دوره ۹، بهار و تابستان ۱۳۹۵
مجله فناوری اطلاعات در طراحی مهندسی
Information Technology on Engineering Design
http://ited.sinaweb.net

تشخیص افراد خبره در شبکه های اجتماعی براساس خوشه بندی اشیاء اجتماعی و ماژولاریتی

سمانه حسن زاده* [۱]، ملیحه ابراهیمی نژاد [۲] ، دکتر مهرداد جلالی [۳]

[۱] دپارتمان مهندسی کامپیوتر فناوری اطلاعات، موسسه آموزش عالی اقبال لاهوری، مشهد،ایران
[۲] دپارتمان مهندسی کامپیوتر، دانشکده ۱۷ شهریور کرج، دانشگاه فنی و حرفه ای استان البرز، ایران
[۳] دپارتمان مهندسی کامپیوتر، دانشگاه آزاد اسلامی ، دانشکده فنی و مهندسی، مشهد، ایران

چکیده:

دهه گذشته توسعه سریعی را در شبکه های اجتماعی و استخراج جوامع شاهد است. نتیجه تحلیل چنین شبکه هایی می تواند کشف الگوهای مهم و پنهانی باشد. وجود حجم گسترده ای از محققان و اطلاعات در چنین شبکه هایی، منجر به تعیین متخصصین در زمینه های مختلف شده است. در این پژوهش یک روش ترکیبی پیشنهاد می کنیم که ابتدا با استفاده از الگوریتم EWKM، موضوعات به اشتراک گذاشته شده بین کاربران را در خوشه هایی قرار می دهد. سپس یک تحلیل ساختاری با استفاده از ماژولاریتی، بر روی کاربران در گیر در هر خوشه اعمال می شود. به علاوه برای دسترسی به متخصص ترین افراد در هر جامعه از الگوریتم Topsis، استفاده شده است. نتایج آزمایش نشان می دهد روش پیشنهادی دقت و صحت قابل قبولی نسبت به روش های پیشین دارد. بنابراین، روش پیشنهادی می تواند برای زمینه های دانشگاهی در جهان واقعی استفاده شود.

واژه‌های کلیدی:

پیدا کردن متخصص، استخراج جوامع، الگوریتم EWKM، ماژولاریتی، الگوریتم Topsis.

* عهده دار مکاتبات
نشانی: مشهد، بلوار سرافرازان ، سرافرازان ۹، موسسه آموزش عالی اقبال لاهوری، مشهد،ایران
تلفن:۰۹۱۵۱۸۲۵۵۱۵، پست الکترونیکی: hasanzadeh@eqbal.ac.ir

۱- مقدمه

با رایج تر شدن شبکه های اجتماعی نظیر فیس بوک تجزیه و تحلیل داده های شبکه به یکی از موضوعات حائز اهمیت در حوزه داده کاوی تبدیل شده است. تجزیه و تحلیل شبکه های اجتماعی علاوه بر درک عمیق ساختار شبکه، باعث کشف ارتباطاتی مانند روابط دوستی، تجاری یا علایق مشترک میان افراد خواهد شد که منجر به ایجاد گروه های منسجم در این شبکه ها می شود. با در اختیار داشتن چنین گروه هایی، دسترسی به افراد متخصص در هر زمینه نیز امکان پذیر خواهد بود[۱], [۲].

به طور معمول شبکه های اجتماعی را می توان در قالب گراف نمایش داد که در این گراف ها، گره ها معادل افراد یا بازیگران در شبکه های اجتماعی هستند و یال ها نشان دهنده ارتباط بین افراد هست. با توجه به ساختار شبکه اجتماعی و یک طرفه یا دوطرفه بودن ارتباطات، گراف متناظر می تواند جهت دار یا بدون جهت باشد. همچنین در صورتی که شدت و ضعف ارتباطات بین افراد در شبکه های اجتماعی یکسان نباشد، گراف متناظر با شبکه را یک گراف وزن دار در نظر گرفت که در آن وزن هر یال متناظر با استحکام ارتباط است.

