۵۱۷۲۳	۶	۰.۶۸۲	۶۱۵۹۱	۵۶	۰.۶۸۲	۶۰۴۱۳	۱۲۳	۲۲														
۰.۷۸۳	۶۲۰۶۶	۵۵	۰.۷۵۹	۶۲۶۸۷	۴۵	۰.۹۴۲	۶۲۰۹۸	۶	۰.۷۷	۶۲۱۹۱	۸۵	۰.۷۲۳	۶۲۰۰۹	۱۷۶	۲۳														
۰.۹۰۹	۶۱۷۲۷	۲۰	۰.۸۶۳	۶۱۷۲۷	۱۸	۰.۹۲۲	۶۸۲۰۱	۵	۰.۷۱۶	۶۸۱۶۸	۷۱	۰.۷۲۴	۶۵۲۱۵	۱۵۳	۲۴														
۰.۹۰۸	۷۵۵۳۶	۲۱	۰.۹۲۵	۶۵۰۸۵	۱۵	۰.۹۳۱	۷۳۲۸۹	۴	۰.۸۵۸	۷۵۱۹۳	۷۲	۰.۸۵۶	۷۴۶۱۶	۱۱۸	۲۵														
۰.۸۴۶	۷۶۲۸۷	۲۸	۰.۸۴۶	۷۶۲۸۷	۲۸	۰.۸۶۸	۷۴۰۴۶	۶	۰.۸۴۵	۷۶۵۷۳	۱۰۵	۰.۸۴۱	۷۷۷۶۵	۲۱۵	۲۶														
۰.۶۰۶	۶۸۵۳۷	۴۳	۰.۶۰۶	۷۸۲۶۶	۴۸	۰.۶۶۶	۷۸۴۴۳	۱۱	۰.۵۵۳	۷۶۴۵۰	۷۶	۰.۵۵۳	۷۸۱۲۳	۱۱۳	۲۷														
۰.۸۴۱	۶۸۶۶۱	۲۴	۰.۸۴۴	۶۸۰۸۷	۲	۰.۸۶۶	۶۸۷۷۱	۱	۰.۷۱۴	۷۵۵۸۵	۱۵	۰.۷۲۴	۸۰۷۲۳	۶۸	۲۸														
۰.۸۱۲	۸۶۹۲۹	۱۰۱	۰.۸۵۸	۸۴۵۰۸	۸۰	۰.۸۷	۸۴۶۵۵	۲۵	۰.۸۱۱	۸۳۳۳۴	۱۰۲	۰.۸۷۷	۸۵۵۵۷	۲۰۹	۲۹														
۰.۷۲۶	۴۲۷۸۷۴	۲۹۱۲	۰.۷۶۴	۴۶۰۹۷۷	۲۵۱	۰.۸۴۴	۴۲۷۱۶۲	۶۱۵۳	۰.۶۵۵	۴۴۴۰۷	۶۵۵۳	۰.۶۹۱	۴۴۹۰۲۴	۱۴۴۵	۳۰														

در بخش روش پیشنهادی بیان شد که پیشنهاد پنجم روشی احتمالی است، از این رو، این واقعیت با برابر شدن پیشنهاد چهارم و پنجم در گروه‌های ۹، ۱۰، ۱۴، ۲۰ و ۲۸ جدول ۲ و گروه‌های ۴، ۱۱ و ۲۶ در جدول ۳، بیشتر آشکار شده است که دلیل آن، نبود کاربر مدیر در بین کاربران ورودی است.

تحلیل نتایج اعتبارسنجی و آزمون برای هر یک از پیشنهاد‌های روش پیشنهادی

در این زیربخش، نتایج حاصل از اعتبارسنجی هر یک از پیشنهادها با نتایج حاصل از آزمون مقایسه شده است. شکل ۲ میانگین خطای حاصل از دو دسته اعتبارسنجی و آزمون را برای هر پیشنهاد نشان می‌دهد. بر اساس این شکل، دو نمودار اعتبارسنجی و آزمون با یکدیگر هم‌جهت هستند و هر دو نسبت به پیش‌بینی کاربران پاسخ یکسانی به دست آورده‌اند. البته، در پیشنهاد سوم این دو دسته از یکدیگر به میزان $0/0305$ فاصله گرفته‌اند که با توجه به اینکه دسته آزمون خطای کمتری را به دست آورده است، می‌توان گفت روش پیشنهادی یکپارچه است. طبق شکل ۲، دسته آزمون در پیشنهاد نخست به میزان $0/0089$ ، در پیشنهاد دوم $0/010$ ، در پیشنهاد سوم $0/030$ ، در پیشنهاد چهارم $0/014$ و در پیشنهاد پنجم به میزان $0/0095$ نسبت به دسته اعتبارسنجی کاهش یافته است. این نتایج، نشان‌دهنده یکپارچگی روش پیشنهادی و پاسخ واحد روش، نسبت به تطابق دو دسته اعتبارسنجی و آزمون هستند.



