

ارزیابی الگوریتم‌های نخبه‌یابی در شبکه‌های اجتماعی به منظور ارائه الگوریتم برتر

احمدآقا کاردان^۱، بهنام بزرگی^۲

^۱استادیار، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران
aakardan@aut.ac.ir

^۲دانشجوی کارشناسی ارشد، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران
behnam.bozorgi71@gmail.com

چکیده

امروزه با گسترش اینترنت و شبکه‌های اجتماعی، انجمن‌های پرسش و پاسخ محیط مناسبی است تا کاربران سوالات خود را درباره مطالب مختلف مطرح کرده و یا به سوالات دیگران پاسخ دهند. این انجمن‌ها به منظور اشتراک دانش طراحی شده تا سطح دانش افراد افزایش یابد، پس ضروریست تا مکانیزمی برای سنجش سطح دانش کاربران و یا یافتن نخبگان وجود داشته باشد. اینجاست که نیاز به وجود الگوریتم‌هایی برای پیدا کردن نخبگان در شبکه‌های اجتماعی و یا هر محیط اشتراک دانش مانند انجمن‌های پرسش و پاسخ احساس می‌شود.

روش‌های مختلفی مبتنی بر تحلیل محتوا و تحلیل پیوند برای پیدا کردن نخبگان در شبکه‌های اجتماعی وجود دارد و همه پژوهش‌ها الگوریتم مدنظر خود را ارائه کرده‌اند. ما در این مقاله سعی می‌کنیم چهار الگوریتم برتر مبتنی بر تحلیل پیوندها را با استفاده از مجموعه داده خود به چالش کشیده و با تحلیل نتایج بدست آمده، آنها را با یکدیگر مقایسه می‌کنیم. بر اساس مطالعات و آزمون‌های انجام شده نتیجه می‌گیریم که الگوریتم امتیاز Z دارای عملکرد بهتری نسبت به سایر الگوریتم‌ها داشته است. دست‌آورد این مقاله برای کسانی که می‌خواهند در حوزه یافتن خبرگان فعالیت کنند، الگوریتم مناسب برای یافتن خبرگان در انجمن‌ها است. در ادامه پس از ارائه مقدمات و کارهای مرتبط، مجموعه داده را معرفی می‌کنیم، توضیحاتی درباره الگوریتم‌های استفاده شده در این پژوهش ارائه می‌دهیم و در انتها نتایج را اعلام می‌کنیم.

کلمات کلیدی

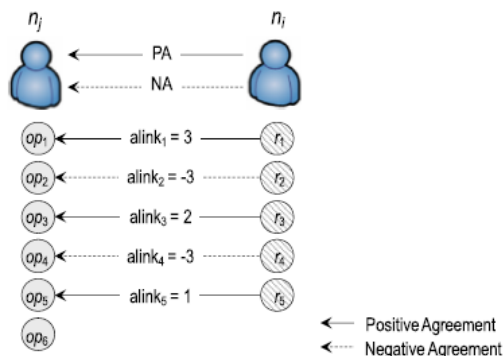
پیدا کردن نخبگان، تحلیل شبکه اجتماعی، انجمن پرسش و پاسخ

تخصصی. همه انسان‌ها به دنبال راه‌حل برای آنها می‌گردند و طبیعتاً برای اینکه هرچه سریعتر به پاسخ سوال و مشکل خود برسند، سعی می‌کنند به دنبال فرد خبره در حوزه مشکل یا سوال خود بگردند. در عصر امروز افراد اولین کاری که می‌کنند این است که کامپیوتر خود را روشن کرده و به دنبال

۱- مقدمه

انسان‌ها در طول زندگی خود به مشکلات زیادی برخورد می‌کنند، چه مشکلات روزمره، مشکلات منحصربه‌فرد، مشکلات تخصصی و غیر

مثبت یا منفی خواهد بود که نمونه ارتباط ۲ کاربر در این پژوهش را در شکل (۱) مشاهده می‌کنیم.