در این تحقیق، برای استخراج متخصص ترین افراد در هر جامعه، ابتدا جوامع را از دو دیدگاه معنایی[۳–۵] و ساختاری [۶][۷] مشخص می کنیم و سپس با استفاده از الگوریتم Topsis[۸]، براساس معیارهای درجه، قدرت و بینابینی به هر کاربر در شبکه رتبه ای اختصاص می دهیم و در نهایت کاربران با بالاترین رتبه ها را به عنوان متخصصین در آن زمینه معرفی می کنیم. ساختار مقاله در ادامه به این شرح است: در بخش ۲ مروری بر کارهای محققان در این حوزه داریم. در بخش ۳ روش ماژولاریتی توضیح داده شده است. در بخش ۴ الگوریتم پیشنهادی را معرفی می‌کنیم. نتایج پیاده سازی روش پیشنهادی بر روی شبکه های اجتماعی واقعی در بخش ۵ بیان می‌شود. بخش ۶ شامل نتیجه گیری و کارهای آینده می باشد.

۲- مروری بر کارهای گذشته

۲-۱- روش‌های استخراج جامعه

بطور کلی، کارهای انجام شده در زمینه تشخیص جوامع را در دو صورت کلی می توان دسته بندی کرد، دسته اول تنها بر ساختار توپولوژی یا الگوهای ارتباطی تمرکز دارند بدون اینکه به موضوعات اشتراکی بین اعضاء توجه کنند و دسته دوم به موضوعات اشتراکی بین اعضاء توجه می کنند. در هر زمینه کارهای بسیاری انجام شده است. البته گروهی از محققین ترکیبی از این دو روش را استفاده کرده اند که مسلما نتایج حاصل از آن دقت بیشتری نسبت به دو روش ابتدا دارد. دسته اول را نیز می توان در دو بخش روش های مبتنی بر بهینه سازی و روش های اکتشافی تقسیم کرد[۹][۱۰]. روش های مبتنی بر بهینه سازی شامل روش های طیفی و روش های مبتنی بر جستجوی محلی است. هدف روش های طیفی به حداقل رساندن تابع برش است. Ruan و همکارش در سال ۲۰۰۷ برای استخراج جوامع، یک الگوریتم طیفی پیشنهاد داده اند[۱۰], [۱۱]. در حالی که هدف روش های مبتنی بر جستجوی محلی، بهینه سازی تابع هدف مانند ماژولاریتی است. تابع ماژولاریتی، برای ارزیابی

صلاحیت یک بخش خاصی از شبکه استفاده می شود. در اکثر کارهای انجام شده معمولا از ماژولاریتی برای ارزیابی صحت کارشان استفاده می کنند[۱۲], [۱۳]. روش های اکتشافی اغلب، الگوریتم های خوشه بندی گراف را مبتنی بر مفروضات حسی طراحی می کنند . Jin و همکارانش در سال ۲۰۱۱، از الگوریتم نیومن[۱۳], [۱۴] برای کشف جوامع استفاده کردند، مزیت کار این محققین استفاده از چندین منبع ناهمگون بود[۱۳]. روش های خوشه بندی گراف نیز روی ساختار توپولوژی تمرکز دارند[۱۵], [۱۶]. Yunhong و همکارانش در سال ۲۰۱۲ از یک روش معنایی در تحلیل شبکه های اجتماعی استفاده کردند و بر این اساس نیز خبره ترین شخص را در شبکه معرفی کردند[۱۷].

۲-۲- روش های تشخیص افراد خبره

پیدا کردن متخصصین در یک زمینه خاص اغلب مورد توجه افراد در بسیاری از حوزه های مختلف مانند دانشگاه و صنعت می باشد. کارهای بسیاری در این زمینه انجام شده است برخی از این روش ها در ادامه توضیح داده شده است.

۲-۲-۱- مدل های زبانی آماری

مدل های زبانی، معمولا توجه بسیاری از متخصصین را به دلیل اینکه مبتنی بر روش های آماری است به خود جلب کرده است. در این مدل عمدتا دو روش زیر مورد استفاده قرار می گیرد[۱۸], [۱۹][۲۰]:

- مدل زبانی ابتدایی
- مدل زبانی وزن دار

در روش مدل زبانی ابتدایی، ایده اصلی آن تخمین زدن یک مدل برای هر موضوع است و سپس این موضوعات را براساس Likelihood رتبه بندی می کنند. چندین روش مختلف برای محاسبه Likelihood ممکن است استفاده شود. در[۲۱] گروهی از محققین از یک موضوع معین با استفاده از یک مدل احتمالی برای تشخیص یک متخصص استفاده کردند که به صورت $P(Ca|q)$ بیان شد.