شکل ۲. میانگین خطای حاصل از دو دسته اعتبارسنجی و آزمون برای هر یک از پیشنهاد‌های روش پیشنهادی

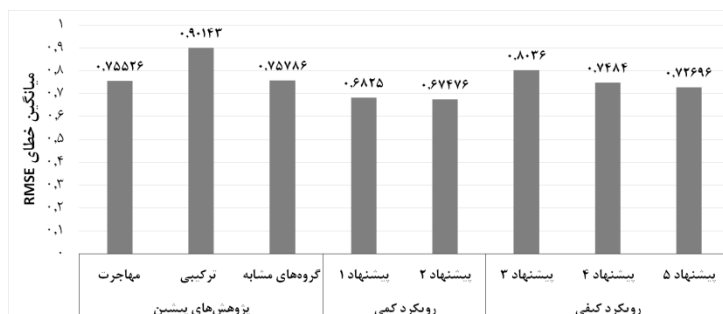
مقایسه خطای روش پیشنهادی با پژوهش‌های پیشین

در این زیربخش، میانگین خطای هر یک از پیشنهاد‌های روش پیشنهادی با سه پژوهش پیشین مقایسه شده است. هر یک از پژوهش‌های پیشین روی هر یک از گروه‌های دسته آزمون ارزیابی شده‌اند و نتایج در جدول ۴ نشان داده شده است. همچنین، میانگین خطای هر یک از پیشنهاد‌های روش پیشنهادی با پژوهش‌های پیشین بر اساس نتایج حاصل از جدول‌های ۳ و ۴ در شکل ۳ نشان داده شده است.