شکل (۱): نمونه ارتباط ۲ کاربر [۱]

سپس با محاسبه امتیاز ارتباط بین آنها و بهره‌گیری از الگوریتم رتبه-صفحه ویرایش شده خود، سطح دانش کاربران را محاسبه کردند. در پژوهش [۲] ابتدا از یک روش صریح معنایی استفاده شده است. بدین صورت که افراد در مجموعه داده انتخابی (Sina Microlab) را به عنوان ورودی سیستم دریافت کرده و آن را با استفاده از الگوریتم TFIDF به بردار تبدیل می‌کند. سپس برای محاسبه میزان مشابهت بین افراد با یکدیگر، از میزان نزدیکی کسینوسی استفاده شده است. سپس برای پیدا کردن نخبگان با استفاده از تحلیل صریح معنایی از تعریفی در شبکه بیضی استفاده شده است. ایده کلی بدین صورت است که سیستم میزان احتمال نخبگی فرد x_i در موضوع Q را تخمین بزند $P(x_i|Q)$. روش‌های متداول مانند مدل‌های زمانی از تعریف شبکه بیضی برای پیدا کردن این احتمال استفاده می‌کنند ولی در این پژوهش، مقدار $P(x_i|Q)$ برابر میزان مشابهت نخبگی فرد در موضوع Q در نظر گرفته شده است. سپس با استفاده از این مشابهت و امتیاز حاصل از تحلیل روابط بین افراد، میزان نخبگی نهایی فرد را محاسبه می‌کنند.

پژوهش [۳] با تغییر در ساختار الگوریتم رتبه-صفحه به منظور قابل استفاده شدن آن در شبکه اجتماعی به پیدا کردن نخبگان در شبکه اجتماعی FriendFeed پرداخته است. در این مقاله ابتدا داده‌ها با استفاده از یک خزنده وب، استخراج شده، در یک پایگاه داده ذخیره شده و سپس داده‌های مورد نیاز جهت پیدا کردن نخبگان، از درون این پایگاه داده استخراج شده و با اعمال الگوریتم معرفی شده، نخبگان مشخص شده‌اند.

در [۴] روشی به منظور یافتن خبرگان با استفاده از محتوای انجمن‌های بر خط ارائه شده است. در این پژوهش سه پارامتر جهت تخمین زدن سطح دانش کاربران بر اساس مدل دامنه دانش داده شده پیشنهاد گردیده است. این سه پارامتر عبارتند از:

- نسبت تعداد مفاهیم کلیدی ذکر شده توسط یک کاربر در پیام-هایش به کل مفاهیم کلیدی موجود در آن بخش از دامنه دانش.
- تعداد مؤلفه‌های همبند ایجاد شده در مدل کاربر بر اساس مفاهیم کلیدی به کار رفته.
- تعداد مفاهیمی که توسط کاربر در یک پیام ذکر شده. این پارامتر به ازای هر کاربر ممکن است متفاوت باشد.

راه‌حل مشکل خود در دنیای مجازی و اینترنت می‌گردند، از دوستان خود در شبکه‌های اجتماعی سوال می‌کنند. ولی برخی افراد که در پیدا کردن راه‌حل مٌصبر هستند، سری به انجمن‌های مختلف می‌زنند و در آنجا با به وجود آوردن یک سرفصل، مشکل خود را مطرح می‌کنند. از جمله انواع انجمن‌ها، انجمن-های پرسش و پاسخ مکان بسیار مناسبی است تا افراد مشکلات و سوالات خود را با دیگران در میان بگذارند و از دیگران مشورت بگیرند. حال این سوال مطرح می‌شود که اگر فردی مشکل خود را در یک انجمن پرسش و پاسخ مطرح کرد و بعد از دریافت چندین پاسخ، از کجا بفهمد که جواب کدام فرد درست است؟ راه حل این مشکل این است که افراد باید متخصصین را در این انجمن بشناسند و فقط به جواب‌های آنها اعتماد کنند. در سال‌های اخیر برای رفع نیاز این افراد توجه زیادی به پیدا کردن نخبگان در حوزه‌های مختلف مانند انجمن‌ها و شبکه‌های اجتماعی شده است. پیدا کردن نخبگان در واقع به این معنی است که فرایندی را انجام دهیم که در انتهای آن کاربران پاسخ-دهنده‌ای که بیشترین امتیاز را از نظر سطح دانش می‌گیرند را پیدا کرده و به عنوان خبره در آن حوزه و موضوع مشخص معرفی کنیم.

امروزه انجمن‌های پرسش و پاسخ به دو دسته کلی تخصصی و غیر تخصصی تقسیم می‌شوند. انجمن‌های پرسش و پاسخ تخصصی انجمن‌هایی هستند که در آنها به پرسش و پاسخ در یک موضوع و حوزه خاص پرداخته می‌شود که از جمله آنها می‌توان به انجمن تخصصی جاوا که برای پرسش و پاسخ در حوزه برنامه‌نویسی جاوا طراحی شده است، اشاره کرد. انجمن‌های پرسش و پاسخ غیر تخصصی مانند یاهو انسر، دارای تعداد بسیار زیاد موضوع برای پرسش و پاسخ می‌باشد.