$$P(Ca|q) = \frac{P(q|Ca)P(Ca)}{P(q)} \qquad (۱)$$

که در آن $P(Ca)$، احتمال یک کاندید است و $P(q)$، احتمال یک پرس و جوی است و یک مقدار ثابت است که می توان در اهداف رتبه بندی از آن چشم پوشی کرد. در این مدل مقدار $P(q|Ca)$، به صورت زیر تعریف می شود:

$$P(q|Ca) = \sum_{d_j \in D_{Ca}} P(q|d_j)P(d_j|Ca) \qquad (۲)$$

(۳)

$$\begin{cases} P(d_j|Ca) = 1 & \text{اگر متخصص Ca با } d_j \text{ در ارتباط باشد} \\ P(d_j|Ca) = 0 & \text{اگر متخصص Ca با } d_j \text{ در ارتباط نباشد} \end{cases}$$

برای هر متخصص مجموعه ای از مستندات وجود دارد که با $D_{Ca} = \{d_j\}$، نشان داده است.

مدل زبانی ابتدایی که در بالا بیان شد تنها رابطه بین یک پرس و جو و یک سند را در نظر می گیرد ولی اعتبار و نفوذ سند را در نظر نمی گیرد. در روش مدل زبانی وزن دار، اعتبار موضوعات را نیز در نظر می گیرد. به عنوان مثال اگر دو سند A و B داشته باشیم که سند A مربوط به نویسنده a و سند B مربوط به نویسنده b است و هر دو سند محتوای مشابهی دارند یعنی $P(q|A) = P(q|B)$. اما اعتبار این دو سند متفاوت

است $I(A) > I(B)$. اما در این روش این مسئله را با دادن وزنی به هر سند در نظر می گیرد.

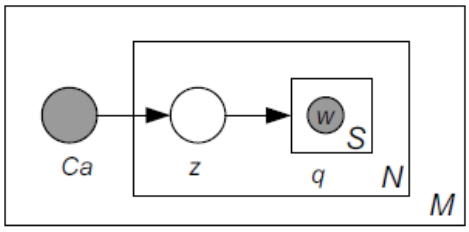
$$P(q|Ca) = \sum_d \{w_d \prod_{t\in q}((1-\lambda)P(t|d) + \lambda P(t))^{n(t,q)}\}P(d|Ca) \quad (۴)$$

۲-۲-۲- مدل های پیشرفته

در مدل های پیشرفته معمولا از یکی از دو روش زیر استفاده می شود[۱۸], [۱۹]:

- مدل متخصص- موضوع
- مدل ترکیبی

در مدل متخصص- موضوع، هر کاندید به صورت یک مجموع وزنی از چندین موضوع نشان داده شده است (زیرا هر متخصص ممکن است دانشی از چندین موضوع داشته باشد). بنابراین پرس و جوهای متفاوت از موضوعات مختلف ممکن است برای یک کاندید تولید شود. مدل متخصص- موضوع در شکل (۱) نشان داده شده است[۱۸].



شکل ۱ نمایش گرافیکی از مدل متخصص- موضوع[۱۸]

که در آن کاندید Ca و پرس و جوی q، از موضوع z مستقل هستند.

$$P_t(q|Ca) = \sum_{z\in Z} P(q|z)P(z|Ca) \quad (۵)$$

که در آن $P(z|Ca)$، احتمال اینکه موضوع z مربوط به کاندید Ca است را نشان می دهد و $P(q|z)$، احتمال اینکه پرس و جو q مربوط به موضوع z است را نشان می دهد. در شکل (۱) M، تعداد کاندید،N تعداد موضوعات تولید شده بوسیله کاندید Ca و S تعداد کلمات یک پرس و جو می باشد. در مدل ترکیبی، برای بهبود امتیاز بندی از یک پرس و جو مشخص در یک کاندید، از یک مدل ترکیبی برای جمع کردن امتیازاتی از مدل زبان و مدل متخصص-موضوع استفاده می شود.

$$P_h(q|Ca) = \mu P(q|Ca) + (1-\mu)P_t(q|Ca) \quad (۶)$$

نتایج تحقیقات انجام شده قبلی نشان می دهد که مدل های ترکیبی دقت بیشتری دارند[۱۸].

۲-۲-۳- سایر روش ها

Jin و همکارانش، از اندازه گیری درجه و بینابینی برای پیدا کردن متخصصان هر گروه استفاده کردند. یک عضو با درجه بالا یعنی ارتباطات محکم و بسیاری با متخصصان دیگر در گروه دارد و یک عضو با بینابینی بالا یعنی نقش مهمی برای ارتباط برقرار کردن با متخصصان دیگر در گروه دارد و یک عضو با درجه و بینابینی بالا یعنی شخص خیلی مهمی در گروه خواهد بود[۱۳][۲۲].