جدول ۴. نتایج ارزیابی سه پژوهش پیشین روی گروه‌های دسته آزمون

روش ترکیبی			روش مهاجرت			روش گروه‌های مشابه			شناسه گروه
E	T	G	E	T	G	E	T	G	
۰/۹۳۳	۶۴۱۸	۵	۰/۷۳۸	۱۵۴۷۶	۱	۰/۷۳۸	۱۵۴۷۶	۱	۱
۰/۹۲	۱۳۴۷۶	۱۳	۰/۶۱۹	۱۵۰۹۲	۱۵	۰/۶۴۲	۷۸۴۱	۱۰	۲
۰/۹۶۸	۴۷۳۶	۹	۰/۶۳۹	۱۵۱۷۰	۱۶	۰/۶۴۶	۱۶۵۶۵	۱۱	۳
۰/۹۹۸	۱۸۹۴۸	۱۴	۰/۲۴۷	۲۱۰۰۵	۲	۰/۳۲۶	۸۰۹۳۷	۱	۴
۰/۹۸	۲۰۶۶۱	۱۵	۰/۹۹۵	۴۶۹۲۷	۱	۰/۹۵۷	۵۵۸۱۷	۱	۵
۰/۸۶۲	۲۲۴۲۴	۱۷	۰/۶۲	۱۷۰۹۴	۴	۰/۶۴۱	۱۶۵۲۰	۳	۶
۰/۹۹۴	۳۱۲۳۹	۷	۰/۶۵۷	۱۹۸۶۱	۴	۰/۶۵۷	۱۹۴۱۲	۳	۷
۰/۶۹۷	۳۸۷۲۵	۳۲	۰/۵۹۲	۳۹۳۱۳	۵	۰/۵۹۹	۳۸۵۴۵	۴	۸
۰/۹۲۱	۳۶۱۳۸	۸	۰/۹۰۸	۳۵۵۶۴	۲	۰/۹۰۷	۵۰۶۱۴	۱	۹
۰/۹۳۵	۴۱۴۷۷	۲۱	۰/۷۵۶	۲۷۰۷۸	۳	۰/۸۵۵	۱۸۳۰۲	۱	۱۰
۰/۹۳۸	۴۱۳۲۰	۳۵	۰/۷۶۵	۴۳۶۲۷	۸	۰/۸۱۶	۳۹۵۸۴	۹	۱۱
۰/۸۷۶	۳۶۹۵۷	۱۱	۰/۸۸۵	۲۶۹۰۳	۲	۰/۸۱۳	۲۰۱۱۹۳	۱	۱۲
۰/۸۲۵	۳۸۰۱۳	۱۹	۰/۸۴۸	۳۴۹۶۴	۲	۰/۸۲۷	۴۱۸۳۰	۱	۱۳
۰/۹۰۷	۴۱۹۹۵	۲۶	۰/۸۸۴	۴۶۱۰۴	۳	۰/۸۶۸	۴۴۶۷۱	۳	۱۴
۰/۸۵۲	۴۴۴۰۲	۳۳	۰/۷۵۷	۵۲۱۴۵	۹	۰/۶۸۷	۴۹۴۴۲	۸	۱۵
۰/۷۵۳	۴۱۴۷۰	۱۷	۰/۶۲۷	۵۳۰۵۶	۹	۰/۶۱۱	۴۸۴۳۸	۹	۱۶
۰/۸۷۵	۵۰۴۸۷	۱۷	۰/۸۱۹	۵۰۸۷۱	۲	۰/۸۳۸	۲۰۱۱۹۳	۱	۱۷
۰/۸۳۶	۳۸۵۲۱	۱۲	۰/۷۸۸	۲۵۲۲۵	۹	۰/۷۷۳	۳۲۶۰۵	۹	۱۸
۰/۹۹۷	۵۴۳۳۵	۳۴	۰/۸۴۵	۵۴۵۰۱	۱۴	۰/۸۴۳	۵۹۵۵۲	۱۲	۱۹
۰/۸۴۱	۵۹۱۹۴	۵	۰/۶۶۶	۲۷۴۳۴	۴	۰/۶۰۶	۵۶۰۰۶	۳	۲۰
۰/۹۵۲	۴۵۴۵۸	۲۷	۰/۹۷۲	۶۹۴۶۳	۱	۰/۹۷۲	۷۶۸۸۹	۱	۲۱
۰/۹۶۴	۶۶۶۲۲	۱۶	۰/۹۴۷	۶۰۰۴۵	۴	۰/۹۶۳	۶۷۳۵۴	۱	۲۲
۰/۹۱۶	۶۷۴۱۲	۲۶	۰/۷۶	۶۸۴۲۴	۹	۰/۷۶۵	۶۸۴۲۴	۹	۲۳
۰/۸۵۴	۷۲۳۱۵	۳۱	۰/۸۱	۷۲۸۹۶	۱۳	۰/۸۴۶	۷۱۳۹۷	۱۰	۲۴
۰/۹۹۷	۷۰۳۵۷	۲۰	۰/۷۱۴	۶۰۰۲۷	۸	۰/۷۳۶	۷۳۷۲۸	۲	۲۵
۰/۹۹۷	۷۴۲۰۰	۱۲	۰/۵۹	۷۷۴۴۳	۱۲	۰/۵۷	۷۶۱۰۲	۱۰	۲۶
۰/۸۳۳	۷۵۲۶۹	۵۸	۰/۶۳۲	۷۸۵۶۶	۱۷	۰/۶۶۵	۷۹۱۶۴	۱۷	۲۷
۰/۹۵۳	۷۹۸۵۱	۵۹	۰/۸۷۸	۷۸۳۳۰	۲	۰/۸۳۲	۷۲۲۴۱	۷	۲۸
۰/۹۳۹	۷۵۲۹۷	۲۹	۰/۷۷۷	۷۸۴۸۳	۸	۰/۷۵۸	۶۷۴۷۰	۲	۲۹
۰/۷۳	۷۶۸۵۱	۱۶	۰/۹۲۳	۸۵۲۶۸	۱	۰/۹۷۹	۷۷۰۳۶	۱	۳۰
۰/۹۰۱	۴۶۱۵۲/۲	۲۱/۴	۰/۷۵۵	۴۶۵۴۵/۱	۶/۳	۰/۷۵۷	۶۰۸۱۱/۶	۵/۰۶	میانگین