ادامه این مقاله در بخش‌های زیر ارائه می‌گردد:

در بخش ۲ به ارائه پژوهش‌های مرتبط با پیدا کردن نخبگان و در بخش ۳ به معرفی مجموعه داده استفاده شده، می‌پردازیم، در بخش ۴ به معرفی الگوریتم-های پیدا کردن نخبگان پرداخته شده است، بخش ۵ تحلیل و اجرای الگوریتم‌ها مورد بررسی قرار گرفته است. بخش ۶ شامل ارزیابی الگوریتم‌های معرفی شده، است. و در بخش ۷ نتیجه‌گیری حاصل از این مقاله ارائه خواهد شد.

۲- کارهای مرتبط

در سال‌های اخیر با گسترش فناوری، حجم زیادی از اطلاعات در اینترنت منتشر شده که این اطلاعات می‌تواند به صورت مفید در راستای اهداف مختلف استفاده شود. از جمله آنها می‌توان به پژوهش [۱] اشاره کرد که در آن پیدا کردن نخبگان به منظور اشتراک دانش استفاده شده است. در این پژوهش ابتدا به بررسی انواع روش‌های پیدا کردن نخبگان پرداخته شده است. پیدا کردن نخبگان شامل ۲ دسته می‌شود، پیدا کردن نخبگان برپایه تحلیل ارتباط بین کاربران مانند روش‌های برپایه الگوریتم HITS و الگوریتم رتبه-صفحه و روش برپایه تحلیل محتوا. پژوهش‌گران این مقاله با بیان اینکه روش تحلیل ارتباط موفق‌تر از روش برپایه محتوا بوده است، ایراداتی نیز به آن گرفته‌اند. آنان به این نتیجه رسیدند که الگوریتم‌هایی مانند HITS این ایراد را دارند که همیشه سعی در افزایش امتیاز کاربران و سطح دانش آنها دارند. البته این ایراد زمانی است که مجموعه داده تحقیقی ذاتاً دارای مکانیزم امتیازدهی منفی به کاربران نباشد. بنابراین ارتباط بین کاربران در شبکه ارتباطی آنها دارای امتیاز

در پژوهش [۸] برخلاف پژوهش‌های قبلی از روشی غیر از روش کلی تحلیل اجتماعی روابط و تحلیل برپایه محتوا، استفاده شده است. این پژوهش در ابتدا به معرفی روش‌های متداول پیدا کردن نخبگان می‌پردازد، سپس مسئله پیدا کردن نخبگان از طریق پیشبینی کردن مقادیر خالی در ماتریس روابط کاربران مدنظر قرار می‌گیرد. در حقیقت این الگوریتم یک الگوریتم بر پایه محاسبات ریاضی در حوزه ماتریس‌ها می‌باشد. فرایند پرکردن ماتریس در حوزه‌های مختلفی مانند دید کامپیوتری [۹] و فیلترینگ همکارانه [۱۰] به کار گرفته شده است که بیشتر به آنها از دید پر کردن مقادیر خالی در ماتریس، نگاه می‌شود. خانه‌های بدون مقدار را با استفاده از مقادیر ماتریس روابط می‌توان بدست آورد ولی طبق بیان پژوهش [۸] مطابق با تابع بدست آورنده رتبه ماتریس، این یک مسئله NP-Hard به حساب می‌آید. رتبه ماتریس مجموع سطرهای غیر صفر یک ماتریس می‌باشد. به همین دلیل این مسئله به یک مسئله بهینه سازی با استفاده از نرم هسته‌ای تبدیل می‌شود. که با استفاده از این نرم، مقادیر ماتریس روابط بدست آمده و میزان نخبگی افراد تعیین می‌شود.

مسئله دیگر در حوزه پیدا کردن نخبگان تعیین تعداد افراد دارای سطح نخبگی بالا است. بدین معنی که اگر قرار باشد k فرد را به عنوان افراد نخبه معرفی کنیم، این k چگونه و بر چه اساسی انتخاب شود [۱۱ و ۱۲ و ۱۳].

ما در این پژوهش قصد داریم چهار الگوریتم پیدا کردن نخبگان را با اعمال روی یک مجموعه داده یکسان، با یکدیگر مقایسه کنیم. بدین منظور در بخش سوم به معرفی مجموعه داده پرداخته و بعد به معرفی الگوریتم‌های متداول می‌پردازیم.