Moreira و همکارش، در روش حسگر چندگانه، برای یافتن متخصص ترین افراد در هر موضوع، هنگامی که کاربر موضوع مورد علاقه خود را بیان می کند در مرحله ابتدا، سیستم همه نویسندگانی که این موضوع، در چکیده و عنوان مقالات آن ها موجود است را بازیابی می کند و سپس با استفاده از سه سنسور، اطلاعات مورد نظر را از آن ها استخراج می کند. این سه سنسور عبارتند از: سنسور متنی، که وظیفه آن محاسبه تعداد تکرار موضوع در مقالات آن نویسنده است و سنسور

پروفایل، کـه تعـداد سـندهای منتشـر شـده از آن نویسنده را محاسبه می کنـد و سنسـور Citation، تعـداد نویسـندگان همکـار را محاسـبه مـی کنـد. براسـاس امتیـازات کسـب شـده از هـر سنسـور متخصصین در آن موضوع معرفی می شـوند امـا گـاهی اوقـات ممکـن اسـت نتـایج مختلفـی از سـه سنسـور حاصـل شـود بـرای رفـع ایـن مشـکل از تئوری دمپستر- شافر همراه بـا آنتروپـی شـانون استفاده شده است تا لیست نهایی از متخصصـین، دقیق تر و قابل اعتمادتر باشد[۲۳].

۳- روش حداکثر ماژولاریتی

ماژولاریتی به عنوان یک معیاری بـرای انـدازه گیری خوبی یک دسته بنـدی گـراف اسـت یکـی از امتیازات ماژولاریتی این است که مسـتقل از تعـداد خوشه هایی است کـه گـراف در آن تعـداد خوشـه تقسیم می شود. اگر یک خوشه معمولی در چنـدین جامعـه تقسـیم شـود، مـاژولاریتی Q^p، بـرای ایـن خوشه به صورت زیر تعریف می شود:

(۷)

$$Q^p = {}^1\!/_{(2m)} \sum_{vw} ((A_{vw} - (k_v k_w)/(2m)) \times \sum_i \delta(C_v, i)\delta(C_w, i) = \sum_i (e_{ii} - a_i^2)$$

که در آن:

- v, w، رئوس در این خوشه هستند.
- A_{vw}، تعداد ارتباطات بـین رئـوس v, w را نشان می دهد.
- $m = \frac{1}{2}\sum_{vw} A_{vw}$
- C_v، جامعـــه ای کـــه راس v، بـــه آن اختصاص داده می شود.
- i، جامعه i را نشان می دهد.
- $\delta(x, y)$، مسـاوی یـک اسـت اگـر $x = y$ باشد در غیر این صورت مسـاوی صـفر است.
- $e_{ij} = \frac{1}{(2m)}\sum_{vw} A_{vw}\delta(C_v, i)\delta(C_w, j)$
- $a_i = \frac{1}{(2m)}\sum_v k_v \delta(C_v, i)$

در روش سنتی ماژولاریتی، در ابتـدا هـر راس را به عنوان یک جامعه در نظر می گیرد و تغییـرات Q^p، محاسبه می شـود و سـپس بزرگتـرین آن هـا انتخاب می شود و ادغام جوامع انجـام مـی شـود. سه ساختار داده بـرای ایـن منظـور نگهـداری مـی شود:۱) ماتریس پراکنده ΔQ_{ij}^p ، کـه حـاوی جوامـع j و i با حداقل یک همکاری می باشـد. ۲) مـاتریس H، که شامل بزرگترین عنصر از هر سطر ماتریس ΔQ_{ij}^p ، همـراه بـا برچسـب هـای جوامـع j و i مـی باشد .۳) آرایه a_i . مقدار دهـی اولیـه بـرای ΔQ_{ij}^p و a_i به صورت زیر است:

(۸)

$$\Delta Q_{ij}^p = \begin{cases} \frac{1}{(2m)} A_{ij} - (k_i k_j)/(2m), & \text{در صورت متصل بودن } i \text{ و } j \\ 0, & \text{در غیر این صورت} \end{cases}$$

(۹) $$a_i = {}^{k_i}\!/_{2m}$$

قوانین به روز رسانی برای مـاتریس ΔQ^p، بـه صورت زیر است:

$$\Delta Q_{ij}^{\acute{p}} = \begin{cases} \Delta Q_{ik}^p + \Delta Q_{jk}^p & \text{اگر جامعه } k \text{ به هر دو } i \text{ و } j \text{ متصل باشد} \\ \Delta Q_{ik}^p + 2a_j a_k & \text{اگر جامعه } k \text{ به } i \text{ و نه } j \text{ متصل باشد} \\ \Delta Q_{jk}^p - 2a_i a_k & \text{اگر جامعه } k \text{ به } j \text{ و نه } i \text{ متصل باشد} \end{cases}$$