بر اساس شکل ۳، پیشنهاد دوم، میانگین خطای RMSE کمتری به نسبت سایر پیشنهادها و سه پژوهش پیشین دارد. با توجه به اینکه پیشنهاد دوم روش گروه‌های مشابه را در بر دارد، اما توانسته است میانگین خطای RMSE را نسبت به روش گروه‌های مشابه به $0/0831$ کاهش دهد. این پیشنهاد نسبت به روش ترکیبی بهبود چشمگیری به دست آورده است، به طوری که میانگین خطای RMSE را نسبت به روش ترکیبی به $0/226$ کاهش داده است. همچنین پیشنهاد دوم میانگین خطای RMSE را نسبت به روش مهاجرت به $0/0805$ کاهش داده است. به دلیل اینکه پیشنهاد دوم هیچ گونه کیفیتی را در نظر ندارد، این احتمال وجود دارد که کاربران پیشنهادی از گروه‌های هر زمانه پیشنهاد شوند. از این رو، در صورتی که بخواهیم نتایج نهایی با کیفیت باشند، پیشنهاد پنجم گزینه مناسبی است. بر اساس شکل ۳، پیشنهاد پنجم توانسته است میانگین خطای RMSE را نسبت به روش گروه‌های مشابه به $0/0309$ ، نسبت به روش ترکیبی به $0/174$ و همچنین نسبت به روش مهاجرت به $0/0283$ کاهش دهد. با توجه به شکل ۳، به ترتیب در پیشنهاد چهارم و پنجم هر قدر کیفیت افزایش می‌یابد، خطا نیز کاهش می‌یابد. رابطه عکس در میان کیفیت و خطا، در پیش‌بینی کاربران هدف، بسیار مهم است.

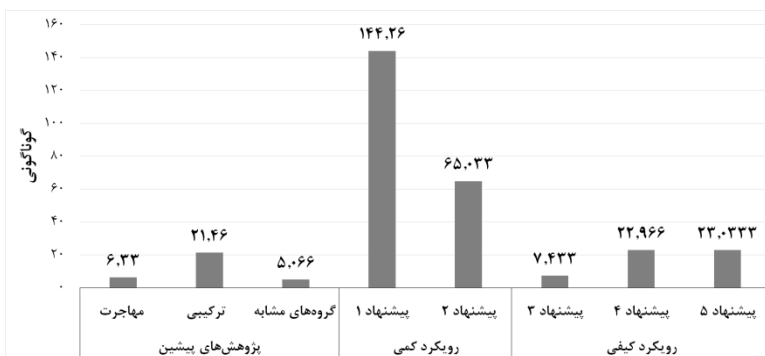


شکل ۳. مقایسه خطای هر یک از رویکردهای روش پیشنهادی با سه پژوهش پیشین

مقایسه گوناگونی روش پیشنهادی با پژوهش‌های پیشین

در سامانه‌های پیشنهاددهنده علاوه بر بررسی میزان دقت یا خطای پیش‌بینی، معیاری برای بررسی اثرگذاری و سنجش عملکرد سامانه، با عنوان گوناگونی مطرح می‌شود و گوناگونی موارد پیشنهادشده را بر اساس تنوع و یکسان نبودن، بررسی می‌کند. در واقع این معیار، سعی در مشابهت نداشتن موارد پیشنهادی دارد. به طور مثال، در سامانه پیشنهاددهنده فیلم، هرچه فیلم‌های پیشنهادی از سبک‌های مختلف بوده و بازیگران و کارگردان‌ها مختلف باشند، گوناگونی بیشتر است. شایان ذکر است که گوناگونی هیچ ارتباطی با خطا یا دقت ندارد و برای مثال ممکن است دقت یک سامانه ۱۰۰ درصد باشد، اما گوناگونی نداشته باشد. برای محاسبه گوناگونی رابطه‌ای خاص و واحدی وجود ندارد و محاسبه آن در هر سامانه با توجه به هدف، متفاوت است.

با توجه به ماهیت این مقاله، کاربران از ادغام اعضای گروه‌های رتبه‌بندی شده پیشنهاد می‌شوند. برای هر پیشنهاد در روش پیشنهادی و هر یک از پژوهش‌های پیشین، این کمیت را بررسی می‌کنیم که هر روش، کاربران را از ادغام اعضای چه تعداد گروه پیشنهاد می‌دهد. هرچقدر کاربران از گروه‌های بیشتری پیشنهاد شده باشند، گوناگونی بیشتر آن روش در پیشنهاد کاربران از گروه‌های مختلف را نشان می‌دهند. در واقع، این روش گوناگونی، بیان می‌کند که کاربران پیشنهادی، از ترکیب اعضای چند گروه پیشنهاد شده‌اند. در جدول‌های ۳ و ۴ ستون G تعداد گروه‌های مورد نیاز را برای هر پیشنهاد و پژوهش‌های پیشین، برای دستیابی به تعداد کاربران هدف، نشان می‌دهد. در واقع، این ستون نشان‌دهنده گوناگونی است. بدین ترتیب که کاربران نهایی پیشنهاد داده‌شده از ترکیب چه تعداد گروه به دست آمده‌اند. شکل ۴ گوناگونی هر یک از پیشنهاد‌های روش پیشنهادی را با سه پژوهش پیشین بر اساس ستون G در جدول‌های ۳ و ۴ نشان می‌دهد. با توجه به شکل ۴، روش گروه‌های مشابه، نسبت به تمامی روش‌ها، گوناگونی کمتری دارد، زیرا این روش، همیشه گروه‌های با تعداد S زیاد را به ابتدای فهرست می‌آورد و همین دلیل باعث می‌شود تا برای دستیابی به هدف، به گروه‌های کمتری نیاز باشد. برخلاف روش گروه‌های مشابه، پیشنهاد نخست در بخش کمی نسبت به تمامی روش‌ها، گوناگونی بیشتری به دست آورده که دلیل آن، رتبه‌بندی گروه‌های با تعداد عضو کم است.