۳- معرفی مجموعه داده

در این پژوهش به یک مجموعه داده مربوط به انجمن پرسش و پاسخ نیاز داریم. بدین منظور از مجموعه داده Stack Exchange، زیرشاخه Stack Overflow آن استفاده شده است. Stack Exchange شبکه‌ای از چندین انجمن پرسش و پاسخ که توسط متخصصان و علاقه‌مندان که به یک موضوع خاص علاقه دارند، ساخته شده است. این انجمن‌ها خزانه‌ای از سوالات و پاسخ‌های با کیفیت مرتبط با موضوع خاص است که می‌توان از بین افراد شرکت کننده بر اساس فعالیت‌های آنها، نخبگان را شناسایی کنیم.

برای استخراج داده از این وبگاه، از واسط کاربری استخراج داده آن وبگاه استفاده شده است. اطلاعات مربوط به افراد، پرسش و پاسخ‌ها و ... در قالب جداول گوناگون پایگاه داده تعبیه شده است. که در شکل (۲) شمای آن قابل مشاهده می‌باشد.

همانطور که در بخش‌های قبل بیان شد، برای تشخیص نخبگی یک فرد باید موضوعی که در آن نخبگی افراد را مشخص می‌کنیم به صورت مشخص معلوم باشد و کلی نباشد. بدین منظور با استفاده از ویژگی برچسب و با زدن دستور و استخراج داده دریافتیم که موضوع سی شارپ^۳ جزء ۳ موضوع مطرح مورد بحث در وبگاه Stack Overflow با بیشترین میزان پرسش و پاسخ می‌باشد. این دست‌آورد نیز توسط این وبگاه تایید شده است. بنابراین، پرسش و پاسخ سال ۲۰۱۴ میلادی، مربوط به موضوع سی شارپ این وبگاه را استخراج کردیم. پس از استخراج و انجام پیش‌پردازش‌هایی مانند حذف

روش پیشنهادی در این پژوهش دارای یک نقطه ضعف بود و آن این است که مبتنی بر موضوع نمی‌باشد. برای مثال در انجمنی درباره شبکه کامپیوتری ممکن است یک فرد درباره حوزه شبکه در میکروسافت بسیار نخبه و با دانش سطح بالا باشد ولی همان فرد در شبکه در حوزه سیسکو دارای دانش پایینی باشد و یا بالعکس. پس این روش قادر است دانش کاربر را به صورت کلی و نه در یک موضوع خاص تعیین کند.

در پژوهش [۵] سه پارامتر برای پیدا کردن نخبگان در یک موضوع خاص و یا سوال خاص در انجمن‌های پرسش و پاسخ یاهو انسر، ارائه شده است که این سه پارامتر شامل نمایه^۴ دانش کاربر، شهرت کاربر و تحلیل ارتباطات است.

در این پژوهش، داده‌های مربوط به گذشته پرسش و پاسخ هر یک از کاربران به وسیله روش‌های داده کاوی بدست آمده و از آنها برای ساختن نمایه دانش کاربران استفاده می‌شود. این نمایه، نشان دهنده موضوعات دانش کاربر خواهد بود. این اطلاعات و همچنین اطلاعات تکمیلی دیگری همانند موضوع سوالات، زمان ارسال و ... از مجموعه داده دریافت شده به فضای برداری برده می‌شود، یعنی هر فرد یک بردار از ویژگی‌های مربوط به دانش خود دارد. پس از ساختن نمایه دانش کاربران، به کاربران امتیاز داده می‌شود. این امتیاز بسته به جایی که این موضوع بیشتر متمرکز شده و مطرح گردیده (در سرتیتر سوال، توضیحات سوال و پاسخ) متفاوت می‌باشد، این امتیاز از طریق میزان مشابهت و نزدیکی دانش فرد با سوالاتی که او پاسخ داده، محاسبه می‌شود.

از سابقه پرسش و پاسخ افراد در حوزه‌های مختلف نیز برای بدست آوردن امتیاز شهرت کاربر استفاده می‌شود. این اطلاعات شامل تعداد پاسخ‌ها و تعداد دفعاتی که به عنوان بهترین پاسخ‌دهنده انتخاب شده است نیز می‌شود. در تحلیل ارتباطات، هر کاربر بعد از رسم شبکه کاربران بر اساس ارتباط بین آنها که در انجمن‌های پرسش و پاسخ، سوال و پاسخ می‌باشد، رابطه آنها بر اساس الگوریتم‌های مختلف تحلیل می‌شود. از جمله الگوریتم‌ها به رتبه صفحه و HITS و یا حتی شاخص‌های شبکه اجتماعی مانند مرکزیت درجه^۵، نزدیکی و سایر شاخص‌ها می‌توان اشاره کرد. در انتها نیز با جمع امتیازات حاصل از این سه پارامتر، امتیاز نخبگی کاربران محاسبه می‌شود.

