

SID



سرویس های ویژه



سرویس ترجمه تخصصی



کارگاه های آموزشی



بلاگ مرکز اطلاعات علمی



سامانه ویراستاری STES



فیلم های آموزشی

کارگاه های آموزشی مرکز اطلاعات علمی جهاد دانشگاهی

کارگاه آنلاین
بررسی مقابله ای متون (مقدماتی)

کارگاه آنلاین
پروپوزال نویسی و پایان نامه نویسی

کارگاه آنلاین آشنایی با پایگاه های اطلاعات علمی بین المللی و ترند های جستجو

دسته‌بندی کاربران هم‌تراز در شبکه‌های اجتماعی با استفاده از خوشه‌بندی گراف

زینب سلامتی

دانشجوی کارشناسی ارشد دانشگاه آزاد بیرجند

Zeynab.salamat@gmail.com

مصطفی سبزه‌کار

دانشجوی دکتری دانشگاه فردوسی مشهد

Sabzekar.mostafa@gmail.com

محمد مهماندوست

کارشناسی ارشد هوش مصنوعی دانشگاه آزاد قزوین

Mohammad_mehmandoost@hotmail.com

چکیده

از مهمترین پدیده‌های نوظهور در فضای مجازی، که امکان تشکیل اجتماع‌های مجازی را فراهم می‌کند، رسانه‌های اجتماعی یا به عبارت بهتر شبکه‌های اجتماعی مجازی است که با رشد دستگاه‌های موبایل کارآمدتر، پیشرفت روزافزونی داشته است. شبکه‌هایی نظیر فیس‌بوک، توییتر، لینکدین و... که امکان گفتگو و تبادل اطلاعات و همچنین عضویت افراد در گروه‌هایی با موضوعات خاص و علایق مشترک را در اختیار کاربران قرار می‌دهد. به دلیل قابلیت اشتراک گذاری علایق و ویژگی‌های رفتاری کاربران در شبکه‌های اجتماعی، میتوان به حجم عظیمی از اطلاعات مربوط به کاربران دست پیدا کرد و با روشهای داده‌کاوی الگوهای رفتاری مشترک کاربران را از پست‌هایی که در شبکه‌های اجتماعی می‌گذارند، استخراج کرد. در این مقاله ما الگوریتم CI-GBI که از قوانین کشف رابطه‌های انجمنی در پیدا کردن زیرگرافهای مشابه با استفاده از شبه قطع‌بندی استفاده میکند، به منظور دسته‌بندی کاربران با خصوصیات رفتاری مشابه در شبکه‌های اجتماعی بهره خواهیم برد.

واژگان کلیدی: شبکه‌های اجتماعی، کاربران هم‌تراز، داده‌کاوی، الگوریتم کشف زیرگراف، خوشه‌بندی گراف