(۱۰)

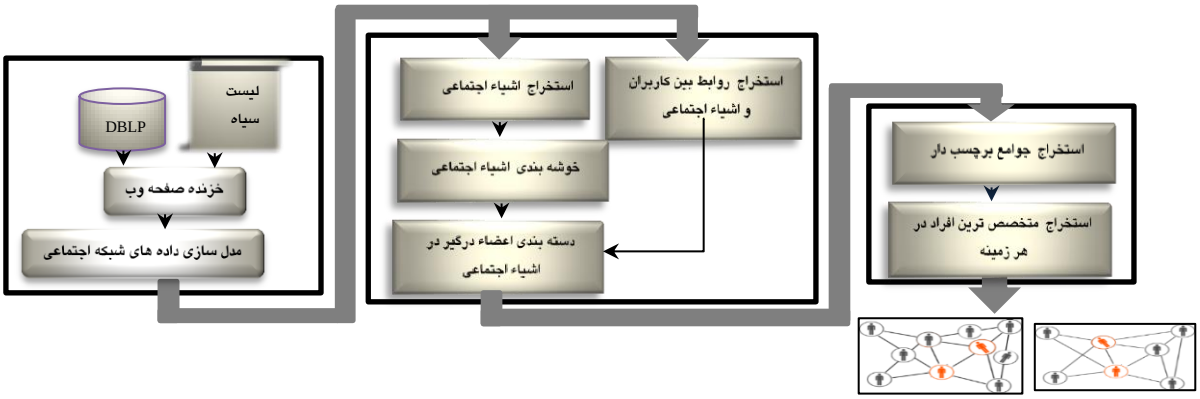
۴-روش پیشنهادی

چارچوب روش پیشنهادی در شکل (۴) نمایش داده شده است. این چارچوب برای تشخیص متخصص ترین افراد در هر جامعه، شامل ۳ بخش کلی زیر می باشد:

۱- آماده سازی مجموعه داده : هدف این بخش، استخراج داده های مورد نیاز از صفحات وب و پاکسازی داده ها می باشد.

۲- استخراج جوامع : هدف این بخش، دسته بندی اشیاء اجتماعی و سپس دسته بندی اعضاء درگیر در آن ها می باشد.

۳- استخراج متخصص ترین افراد در هر زمینه : در این بخش، افراد براساس ویژگی هایشان در هر زمینه موضوعی رتبه بندی می شوند و سپس افراد با بالاترین رتبه به عنوان متخصص ترین افراد معرفی می شوند.



شکل ۴ چارچوب روش پیشنهادی

۴-۱- آماده سازی مجموعه داده

پس از استخراج اطلاعات مورد نیاز با استفاده از خزنده صفحه وب، از سایت های معتبر علمی و انجام عمل پاک سازی بر روی آن ها، این اطلاعات در قالب یک مدل نمایش داده می شود. با توجه اینکه افراد در شبکه های اجتماعی براساس اشیاء اجتماعی، با هم ارتباط برقرار می کنند، یک گراف رسمی برای توضیح شبکه اجتماعی در نظر گرفته می شود. گراف مربوطه به صورت یک سه تایی تعریف می شود.

$$EG = (U, O, E) \qquad (۱۱)$$

که در آن:

U، مجموعه ای از کاربران درگیر، در فعالیتهای اجتماعی هستند و O، مجموعه ای از اشیاء اجتماعی و E، مجموعه ای از لبه هاست که ارتباطاتی که در EG موجود است را نشان می دهد.

$$E = E_{UU} \cup E_{UO} \qquad (۱۲)$$

۴-۲- استخراج جوامع برچسب دار

۴-۲-۱- استخراج اشیاء اجتماعی

اشیاء اجتماعی در واقع موضوعات به اشتراک گذاشته

بین افراد می باشد. در این مرحله، در ابتدا محتوای اشیاء اجتماعی به صورت جفت های (t_i, mea_i)، نمایش داده می شود که در آن t_i، کلمه و mea_i، اندازه t_i است. برای نشان دادن هر شی از مدل فضا بردار استفاده می شود[۳].

$$V_j = ((t_1, mea_1), \dots, (t_n, mea_n)) \quad j = 1, \dots, m \qquad (۱۳)$$

که در آن

- m، تعداد کل اشیاء اجتماعی است.
- n، تعداد کل کلمات موجود بعد از پاک سازی است.