پژوهش [۶] نشان می‌دهد که پیدا کردن نخبگان بر پایه موضوع، می‌تواند بر اساس روش‌های نهان مدل کردن موضوعات درباره محتوای سوالات مطرح شده باشد. به عنوان مثال می‌توان میزان نخبه بودن کاربران را بر اساس پاسخ‌ها و نظراتی که به سوالات مختلف در حوزه‌های مختلف می‌دهند، مدل کرد. در این روش موضوعات مورد علاقه کاربر از روی نمایه کاربری او استخراج می‌شود. در این روش حجم مجموعه داده اهمیت زیادی دارد چراکه اگر مجموعه داده بزرگی مدنظر باشد، نمی‌توان از روش‌های بدون ناظر و یا با ناظر ضعیف برای بدست آوردن موضوع مورد علاقه کاربر به صورت خودکار، استفاده کرد.

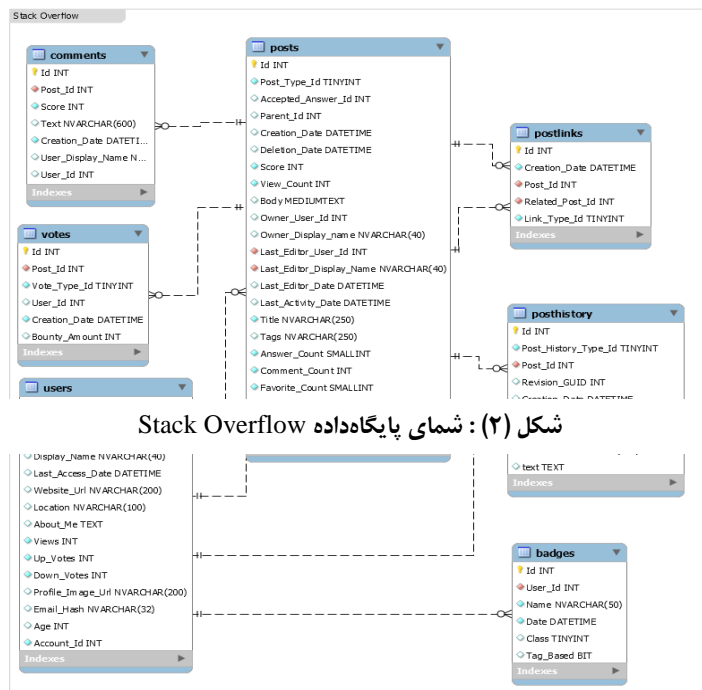
روش CRAR^۶ معرفی شده در [۷]، اعتبار کاربران را بر اساس آنالیز ارتباطات آنان بدست می‌آورد. این کار را بر اساس طبقه بندی که پرسش مورد نظر در آن قرار می‌گیرد، و همچنین بر اساس موضوعات مرتبط بدست آمده از مدل موضوعی، انجام می‌دهد. به بیانی دیگر این الگوریتم از هردو روش برپایه محتوا و تحلیل روابط برای پیدا کردن نخبگان بهره می‌گیرد.

• وقتی که گره‌ای هیچ پیوندی را به گره‌ای دیگر نداشته باشد، آنگاه عمل پرسش انجام می‌شود.
• اگر گره دارای پیوند خروجی به گره‌ای دیگر باشد، آنگاه جست‌وجو کننده عمل پرسش را با احتمال p ($0 < p < 1$) اجرا می‌کند و با احتمال $1 - p$ به دنبال کردن پیوندها ادامه می‌دهد که P یک مقدار ثابت بوده و معمولاً برابر با ۰.۱ در تحقیقات علمی در نظر گرفته می‌شود.
این الگوریتم یک ویژگی منحصر به فرد دارد و آن این است که امتیازی

کاربران بدون شناسه، ۲۳۴.۸۷۷ پاسخ توسط ۸۹.۵۱۹ کاربر در سال ۲۰۱۴ میلادی به پرسش‌های مربوط به سی‌شارپ داده شده است.

۴- الگوریتم‌های پیدا کردن نخبگان

همانطور که در بخش قبل بررسی شد، الگوریتم‌های مختلفی برای پیدا کردن نخبگان در شبکه‌های اجتماعی استفاده شده است که از جمله مطرح‌ترین آنها



که کاربر می‌گیرد نه تنها به تعداد یال‌های ورودی بلکه به تعداد یال‌های ورودی به نودی که به آن کاربر متصل است، بستگی دارد، یعنی امتیاز رتبه افراد به صورت بازگشتی محاسبه می‌شود.