مقدمه

قدرت زیرساختار Big Data، جمع‌آوری انبوهی داده از جنبه‌های زندگی روزانه ما به واسطه شبکه‌های اجتماعی است. شبکه‌های اجتماعی بازتابی از زندگی فیزیکی ماست که روز به روز با جزئیات بیشتری اطلاعات مربوط به علایق فردی و روابط گوناگون انسانها را با یکدیگر جمع‌آوری میکند. اگرچه انبوه داده‌هایی که از شبکه‌های اجتماعی جمع‌آوری شده همچنان به صورت یک چالش حل‌نشده باقی مانده است، اما از جهتی میتوان به دورنمایی واقعی از زندگی کاربران پی برد. داده‌های اجتماعی که با رشد موفق سرویسهای شبکه اجتماعی به سرعت در حال افزایش هستند اطلاعاتی از کاربران نظیر روابط، اشتراکها، بلاگها و نظرات تولید میکنند و تجربیاتی از زندگی واقعی افراد در زمینه‌های مختلف را بازتاب میکنند مثل Facebook، Twitter، LinkedIn، Pinterest، Foursquare و غیره. چنین شبکه‌های اجتماعی فرصت مناسبی برای محققان است تا طیف کاملی از پدیده‌های جدید مربوط به کاربران را بفهمند. مثلاً متخصصان داده میتوانند ارتباطات و مشخصه تاثیرگذار بر افراد یک گروه جمعیتی بزرگ و یا ارتباطات استاتیک یا پویای بین کاربران را کشف کنند. شبکه‌های اجتماعی مشخصاً همه دارای خصوصیات مشترکی هستند. کاربران میتوانند با عضویت در این شبکه‌ها از امکانات آن استفاده کنند، صفحه‌ای مربوط به پروفایلشان در اختیارشان قرار داده میشود که در آن میتوانند خصوصیات شخصی خود و علایقشان را بیان کنند. میتوانند اطلاعات مربوط به تماس خود مثل ایمیل، شماره‌های تماس و... را با دیگران به اشتراک بگذارند. با دوستان خود در تماس باشند و در گروههایی مطابق با علاقه خود در کنار افراد مشابه خود گرد هم آیند و راجع به مسائل مختلف گفتگو و ابراز رضایت یا عدم رضایت کنند. نظرات خود را بیان کنند و آن را با اشتراک بگذارند. مثلاً افراد علاقه مند به موسیقی سنتی یا علاقه مند به یک شخصیت سیاسی یا هنری خاص و... میتوانند در یک گروه عضو شوند و نظرات خود را بیان کنند. شبکه‌های اجتماعی برخط، بخصوص آنهایی که کاربردهای معمولی و غیرتجاری دارند، مکانهایی در دنیای مجازی هستند که مردم در آنها خود را به طور خلاصه معرفی میکنند و امکان برقراری ارتباط بین خود و همفکران خود در زمینه‌های مختلف مورد علاقه فراهم میسازد. با توجه به ابعاد مختلف شبکه‌های اجتماعی و تحقیقات گسترده‌ی انجام شده در این زمینه بررسی جامع همه‌ی این ابعاد به زمان و مطالعات بیشتری نیاز دارد، از این رو در این مقاله بر آن شدیم تا بخشی از این تحقیقات را مورد مطالعه قرار دهیم. با گسترش و وسعت شبکه‌های مجازی، کنترل و پردازش اطلاعات دریافتی تقریباً امری ناممکن شده است، نیاز به وجود سامانه‌ای که بتواند بر مشکلات موجود برای دسترسی کاربران به اطلاعاتشان غلبه کند، بیش از هر زمانی حس می‌گردد. سیستم‌هایی که بتوانند از میان خیل عظیمی از اطلاعات و محصولات گوناگون، مناسب‌ترین و مورد پسندترین‌ها را بنا به شرایط و ویژگیهای خاص هر کاربر به وی پیشنهاد کنند، طرفداران بسیاری یافته‌اند. "سیستم‌های پیشنهاد گر"^۱ سیستم‌های هوشمندی هستند که در فضای اینترنت با شناسایی علایق و اولویتهای کاربر، اطلاعات موجود را پالایش کرده و پیشنهادات مناسب و مرتبط را به تک تک کاربران ارائه می‌کنند. به همین دلیل دسته‌بندی کاربران بر اساس علایق (شباهت) امر بسیار مهمی میباشد چرا که ارائه پیشنهاد مرتبط با علایق (شباهت) کاربران زمانی امکان پذیر میباشد که بتوانیم کاربران را دسته‌بندی نماییم. یکی از نیازهای مهم سیستم‌های پیشنهاددهنده کشف الگوهای تعاملی کاربران به دست جست‌وجوی کاربران شبکه‌های اجتماعی و علاقه‌مندی‌ها و نرفتهای مشترک آنهاست. تحلیل شبکه‌ی اجتماعی در این سیستم‌ها نیز سودمند است تا به روابط بین افراد مختلف در سیستم رسیدگی کند و ترجیحات کاربران را پیش بینی کند. بنابراین تحلیل شبکه‌ی اجتماعی میتواند برای تعیین روابط بین کاربران به کار برده شود تا در ساخت پیشنهادهای کمک کند. کارایی سیستم‌های پیشنهادی با شگردهای تحلیل شبکه‌های اجتماعی ارزیابی میشود. برای تسهیل این