شکل ۴. مقایسه گوناگونی هر یک از رویکردهای روش پیشنهادی با سه پژوهش پیشین

بر اساس شکل ۴، به دلیل اینکه روش گروه‌های مشابه نقطه شروع در پژوهش حاضر است، تعداد گروه مورد نیاز یا گوناگونی از این روش تا پیشنهاد نخست به‌طور میانگین از $5/06$ به $144/27$ افزایش یافته است. دلیل افزایش گوناگونی این است که وقتی متغیر C به‌دست آمده از پیشنهاد نخست، تقسیم بر S می‌شود، گروه‌های با تعداد S کم به ابتدای فهرست می‌آیند که موجب می‌شود برای دستیابی به هدف به گروه‌های بیشتری نیاز باشد. گوناگونی از پیشنهاد نخست تا پیشنهاد دوم به‌طور میانگین از $144/27$ به $65/03$ کاهش یافته است. دلیل کاهش این است که وقتی کسر C/S به‌دست آمده از پیشنهاد نخست در $\log_{10} S$ ضرب می‌شود، باعث خنثی شدن اثرگذاری زیاد متغیر S در مخرج کسر C/S می‌شود که موجب از

بین رفتن رتبه‌بندی حداکثری گروه‌های با S کم و زیاد می‌شود. گوناگونی از پیشنهاد دوم به سوم به‌طور میانگین از $۶۵/۰۳$ به $۷/۴۳$ کاهش یافته است. کاهش گوناگونی در پیشنهاد سوم نشان می‌دهد که رتبه‌بندی گروه‌ها بر اساس تعداد مدیر کم، همانند روش گروه‌های مشابه، گروه‌های با تعداد S زیاد را به ابتدای فهرست می‌آورد. این کاهش بیانگر این است که اکثر گروه‌های با تعداد S زیاد دارای S_A کمتری هستند. گوناگونی از پیشنهاد سوم به چهارم به‌طور میانگین از $۷/۴۳$ به $۲۲/۲۹۶$ افزایش یافته است. دلیل افزایش گوناگونی در این پیشنهاد این است که وقتی کسر حاصل از پیشنهاد سوم نرمال شده و اثرگذاری خطی آن خنثی می‌شود، گروه‌های با S کم به ابتدای فهرست می‌آیند. همچنین از پیشنهاد چهارم به پنجم گوناگونی به‌طور میانگین از $۲۲/۹۶$ به $۲۳/۰۳$ افزایش یافته است. این افزایش بیانگر این است که گروه‌های کاربران مدیر ورودی، نسبت به پیشنهاد چهارم، S کمتری دارند که موجب می‌شود برای دستیابی به مجموعه هدف به گروه‌های بیشتری نیاز باشد.

معرفی روش نهایی پیشنهادی پژوهش با تحلیل نتایج خطا و گوناگونی

در این زیربخش، بهترین پیشنهاد روش پیشنهادی بیان شده و پاسخ مسئله این پژوهش، بررسی و تحلیل می‌شود. به‌طور کلی، مسئله و مشکل اصلی در پژوهش حاضر این است که با توجه به نبود ابزار جامعی در تلگرام برای جست‌وجوی کاربر، چگونه می‌توان کاربرانی مشابه با مجموعه‌ای کاربر را در پیام‌رسان تلگرام یافت. به بیان دیگر، چگونه می‌توانیم جامعه هدف جدیدی را برای یک بازاریاب، مشابه با مشتریان قدیمی او بیابیم؟

برای پاسخ به مسئله، در این پژوهش روشی را طی پنج پیشنهاد در دو بخش کمی و کیفی بررسی کرده‌ایم. هر یک از پیشنهادها، پیشنهاد قبل از خود را تکمیل می‌کنند. روش پیشنهادی در این پژوهش با دریافت مجموعه‌ای کاربر از فرد بازاریاب در هر یک از پیشنهادها، گروه‌های کاربران را به‌منظور دستیابی به کاربرانی مشابه، بررسی کرده است.