در انجمن‌های پرسش و پاسخ برای استفاده از این الگوریتم می‌بایست کمی تغییرات در این الگوریتم داده شود. مشکل در آنجایی است که اگر یک فرد به بیش از یک سوال فرد دیگر پاسخ دهد، بین آن دو همچنان یک یال وجود خواهد داشت و این الگوریتم همچنان یک پاسخ را می‌بیند (روی هم افتادگی). برای رفع این مشکل تعداد پاسخ‌های هر فرد به فرد دیگر (در صورت وجود) بر روی یال مربوطه قرار دادیم تا در الگوریتم لحاظ شود که هرچه تعداد پاسخ‌های فرد بیشتر باشد امتیاز او بیشتر خواهد بود. رابطه (۱) نحوه محاسبه امتیاز رتبه‌صفحه هر کاربر را نشان می‌دهد که در آن $PR(u)$ امتیاز گره N تعداد گره‌های موجود در شبکه، d پارامتر تعدیل نام دارد که احتمال عبور از یک گره به گره بعدی را نشان می‌دهد و به طبع $1-d$ احتمال پرسش را نشان می‌دهد، B_u مجموعه گره‌هایی است که به گره u متصل هستند و $L(v)$ تعداد یال‌های خروجی از نود v می‌باشد که پس از ویرایش برای استفاده در انجمن‌های پرسش و پاسخ، وزن یال‌های خروجی v با یکدیگر جمع شده تا $L(v)$ بدست آید.

$$PR(u) = 1 - d + d(\sum_{v \in B_u} \frac{PR(v)}{L(v)}) \quad (1)$$

می‌توان به رتبه‌صفحه، HITS، درجه ورودی و امتیاز Z اشاره کرد. در ادامه به معرفی این الگوریتم‌ها پرداخته و نحوه ارائه امتیازدهی به کاربران برای نمایش خبرگی را بررسی می‌کنیم.

۴-۱- الگوریتم رتبه صفحه

این الگوریتم برای اولین بار در پژوهش [۱۴] ارائه شد. هدف از این الگوریتم در ابتدا تعیین اهمیت وبگاه‌های مختلف بر اساس میزان مراجعه به آنها بود ولی بعدها برای کاربردهای مختلف مانند پیدا کردن نخبگان مورد استفاده قرار گرفت. در ادامه به توضیح روش کار آن می‌پردازیم.

الگوریتم رتبه صفحه الگوریتمی برپایه حرکت تصادفی یک فرد بین صفحات اینترنت می‌باشد. بدین صورت که فردی وارد وبگاهی شده و از آن وبگاه روی لینک‌های موجود به صورت تصادفی کلیک کرده و به صفحه بعدی می‌رود و این کار را ادامه می‌دهد و بر اساس میزان صفحه‌های که بیشتر ببیند، آن صفحه اهمیت بیشتری دارد. ولی این احتمال وجود دارد که فرد درون حلقه‌ای از صفحات به تله بیفتد. برای رهایی از این تله، یک احتمال پرسش برای کاربر در نظر گرفته شده است تا در صورت گیر کردن در تله، پرسش را انجام دهد ولی این بدین معنی نیست که فرد فقط زمانی پرسش کند که در تله بیفتد. فرد ممکن است با یک احتمالی در هر شرایطی پرسش کند ولی زمانی که به تله می‌افتد این احتمال به ۱۰۰٪ افزایش می‌یابد. در واقع روش کار این پرسش به صورت زیر است:

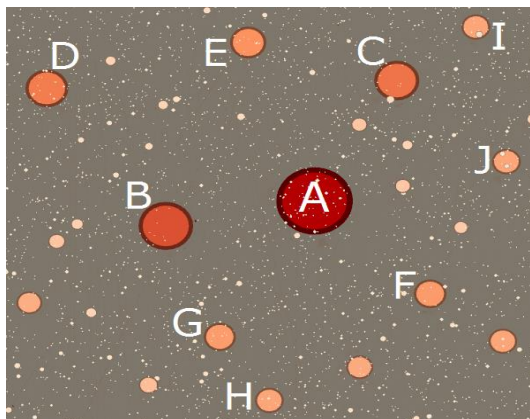
پرسش و پاسخ به تعداد q سوال و به تعداد a پاسخ به سوالات داشته باشد، آنگاه میزان نخبگی او بر اساس رابطه (۴) محاسبه می‌شود.

$$Z = \frac{a-q}{\sqrt{a+q}} \quad (4)$$

در این روش، اگر تعداد سوالات و پاسخ‌ها برابر باشد به طبع مقدار امتیاز Z برابر صفر خواهد شد و در صورت بیشتر بودن a از q مقدار امتیاز Z مثبت، و در غیر این صورت، منفی خواهد شد. اگر در این رابطه از تعداد سوالات و پاسخ‌های فرد استفاده شود، آن را عدد Z^{۱۳} و اگر برابر تعداد کاربرانی که آن کاربر از آنها پرسش کرده و یا به آنها پاسخ داده محاسبه شود، درجه Z^{۱۴} نامیده می‌شود.