¹-Recommender system/Recommending systems

کار، درصد پیش‌بینی داده شده به دست سیستم می‌تواند با استنتاج تحلیل شبکه‌ی اجتماعی از مشتری‌های شبکه پیش‌بینی شود. در این مقاله هدف ما دسته بندی کاربران هم‌تراز به منظور حصول نسبتاً دقیق به کاربران با ویژگی‌های مشابه می‌باشد. مثلاً سعی خواهیم کرد کاربران را در دسته‌هایی با علایق مشترک مانند هنر، سیاست، مسائل اجتماعی، ورزش، مذهب،... طبقه بندی کنیم. به این منظور به کشف الگویی در روابط بین کاربران نیاز داریم. ما از این ویژگی که کاربران هم‌تراز به احتمال زیاد دارای محاوره بیشتر نسبت به سایر اعضا خواهند بود و همچنین در گروه‌هایی با موضوع مشترک گرد هم جمع خواهند شد استفاده خواهیم کرد. برای کشف الگو احتیاج به مجسم‌سازی شبکه‌های اجتماعی داریم. (ساعده، 1393) در مقاله- ای شبکه‌های اجتماعی را بوسیله گراف مجسم سازی کردند که در آن کاربران گره‌های گراف و یالها ارتباط بین آنان است. ارتباطات رکن اصلی سیستم‌های اجتماعی بوده و ارتباطات برخط دارای ویژگی‌هایی نظیر حصول (به دست آوردن شنوندگان جهانی)، دسترسی، قابلیت کاربرد مجدد، پاسخ دهی سریع و پایداری است. همچنین این ارتباطات در شبکه‌های اجتماعی برخط با سبک‌هایی همچون پیام‌ها، نظرها، گفت و گو در رسانه‌های اشتراک مصنوعی، فعالیت‌های اجتماعی ظاهر میشوند. (کجباف، 1391)

سوابق

در سال 2002 در مقاله‌ای به بررسی تئوری و عملی الگوریتم GBI و برنامه‌های کاربردی پیش پرداخته و نشان داده که این الگوریتم می‌تواند روی هر نوع گرافی مثل مستقیم یا غیرمستقیم با لینکها و نودهای رنگی یا غیررنگی، با حلقه یا بدون بی‌حلقه کار کند. پیچیدگی زمانی این الگوریتم اکثراً خطی است و به درجه گراف بستگی دارد و در data set از DNA تعدادی از افراد آزمایش شده است و به نتایج رضایت‌بخشی رسیده است. (Matsuda, Motoda, & Washio, 2002)

در سال 2005 در مقاله‌ای الگوریتم جدیدی از خانواده GBI معرفی کرده که در آن ضعف ناشی از عدم شناسایی زیرگراف‌های مشابه را پوشانده است. در این روش الگوریتم جدیدی به نام CI-GBI معرفی کرده که در ابزار جستجوی گراف قابلیت جدیدی برای پیدا کردن زیرگراف‌های مشابه را دارا می‌باشد. این الگوریتم در مرحله‌ای که زیر گراف قابلیت تکه‌تکه شدن ندارد نودهای مشابه را شناسایی و خارج می‌کند. (Nguyen, Ohara, Motoda, & Washio, 2005)

در سال 2012 نتایج جستجوی حاصل از وب را خوشه بندی کرده‌اند و با استفاده از گراف از یک ردیف متنی کلمات مشابه را خارج کردند ولی چون امکان کاربرد کلمات در مباحث مختلف وجود دارد، ردیفی از کلمات مربوط به کلمه مورد نظر را ایجاد کردند و با محدودیت جستجو به نتایج دقیق‌تری رسیدند. در یک مرحله کلمه را جستجو کردند و زیرگراف‌های مربوط به آن را خارج کردند و در زیرگراف بعدی کلمه‌ای نزدیک به کلمه مورد نظر را جستجو کردند. در نهایت به زیرگرافی خواهند رسید که قابلیت شکسته شدن ندارد. ایده پیشنهادی این مقاله ارائه الگوریتمی به نام WSI است که هدفش کشف خودکار معانی کلمه ای است که دروب جستجو شده است و توسط خوشه‌بندی در خوشه‌ای که مناسب ترین جواب است قرار داده میشود. (Di Marco & Navigli, 2012)