با توجه به نتایج بررسی خطا و گوناگونی در دو زیربخش پیشین به این نتیجه می‌رسیم که پیشنهاد دوم دارای کمترین خطا، پیشنهاد پنجم دارای بیشترین کیفیت و پیشنهاد نخست دارای بیشترین گوناگونی است. پیشنهاد نخست دارای بیشترین تنوع و گوناگونی در کاربران پیشنهاد شده است. اما مشکل این پیشنهاد، همان‌طور که در بخش روش پیشنهادی نیز بیان شد، رتبه‌بندی همیشگی گروه‌های با تعداد عضو کم است. پیشنهاد دوم پس از پیشنهاد نخست، از جنبه بیشترین گوناگونی، دارای تنوع در کاربران پیشنهاد شده است. این پیشنهاد علاوه بر گوناگونی دارای کمترین خطا نیز هست. پیشنهاد اول و دوم نسبت به پیشنهاد پنجم که کیفیت را در نظر دارد، هیچ‌گونه کیفیتی را در نظر ندارند. در این دو پیشنهاد این امکان وجود دارد که کاربران از گروه‌های بی‌کیفیت یا هرزنامه پیشنهاد شوند، اما گوناگونی این دو پیشنهاد نسبت به پیشنهاد پنجم بیشتر است.

سه نکته مهم برای بررسی نهایی پیشنهادهای ارائه شده در این پژوهش، خطا، گوناگونی و کیفیت است. به‌طور کلی، نتیجه‌گیری نهایی این است که پیشنهاد دوم از جنبه کمترین خطا و بیشترین

گوناگونی، توانسته است کاربرانی مشابه با مجموعه کاربران ورودی را پیش‌بینی کند. همچنین پیشنهاد پنجم از جنبه بیشترین کیفیت، توانسته است کاربرانی باکیفیت و مشابه با مجموعه کاربران ورودی، پیش‌بینی کند. پیشنهاد دوم نسبت به پیشنهاد پنجم که کیفیت را در نظر دارد، هیچ‌گونه کیفیتی را ارائه نمی‌دهد. از این رو، در صورتی که برای بازاریابان پارامتر کیفیت مهم باشد، پیشنهاد پنجم و در صورتی که خطا و گوناگونی مد نظر آنان باشد، پیشنهاد دوم بهترین گزینه است. شایان ذکر است که اختلاف خطای پیشنهاد پنجم نسبت به پیشنهاد دوم، ۰/۰۵۲ است.

نتیجه‌گیری و پیشنهادهای آینده

در این مقاله با استفاده از گراف ارتباطی کاربران که بیانگر ارتباط کاربران از طریق گروه‌های مشترک در تلگرام است، روشی برای پیشنهاد کاربران ارائه شده است. روش پیشنهادی دارای دو بخش کمی و کیفی طی پنج پیشنهاد ایجاد شده است. بخش کمی، گروه‌ها را بر اساس تعداد و درصد کاربران اشتراکی با مجموعه کاربر ورودی رتبه‌بندی می‌کند. بخش کیفی در ادامه بخش کمی ارائه شده است که گروه‌ها را بر اساس ویژگی تعداد مدیر در هر گروه رتبه‌بندی می‌کند. همچنین بخش کیفی، کاربران ورودی را از جهتی دیگر و بر اساس مدیر بودن در گروه‌ها بررسی کرده است و به گروه‌های این نوع از کاربران، اولویت بیشتری در رتبه‌بندی می‌دهد. هر یک از پیشنهادهای ارائه‌شده در بخش کمی و کیفی، پیشنهاد قبل از خود را تکمیل کرده و می‌توان هر یک را روشی جداگانه برای پیشنهاد کاربر در نظر گرفت. در این مقاله، به‌منظور ارزیابی پیشنهادهای ارائه‌شده، از گروه‌های محبوب در تلگرام استفاده شده است. با بهره‌مندی از خبره دو دسته از گروه‌ها را که هر دسته شامل ۳۰ گروه محبوب در تلگرام است، برای اعتبارسنجی و آزمون شناسایی کرده‌ایم. میزان پیش‌بینی و خطای پیشنهادهای ارائه‌شده به‌صورت جداگانه توسط هر گروه محبوب از طریق تقسیم‌بندی داده‌ها به دو بخش آموزش و آزمون سنجیده شده و نتایج به‌دست‌آمده با سه پژوهش پیشین مقایسه شده‌اند. به‌طور کلی، نتایج حاصل از آزمایش‌ها، به‌میزان قابل توجهی بیانگر کاهش خطای پیشنهادهای ارائه‌شده نسبت به سه پژوهش پیشین هستند. پس از تحلیل نتایج دریافتیم، در صورتی که گروه‌ها را فقط بر اساس درصد کاربران اشتراکی با مجموعه کاربران ورودی در نظر بگیریم و اثرگذاری خطای درصد حاصل را به‌وسیله لگاریتم خنثی کنیم پیش‌بینی بهتر، خطای کمتر و گوناگونی بیشتری خواهیم داشت. همچنین با توجه به اینکه در این پژوهش، کیفیت گروه‌ها بر اساس تعداد مدیر کم در نظر گرفته شده است، به این نتیجه رسیدیم که بسیاری از گروه‌های باکیفیت تعداد اعضای زیادی دارند. علاوه بر این، طبق بررسی نتایج حاصل از بخش کیفی، در صورتی که گروه‌ها به‌صورت تعداد مدیر کم رتبه‌بندی شوند و اثرگذاری خطای درصد حاصل از آن، به‌وسیله جذر خنثی شود، علاوه بر خطای کمتر، کیفیت کاربران پیشنهادشده نیز تضمین خواهد شد. با توجه به نتایج به‌دست‌آمده از دو پیشنهاد ابتدایی در بخش کیفی، می‌توان دریافت که هرچقدر کیفیت گروه‌ها افزایش می‌یابد، خطا نیز کاهش می‌یابد. به‌طور کلی، در سامانه‌های پیشنهاددهنده در صورتی که با افزایش کیفیت، خطا کاهش یابد، آن