۵- تحلیل و اجرا

ابتدا شبکه ارتباط بین کاربران را تعریف می‌کنیم. این شبکه شامل گره و یال-هایی است که در اینجا گره‌ها نمایانگر کاربران و یال‌ها نشان دهنده پاسخی است که یک کاربر به کاربر دیگر داده است. شکل (۳) نشان دهنده بخشی از شبکه ارتباط کاربران و ۱۰ کاربر برتر با تمرکز بر الگوریتم درجه ورودی است. بدین منظور که آن گره‌ای که از همه بزرگ‌تر و پررنگ‌تر است (A)، دارای بیشترین درجه ورودی در شبکه ارتباط کاربران در این پژوهش است. به دلیل تراکم یال‌ها پس‌زمینه شکل تیره می‌باشد. لازم به ذکر است که شبکه از نوع پرسشگر-پاسخ‌دهنده است و یال‌ها از پرسش‌کننده به پاسخ‌دهنده ترسیم شده است.



شکل (۲): بخشی از شبکه ارتباط کاربران

حال بر اساس اطلاعات خروجی از مجموعه داده و الگوریتم‌های مطرح شده، الگوریتم‌ها را بر روی پرسش و پاسخ‌های مطرح شده توسط کاربران اعمال کرده و برای هر یک از کاربران امتیازات مرتبط با هر یک از شاخص‌های معرفی شده را بدست می‌آوریم. ابتدا به بررسی مفهوم هر الگوریتم در این پژوهش می‌پردازیم.

الگوریتم رتبه صفحه ماهیتاً گره‌ها با اهمیت بالا از نظر مراجعه به آن گره را نشان می‌دهد و این مراجعه در انجمن پرسش و پاسخ بدین مفهوم است که آن گره (فرد) در پاسخ‌گویی به سوالات فعالیت زیادی داشته است که معیار نخبگی فرد به حساب می‌آید. شاخص اعتبار در HITS برای یک گره به

۴-۲- الگوریتم HITS

این الگوریتم که یک الگوریتم بر اساس تحلیل پیوندها می‌باشد، دارای دو پارامتر برای امتیازدهی می‌باشد که این دو پارامتر هر دو باید بروز شوند تا به محض اضافه شدن گره (کاربر) و یال (پاسخ) تحت یک سری قوانین بتوان امتیازات را محاسبه کرد. این دو پارامتر قطب^{۱۱} و اعتبار^{۱۲} نام دارند. در این الگوریتم از ماتریس ارتباط بین کاربران نیز استفاده می‌شود که بعد از تکرار بینهایت بار این الگوریتم این ماتریس و پارامترها بروز می‌شوند. برای جلوگیری از گرفتن ارزش‌های واگرا توسط مقادیر این ماتریس، نیاز به نرمال کردن مقادیر وجود دارد که این مکانیزم نیز در این الگوریتم وجود دارد.

الگوریتم HITS برای پیدا کردن نخبگان در انجمن‌های پرسش و پاسخ بدین صورت عمل می‌کند که هر گره دارای ۲ پارامتر قطب و اعتبار می‌باشد و در صورتی که آن گره (کاربر) سوالی را مطرح کند، "امتیاز قطب" او افزایش و هر بار که به سوالی پاسخ دهد، "امتیاز اعتبار" او افزایش می‌یابد. نحوه محاسبه قطب و اعتبار به این صورت است که مطابق رابطه (۲)، اعتبار یک نود برابر مجموع قطب نودهای متصل به آن است یا به بیان دیگر اعتبار کاربر p برابر مجموع قطب کاربرانی (i) است که p به آنها پاسخ داده است و مطابق رابطه (۳)، قطب یک نود برابر مجموع اعتبار نودهایی است که آن نود به آنها متصل است یا به عبارت دیگر قطب کاربر p برابر مجموع اعتباری کاربرانی (i) است که به سوال p پاسخ دادند.

$$auth(p) = \sum_{i=1}^n hub(i) \quad (2)$$

$$hub(p) = \sum_{i=1}^n auth(i) \quad (3)$$

در ابتدا که شبکه ارتباط کاربران تشکیل شد، مقدار قطب و اعتبار هر گره برابر با ۱ در نظر گرفته می‌شود سپس با استفاده از روابط (۱) و (۲) مقادیر این دو پارامتر بروز می‌شوند. در انتها نیز چون پاسخی که افراد می‌دهند در بدست آوردن نخبگی آنها نقش بسزایی دارد، پس امتیاز اعتبار برای رتبه بندی نهایی افراد مدنظر قرار خواهد گرفت.