در سال 2012 با فرض تحت تاثیر قرارگرفتن فرایند یادگیری دانشجویان توسط فاکتورهای موثر و احساسی، محققان انگیزه این را پیدا کردند تا برای بهبود عملکرد و انگیزه دانشجویان با استفاده از یک ماشین یادگیری محتویات متن‌های مربوط به این دانشجویان را بررسی کنند. در یک مطالعه موردی از 12 مقاله که توسط دانشجویی نوشته شده بود به این نتیجه رسیدند که احساسات دانشجو در هنگام نوشتن در متنها بازخورد پیدا میکند. روش کار به این صورت بود که چهار احساس اصلی انسانها را ترس، شادی، ناراحتی و عصبانیت را در کنار دو رویکرد در نظر گرفتند. اولین رویکرد استفاده از یک فرهنگ لغت مربوط به کلماتی که در دامنه چهار احساس اصلی بود و رویکرد دوم کشف کلمات در این دامنه است. در این مقاله از یک

ماشین یادگیری به نام CoMoLE² استفاده شده که از دو گزینه زیر حمایت میکند: ابتدا با استفاده از محتوای متن مربوط به آنها تشخیص فعالیت‌های شخصی یا مشترکی که دارند، دوم پیشنهاد به آنها بر اساس گزینه اول. (Rodriguez, 2012)

در سال 2014 شبکه اجتماعی توییتر با استفاده از یک گراف تجزیه و تحلیل شد تا به این سوال که آیا توییتر شبکه اجتماعی است یا شبکه اطلاعاتی پاسخ داده شود. در این مقاله ویژگی‌های توپولوژیکی گراف توییتر مثل توزیع درجه، مولفه‌های مرتبط، کوتاهترین طول مسیر، ضرایب خوشه‌بندی و درجه وابستگی تحلیل داده شد. در هر زمان که گراف متناظر دچار برخوردگی شده بود رفتار توییتر بیشتر شبیه شبکه اطلاعاتی بود. هر یک از این ویژگی‌ها با داده‌های به دست آمده از شبکه‌های اجتماعی دیگر مقایسه شد و در نهایت به آمار موثقی به عنوان مرجع در جامعه دست پیدا کرد. در پایان مقاله به این نتیجه رسیدیم که توییتر با آنکه شبیه یک شبکه اجتماعی است ولی بیشتر شبیه شبکه اجتماعی رفتار میکند و حاوی داده‌های مفیدی است که میتوان به عنوان یک منبع آماری از جامعه مورد استفاده قرار گیرد. (Myers, Sharma, Gupta, & Lin, 2014)

در سال 2015 به بررسی شبکه‌های اجتماعی مبتنی بر مکان پرداخته‌اند. این شبکه‌های اجتماعی مثل Facebook، FourSquare و Places به وسیله GPS مکان کاربران را ردیابی میکنند. ولی در این مقاله با استفاده از رصد فعالیت کاربران و محتوای مکالمات، مکانهایی که امکان بودن کاربران را در آن لحظه میدهد را فراهم میکنند. مثالی از مزیت این روش، سیستمهای پیشنهاددهنده هستند که مثلاً اگر کاربری به شهر دیگری مسافرت کرده باشد احتمالاً مکالمات و محاوراتی در آن روزها انجام میدهد که حاوی اطلاعاتی راجع به مکانش میباشد مثلاً از موزه‌های خاص یا مکانهایی خاص با دوستانش صحبت کند یا پستی در اکانت خود قرار دهد. با داده‌کاوی این اطلاعات میتوان حدس زد که به چه شهری مسافرت کرده است و مثلاً سیستم پیشنهاددهنده میتواند اطلاعات مربوط به آن منطقه مثل آدرس موزه‌ها، رستورانها... را به کاربر بدهد. (Gao & Liu, 2015)