سامانه در مدل‌سازی کاربران بسیار ارزشمند است که این مورد در این سامانه پیشنهادی، اتفاق افتاده است. در نهایت، ارزیابی روش پیشنهادی به‌صورت جداگانه روی دو دسته اعتبارسنجی و آزمون، تکرار نتایج و یکپارچگی روش را نشان داده است. همچنین، با محاسبه گوناگونی به این نتیجه دست یافتیم که روش پیشنهادی در کاربران پیشنهادشده نسبت به سه پژوهش پیشین، تنوع قابل توجهی دارد.

برای بهبود روش پیشنهادی، در کارهای آتی با توسعه گراف ارتباطی، می‌توان اطلاعات بیشتری از کاربران و گروه‌ها را در نظر گرفت. در روش پیشنهادی، فقط ارتباط گرافی کاربران در نظر گرفته شده است و در آینده می‌توان از محتوای نمایه هر کاربر از جمله نام، نام خانوادگی، شناسه کاربری و پیام‌های به‌اشتراک گذاشته کاربران در گروه‌های عمومی استفاده کرد. همچنین، می‌توان روش‌های کیفیت‌سنجی بیشتری را برای رتبه‌بندی بهتر گروه‌ها در نظر گرفت. در این مقاله، سه پژوهش پیشین را مطرح کرده و با این مقاله مقایسه کرده‌ایم. در پژوهش‌های آینده می‌توان با ترکیب هر یک از این پژوهش‌ها با روش پیشنهادی و در نظر گرفتن تجربه‌های کاربران، نتایج بهتر و مؤثرتری به دست آورد.

فهرست منابع

- بیگدلو، مهدی؛ هادیان، ناصر (۱۳۹۷). تأثیر کارکردهای شبکه‌های اجتماعی بر سمت‌گیری‌های فرهنگ سیاسی کاربران دانشگاهی در ایران (مطالعه موردی: وایبر و تلگرام). *جامعه‌شناسی سیاسی جهان اسلام*، ۶ (۱۲)، ۷۳-۱۰۴.
- هاشمی، سید علی؛ زارع چاهوکی، محمدعلی (۱۳۹۷). توسعه بازاریابی با توصیه‌گر گروه‌های پیام‌رسان‌ها. *دوفصلنامه محاسبات و سامانه‌های توزیع‌شده*، ۱ (۱)، ۲۱-۳۰.
- Al Momani, M. A. M. (2020). The Effectiveness of Social Media Application "Telegram Messenger" in Improving Students' Reading Skills: A Case Study of EFL Learners at Ajloun University College/Jordan. *Journal of Language Teaching and Research*, 11(3), 373-378.
- AlSuwaidan, L., & Ykhlef, M. (2018). Interest-based clustering approach for social networks. *Arabian Journal for Science and Engineering*, 43(2), 935-947.
- Asnafi, A. R., Moradi, S., Dokhtesmati, M., & Naeini, M. P. (2017). Using mobile-based social networks by Iranian libraries: The case of telegram messenger. *Library Philosophy and Practice*, no. 1.
- Ghaffari, M., Rakhshanderou, S., Mehrabi, Y., & Tizvir, A. (2017). Using social network of telegram for education on continued breastfeeding and complementary feeding of children among mothers: a successful experience from Iran. *International Journal of Pediatrics*, 5(7), 5275-5286.
- Ghorbanzadeh, D., & Saeednia, H. R. (2018). Examining telegram users' motivations, technical characteristics, trust, attitudes, and positive word-of-mouth: evidence from Iran. *International Journal of Electronic Marketing and Retailing*, 9(4), 344-365.