۴-۳- الگوریتم درجه ورودی^{۱۱}

شاخص درجه از جمله شاخص‌های مرکزیت در شبکه‌های اجتماعی می‌باشد که برابر با تعداد یال‌های ورودی به یک گره است. برای بدست آوردن آن، ماتریس مجاورت شبکه پرسش و پاسخ کاربران را بدست آورده و مجموع درایه‌های یک ستون برابر با امتیاز درجه ورودی گره مربوط به آن ستون خواهد بود. با این تفاوت که در ماتریس مجاورت شبکه پرسش و پاسخ، در صورت ارتباط ۲ گره، وزن یال که نشان دهنده تعداد پاسخ گره (کاربر) پرسش‌گر به گره پاسخ‌دهنده است، قرار می‌گیرد.

۴-۴- الگوریتم امتیاز Z^{۱۳}

در حالیکه پاسخ دادن به سوالات زیاد نشانه این است که یک کاربر دارای تخصص زیادی می‌باشد، پرسش چندین سوال در یک موضوع نشانه کمبود و یا فقدان دانش کاربر در آن موضوع است. بنابراین، در روش امتیاز Z^{۱۳} الگوهای پرسش و پاسخ کاربر با هم ترکیب می‌شوند. اگر کاربر در انجمن

مراجع

زیر نویس ها

- 1 Yahoo Answer
- 2 HITS (Hyperlink Induced Topic Search)
- 3 Profile
- 4 Degree
- 5 CRAR (Category Relevancy based Authority Ranking)
- 6 Tag
- 7 C#
- 8 Overlap
- 9 Hub
- 10 Authority
- 11 In-Degree
- 12 Z-Score
- 13 Z-Number
- 14 Z-Degree

- [1] Wei, Chih-Ping, et al. "Finding experts in online forums for enhancing knowledge sharing and accessibility." *Computers in Human Behavior* 51 (2015): 325-335.
- [2] Li, Xiu, et al. "A Service Mode of Expert Finding in Social Network." *Service Sciences (ICSS)*, 2013 International Conference on. IEEE, 2013.
- [3] Kardan, Ahmad, Amin Omidvar, and Farzad Farahmandnia. "Expert finding on social network with link analysis approach." *Electrical Engineering (ICEE)*, 2011 19th Iranian Conference on. IEEE, 2011.
- [4] Kardan, Ahmad, Mehdi Garakani, and Bamdad Bahrani. "A method to automatically construct a user knowledge model in a forum environment." *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2010.
- [5] Liu, Duen-Ren, et al. "Integrating expert profile, reputation and link analysis for expert finding in question-answering websites." *Information Processing & Management* 49.1 (2013): 312-329.
- [6] Zhou, Guangyou, et al. "Topic-sensitive probabilistic model for expert finding in question answer communities." *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012.
- [7] Zhu, Hengshu, et al. "Ranking user authority with relevant knowledge categories for expert finding." *World Wide Web* 17.5 (2014): 1081-1107.
- [8] Zhou Zhao; Lijun Zhang; Xiaofei He; Ng, W., "Expert Finding for Question Answering via Graph Regularized Matrix Completion," *Knowledge and Data Engineering, IEEE Transactions on*, vol.27, no.4, pp.993,1004, April 1 2015.
- [9] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2233–2246, 2012.
- [10] R. Salakhutdinov and N. Srebro. Collaborative filtering in a nonuniform world: Learning with the weighted trace norm. *tc (X)*, 10:2, 2010.
- [11] Bouguessa, Mohamed, Benoît Dumoulin, and Shengrui Wang. "Identifying authoritative actors in question-answering forums: the case of yahoo! answers." *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008.
- [12] C.S. Campbell, P.P. Maglio, A. Cozzi and B. Dom. Expertise Identification using Email Communication. *Proceedings of the 12th ACM International Conference on Information and Knowledge Management (CIKM'03)*, pages 528-531, 2003.
- [13] B. Dom, I. Eiron, A. Cozzi and Y. Zhang. Graph-Based 8th Ranking Algorithms for E-mail Expertise. *Proceedings of ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'03)*, pages 42-48, 2003.
- [14] Page, L., Brin, S., Motwani, R. and Winograd, T. "The pagerank citation ranking: Bringing order to the web". Technical report, Stanford Digital. Library Technologies Project, 1998.

Archive