در مقاله‌ای در سال 2015 نویسندگان متغیرهایی به چهار نقش کلیدی کاربران اختصاص دادند. شخصیت، نگرش، خود اثربخشی و انگیزش این چهار متغیر کلیدی شدند. بعد از آن 5 فرضیه مطرح کردند: فرضیه 1 به این صورت بود که اجتماعی بودن افراد میتواند توسط فعالیت کاربران در شبکه‌های اجتماعی مشخص شود. فرضیه 2 این بود که خجالتی بودن افراد با میزان عضویتشان در شبکه‌های اجتماعی مرتبط نیست. فرضیه 3 این بود که میزان واکنشی که افراد خجالتی در شبکه‌های اجتماعی نشان میدهند نسبت به دنیای واقعی بیشتر است. فرضیه 4 این بود که افراد تنبل انگیزه بیشتری برای عضویت و فعالیت در شبکه‌های اجتماعی دارند و فرضیه 5 این است که نگرش افراد در تمایل به استفاده از اینترنت و شبکه‌های اجتماعی تاثیر میگذارد. در نهایت به این نتیجه رسیدند که شخصیت و سبک زندگی بر استفاده از اینترنت و شبکه‌های اجتماعی تاثیر میگذارد. (Wang, 2015)

در مقاله‌ای در سال 2015 محققان با استفاده از فعالیتهای روتینی (فعالیهایی که کاربران در سیکل های زمانی در مکان هایی مشخص مرتباً تکرار میکنند) که کاربران شبکه‌های مجازی انجام میدهند تشابه بین آنها را خارج کردند. در این مقاله فعالیت کاربران دو گزینه داشت: اول فعالیتهای بلند مدت و دوم فعالیت های کوتاه مدت. در مقالات قبلی بر روی داده هایی که از فعالیت های کوتاه مدت کاربران در شبکه های اجتماعی به دست آمده بود کاوش انجام داده بودند اما در این مقاله بر روی فعالیت های بلند مدت تحقیق شده است. با استفاده از فعالیتهای روتینی که هر کاربر هرروزه انجام میدهد به فعالیت هایی

². Context-based adaptive Mobile learning Environments

که در دراز مدت انجام میدهند پی برده میشود و با رصد این فعالیت های روتین تشابه بین کاربران تخمین زده میشود. مراحل انجام کار به صورت زیر است:

در مرحله اول مسیری که یک کاربر مثلاً یک دانشجو در طی روز میپیماید. دوم کاوش این مسیر برای پی بردن به فعالیتهای روتینی که این دانشجو انجام میدهد (اما باید این را در نظر داشته باشیم که یک دانشجو هرروز کارهای مشابه انجام نمیدهد). سوم انتخاب معیاری برای استخراج تشابه از فعالیتهای روتین. در نتیجه این مراحل به: 1. مدلی برای فعالیتهای روزانه احتیاج است که بر اساس مکانهایی که از رصد کاربران به دست میاید. 2. یک الگوریتم برای خوشه بندی فعالیتهای روزانه یک کاربر برای فعالیتهای بلند مدتی که مرتباً تکرار میکنند. 3. پیدا کردن یک معیار تشابه کاربران با استفاده از مفهوم فعالیتهای روتین. 4. انجام یک آزمایش جامع و نشان دادن نتایج برای ایجاد یک روشی که بتوان فعالیتهای بلند مدت و شباهت را از آن استنباط کرد. بعد از آن چهار تعریف داده میشود: 1. نقطه و مسیر GPS 2. نقطه ملاقات و مکان منبع 3. فعالیت در طی یک روز 4. فعالیت روتین. برای داده کاوی فعالیت های روتین از یک الگوریتم خوشه بندی مبتنی بر زمان استفاده شده است به این صورت که الگوریتم خوشه بندی تمام فعالیتهای یک روز را به K گروه تقسیم میکند به صورتی که فعالیتهای مشابه در یک گروه قرار میگیرند. با توجه به این که میزان فعالیت کاربران در روزهای مختلف متفاوت است از یک الگوریتم دیگر برای خوشه بندی فعالیتهای یک روز برای داده کاوی فعالیتهای روتین استفاده میکنند. در مرحله بعد با ارائه یک تعریف دیگر با عنوان مکانی که احتمال وقوع یک فعالیت روتین در آن در یک دامنه خاصی قرار گیرد. در مرحله مدلسازی یک ماتریس برای تخمین ماکزیمم شباهت کاربران به یکدیگر با استفاده از تعریف ششم مقاله که جفت دوتایی هایی از فعالیت های روتینی که در دو مکان انجام میشود را میدهد. در نهایت الگوریتم سوم مقاله ارائه میشود که جفت مکانهایی که در تعریف ششم آورده شده کشف میکند. در این مقاله مسیری که یک کاربر در طی یک روز میپیماید را رصد میکند و به جفت مکان تقسیم میکند و با استفاده از الگوریتم سوم احتمال فعالیت روتینی که در این مکانها روی میدهد را برآورد میکند. نهایتاً روش پیشنهادی بر روی دو دیتاست واقعی و مجازی آزمایش شد، این روش نسبت به روش معیار تشابه MSTP که در مقاله معرفی شده نتایج بهتری داشت و به نرخ درستی 88.4٪ رسید (Lv, Chen, & Chen, 2013).