- Han, X., Wang, L., Farahbakhsh, R., Cuevas, A., Cuevas, R., Crespi, N. & He, L. (2016). CSD: A multi-user similarity metric for community recommendation in online social networks. *Expert Systems with Applications*, 53(1), 14-26.
- Hashemi, A., & Zare Chahooki, M. A. (2019). Telegram group quality measurement by user behavior analysis. *Social Network Analysis and Mining*, 9(1), 33.
- Hashemi, A., & Zare Chahooki, M. A. (2021). GroupRank: Ranking Online Social Groups Based on User Membership Records. *Journal of AI and Data Mining*, 9(1), 45-57.
- Iqbal, M. (2021). Telegram Revenue and Usage Statistics (2020). Accessed 13 February 2021, URL: <https://www.businessofapps.com/data/telegram-statistics>.
- Jiang, F., Leung, C. K., & Pazdor, A. G. (2016). Big data mining of social networks for friend recommendation. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 921-922.
- Jibouni, A., Lotfi, D., El Marraki, M., & Hammouch, A. (2018). A novel parameter free approach for link prediction. In *2018 6th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 1-6.
- Karimpour, D., Zare Chahooki, M. A., & Hashemi, A. (2021). Telegram group recommendation based on users' migration. *26th International Computer Conference, Computer Society of Iran (CSICC)*, 1-6.
- Karimpour, D., Zare Chahooki, M. A., & Hashemi, A. (2021). User recommendation based on Hybrid filtering in Telegram messenger. *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, 1-7.
- Kumar, P., & Reddy, G. R. M. (2018). Friendship recommendation system using topological structure of social networks. In *Progress in Intelligent Computing Techniques: Theory, Practice and Applications*, 237-246.
- Li, S., Song, X., Lu, H., Zeng, L., Shi, M., & Liu, F. (2020). Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm. *Expert Systems with Applications*, 139(1), 112839.
- Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, 390(6), 1150-1170.
- Nobari, A. D., Reshadatmand, N., & Neshati, M. (2017). Analysis of Telegram, an instant messaging service. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2035-2038.
- Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, 1-35.
- Silva, N. B., Tsang, R., Cavalcanti, G. D. C., & Tsang, J. (2010). A graph-based friend recommendation system using genetic algorithm. In *IEEE congress on evolutionary computation*, 1-7.

Telegram Users' Recommendation Based on the Membership Graph and Measuring Groups with Users from the Target Community

Davod Karimpour

MSc. Student of Computer Engineering, Faculty of Computer Engineering, Yazd University, Yazd, Iran

Mohammad Ali Zare Chahooki¹

Assistant Prof., Faculty of Computer Engineering, Yazd University, Yazd, Iran

Abstract

Telegram is a cloud-based messenger with more than 500 million monthly active users. Telegram's features include the creation of supergroups, channels, bots, and secret chats that provide users with more capabilities than a messenger. Telegram is used as a social network in Iran; but it does not offer all the features of a social network, including finding a collection of users. Finding users is very useful for marketers to find the target audience. In this paper, a method in two parts is presented by users' communication graph. The first part is based on graphs and model groups based on the percentage of subscriptions with incoming users. The second section ranks the groups based on the number of managers. The data of this study includes information about 70 million users and 700,000 supergroups in Telegram. The information obtained has been crawled by the IdeKav System, and it is worth noting that this information has been obtained quite legally by the tool provided by Telegram to the public. In this article, the proposed method is evaluated based on popular groups in Telegram. Popular groups are groups in which a single topic is discussed. In this paper, the evaluation by popular Telegram groups has been done by two categories of validation and test, and the results of experimental experiments have shown the repetition of the results and the integrity of the proposed method. The proposed method with the aim of covering the whole graph in two parts has shown the effectiveness of the suggestions and reduction of forecasting error compared to the previous three studies.

Keyword: Find users, Social Networks, Telegram-based marketing, Find the target community of marketers