خوشه‌بندی گراف

به فرایند گروه بندی کردن راس‌های گراف به داخل خوشه‌ها به طوری که در داخل هر خوشه تراکم یال‌ها زیاد و بین خوشه‌ها تراکم نسبتاً کم باشد خوشه‌بندی کردن گراف گفته میشود. هدف ما استخراج اجتماعات کاربران است که در مقاله (ساعده، 1393) در مورد آن تحقیق شده است: در این مقاله شبکه اجتماعی با استفاده از یک گراف مجسم سازی شده است که کاربران نودها و ارتباطات بین آنها همان یال‌ها هستند، برای تک‌تک نودها سه معیار مرکزیت میانگی، مرکزیت نزدیکی و مرکزیت ویژه بردار را محاسبه میکند و تمامی این مقادیر را نرمال میکند و در نهایت برای هر گره یک امتیاز محاسبه میکند و آن را مبنایی برای خوشه‌بندی گرافها قرار میدهد. با توجه به اینکه این روش برای نودهای زیاد روش کارآمدی نمیباشد ما سعی خواهیم کرد از الگوریتم خانواده GBI³ استفاده کنیم که با توجه به حجم عظیم داده در محیط ابری نیز قابل استفاده باشد.

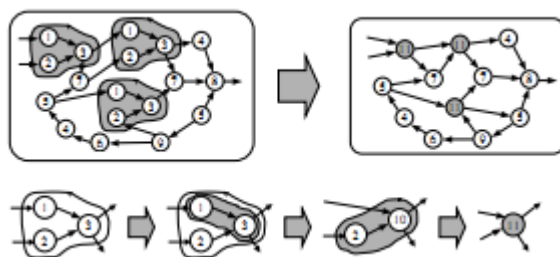
الگوریتم‌های خانواده GBI

الگوریتم GBI نوعی از ماشین‌های یادگیری گراف محور هستند که میتواند گام‌به‌گام زیرگرافهایی از ساختار داده گرافها را به صورت ترکیبی با استفاده از گسترش جفت نودها به دست بیاورد. گسترش گام‌به‌گام جفت نودها عملیاتی ضروری در الگوریتم GBI است که بصورت بازگشتی نودهای جدیدی را توسط انتخاب دو نود مجاور طبق یک معیار خاص مثل فرکانس و

³. Graph_Based Induction

جایگزینی همه وقایعشان در گرافها در یک سطح تخصیص میدهد. خروجی این الگوریتم چندین خوشه از ادها در قالب زیرگراف میباشد. (شکل 1) نوعی از این الگوریتم با نام $B-GBI^4$ ، با انتخاب اختیاری یک زیرگراف از ساختار داده گراف برپایه دستوالعمل گسترش جفت نود، دو نود مجاور را تکه تکه میکند ولی این روش زیرگرافهای مفید را در اختیار ما قرار نمیدهد. اما در مقابل الگوریتم داده کاوی گراف $CI-GBI^5$ که بعد از آن معرفی شد از عمل تکه تکه کردن استفاده نمیکند بلکه در عوض یک جفت نود به عنوان نودهای کاذب⁶ در نظر میگیرد و یک سطح جدید به آن تخصیص میدهد.

این الگوریتم از قوانین انجمنی در داده کاوی برای کشف زیرگرافها استفاده میکند و از بزرگترین گروه شروع می کند و تا زمانی که کمینه ضریب پشتیبان برقرار شود، به صورت بازگشتی به گروههای کوچکتر تقسیم می شود. روش کار به این صورت است که برای به دست آوردن زیر گرافهای با تکرار بالا از بزرگترین گراف ممکن شروع می کند و تا زمان اجرای شرط به- صورت بازگشتی، زیر مجموعه زیر گرافی که دارای شرط نباشد را بررسی می کند. این کار تا زمانی که دو نود در هر زیر گراف باقی مانده باشد، ادامه پیدا می کند. این الگوریتم دارای ۲ شرط است. یکی بر اساس اندازه فراوانی برای قطعه کردن و دیگری پیدا کردن زیرگرافهای متمایز بعد از قطعه بندی. اگرچه دارای پیچیدگی زمانی بیشتری نسبت به روش قبلی میباشد اما در استخراج زیرگرافهای مشابه از یک گراف بزرگ بسیار کارآمدتر میباشد.



شکل 1- الگوریتم GBI

روش پیشنهادی

کاربران شبکه های اجتماعی نودها و ارتباط بین آنها نودهای گراف ما خواهند بود، الگوریتم $CI-GBI$ بر آن اعمال خواهیم کرد تا زیرگرافهایی با تکرار آستانه θ را خارج کنیم. الگوریتم $CI-GBI$ به این صورت میباشد:

ورودی: یک پایگاه داده گراف G ، یک شعاع عرضی b ، بیشینه تعداد سطوح L ، یک معیار برای رتبه بندی جفتها به یک تکه کاذب C ، یک شرط ضروری که بگوید زیرگرافها θ را باید برآورده سازند.

خروجی: یک مجموعه از زیرگرافهای معمولی S ;

گام اول: گسترش همه جفت های شامل دو نود متصل در G ، ثبت نام موقعیتشان با استفاده از شناسه نود (از سطح دوم همه جفت های شامل دو نود متصل با حداقل یک نود که نود کاذب جدید باشد).

گام دوم: شمردن تکرار جفت های بسط داده شده و از بین بردن آنهایی که شرط ضروری θ را برآورده نمیکنند.

⁴ . Beam-wise Gaph Based

⁵ . Chunking Less Graph-Based Induction

⁶ . Pseudo-node

گام سوم: انتخاب بهترین جفتهای b طبق C از بین جفتهایی که در گام دوم باقی مانده است (از سطح دوم از بین جفتهای انتخاب نشده در سطوح قبلی د جفتهای بسط داده شده جدید). هریک از جفتهای b انتخاب شده به عنوان یک نود کاذب ثبت میشود. اگر یکی یا دوتا از تودههای انتخاب شده یک جفت اصلی نباشند اما نود کاذب باشند، آنها به زیرگرافهایی که قبلاً ثبت شده‌اند بازگردانده میشوند.

گام چهارم: تخصیص یک سطح جدید به هر جفت انتخاب شده در گام سوم، اما گراف دوباره نوشته نمیشود. برگشت به گام اول.

این گام‌ها را به تعداد N (تعداد سطوح) تکرار میکنیم. خروجی الگوریتم GBI مجموعه‌ای از الگوهای معمولی مرتبی هستند که در موقعیت هر پیشامد در هر تراکنش پایگاه داده در کنار یکدیگر قرار میگیرند و همچنین تعداد این پیشامدها در هر تراکنش را میدهد.

بحث و نتیجه‌گیری

در این مقاله ما با استفاده از الگوریتم CI-GBI که نمونه‌ای از الگوریتم خانواده GBI میباشد بر روی گراف مجسم‌سازی شده‌ی کاربران شبکه اجتماعی، خوشه‌بندی انجام دادیم تا کاربرانی که دارای تشابهات رفتاری در این شبکه‌ها هستند را کشف کنیم.

با توجه به حجم عظیم داده‌های حاصل از شبکه‌های اجتماعی، استفاده از الگوریتمهای سنتی ما را به بهینه‌ترین جوابها نمیرساند. استفاده از تکنیک‌های داده‌کاوی ابرکامپیوتورها مانند مدل MapReduce در چهارچوب Hadoop به منظور کاهش داده‌ها برای رسیدن به داده‌های معقول، به عنوان کارهای آینده پیشنهاد داده میشود.

منابع

ساعد، نوشین و مهدی صادق زاده، ۱۳۹۳، ارائه الگوریتمی جهت خوشه بندی گراف شبکه های اجتماعی مبتنی بر مرکزیت گره ها، اولین همایش ملی مهندسی برق و کامپیوتر در شمال کشور، بندر انزلی، موسسه آموزش عالی موج
کجباف، محسن و محمد صادق رحمانیان، ۱۳۹۱، پیشنهاد یک الگوریتم برای خوشه بندی گراف شبکه های اجتماعی مبتنی بر گره های مرکزیت، یازدهمین کنفرانس سراسری سیستم های هوشمند، انجمن سیستمهای هوشمند ایران، دانشگاه خوارزمی

- Di Marco, A., & Navigli, R. (2012). Clustering and Diversifying Web Search Results with Graph-Based Word Sense Induction. *Computational Linguistics*, 39(3), 709-754. doi: 10.1162/COLL_a_00148
- Gao, H., & Liu, H. (2015). Mining Human Mobility in Location-Based Social Networks. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 7(2), 1-115. doi: 10.2200/S00630ED1V01Y201502DMK011
- Lv, M., Chen, L., & Chen, G. (2013). Mining user similarity based on routine activities. *Information Sciences*, 236, 17-32. doi: <http://dx.doi.org/10.1016/j.ins.2013.02.050>
- Matsuda, T., Motoda, H., & Washio, T. (2002). Graph-based induction and its applications. *Advanced Engineering Informatics*, 16(2), 135-143. doi: [http://dx.doi.org/10.1016/S1474-0346\(02\)00005-8](http://dx.doi.org/10.1016/S1474-0346(02)00005-8)
- Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). *Information network or social network?: the structure of the twitter follow graph*. Paper presented at the Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea.
- Nguyen, P., Ohara, K., Motoda, H., & Washio, T. (2005). CI-GBI: A Novel Approach for Extracting Typical Patterns from Graph-Structured Data. In T. Ho, D. Cheung & H. Liu (Eds.), *Advances in Knowledge Discovery and Data Mining* (Vol. 3518, pp. 639-649): Springer Berlin Heidelberg.
- Rodriguez, P. E. P. S., Univ. Autonoma de Madrid, Madrid, Spain ; Ortigosa, A. ; Carro, R.M. (2012). Extracting Emotions from Texts in E-Learning Environments. *ON Sixth International Conference Complex, Intelligent and Software Intensive Systems (CISIS)*.
- Wang, J.-L., Jackson, Linda A, Wang, Hai-Zhen, Gaskin, James. (2015). Predicting Social Networking Site (SNS) use: Personality, attitudes, motivation and Internet self-efficacy. *Personality and Individual Differences*, 80, 119-124. doi: <http://dx.doi.org/10.1016/j.paid.2015.02.016>

Board of SID
INTERNATIONAL CONFERENCE ON
**RESEARCH IN SCIENCE
AND TECHNOLOGY**

BERLIN . GERMANY 9 July, 2016



Visit  Berlin

سومین کنفرانس بین المللی

ی پژوهش در علوم و تکنولوژی

برلین - آلمان ۱۹ تیرماه ۱۳۹۵

SID



سرویس های ویژه



سرویس ترجمه تخصصی



کارگاه های آموزشی



بلاگ مرکز اطلاعات علمی



سامانه ویراستاری STES



فیلم های آموزشی

کارگاه های آموزشی مرکز اطلاعات علمی جهاد دانشگاهی

توجه: بررسی مقاله ای متون (مقدماتی)

کارگاه آنلاین
بررسی مقابله ای متون (مقدماتی)

PROPOSAL
پروپوزال

توجه: پروپوزال نویسی و پایان نامه نویسی

کارگاه آنلاین
پروپوزال نویسی و پایان نامه نویسی

ISI
Scopus

توجه: آشنایی با پایگاه های اطلاعات علمی بین المللی و ترند های جستجو

کارگاه آنلاین آشنایی با پایگاه های اطلاعات علمی بین المللی و ترند های جستجو