مقدار mea_i، با توجه به الگوریتم TF/IDF، محاسبه می شود. به مجموع کل کلمات منحصر به فرد در کل اشیاء اجتماعی بعد گفته می شود. در نهایت، همه اشیاء اجتماعی بوسیله ماتریس $M_{m\times n}$، مشخص می شود که m تعداد اشیاء اجتماعی و n تعداد کل ابعاد می باشد.

۴-۲-۲- خوشه بندی اشیاء اجتماعی

براساس محتوای اشیاء اجتماعی به دست آمده از مرحله قبل، با استفاده از الگوریتم EWKM[۲۴]، اشیاء اجتماعی در خوشه هایی دسته بندی می شوند. این الگوریتم نسخه ای از الگوریتم[۲۵], [۲۶] K-means است با این تفاوت که علاوه بر ایجاد خوشه ها، وزنی را به هر بعد اختصاص می دهد. الگوریتم۱، روند این الگوریتم را نمایش می دهد.

الگوریتم ۱: الگوریتم EWKM : برای خوشه بندی اشیاء اجتماعی براساس محتوای اشیاء[۳], [۲۴]

ورودی: ماتریس $M_{m\times n}$ و تعداد خوشه ها K=7.

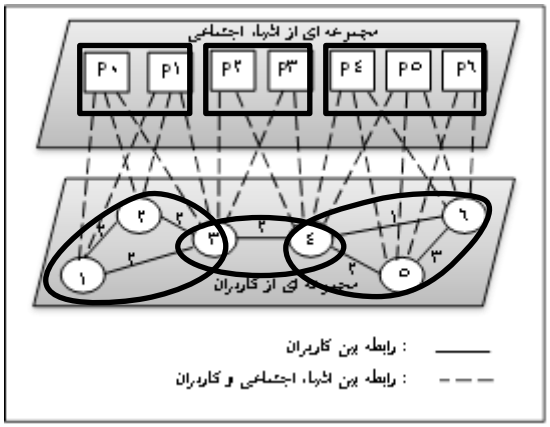
خروجی: خوشه هایی از اشیاء اجتماعی و وزن ابعاد برای هر خوشه.

۱. k نقطه به عنوان مراکز اولیه خوشه ها به طور تصادفی انتخاب می شود و در ابتدا ماتریس وزن ابعاد با مقدار $\frac{1}{m}$ مقدار دهی می شود.
۲. گام های ۳ تا ۵، تا زمانی که مراکز خوشه ها تغییری نمی کنند و یا تا رسیدن به حداکثر تکرار و یا تا رسیدن به حداقل مقدار تابع هدف و یا تا رسیدن به حداکثر وزن اجرا می شوند.
۳. نقاط را به هر یک از k خوشه به نحوی اختصاص می یابند که فاصله نقاط تا مراکز خوشه ها حداقل شود.
۴. مراکز خوشه ها با استفاده از نقاطی که به هر خوشه اختصاص یافته اند به روز آوری می شوند.
۵. ماتریس وزن ابعاد به روز آوری می شود.
۶. پایان.

علاوه بر خوشه ها، وزن تمام ابعاد در خوشه ها مشخص می شود و کلماتی که دارای بیشترین وزن هستند برای برچسب گذاری خوشه های استخراج شده، انتخاب می شوند.

۴-۲-۳- دسته بندی اعضاء درگیر در اشیاء اجتماعی

در این بخش با توجه به خوشه های اشیاء اجتماعی که از مرحله ی خوشه بندی اشیاء اجتماعی بدست آوردیم و همچنین با توجه به ارتباط بین کاربران که در بخش مدل سازی مشخص کردیم، کاربران در گیر را در خوشه هایی دسته بندی می کنیم. هر خوشه یک زیر گراف بدون جهت از گراف اصلی است که توسط یک ماتریس نمایش داده می شود. سطر و ستون های این ماتریس مشخص کننده کاربرانی است که با هم در آن خوشه در ارتباط هستند و محتوای هر سلول این ماتریس، نشان دهنده تعداد همکاری های بین دو کاربر درآن خوشه می باشد.



شکل ۵ تخصیص کاربران به خوشه ها براساس خوشه بندی اشیاء اجتماعی

در شکل (۵)، نحوه ی تخصیص کاربران به خوشه ها نمایش داده شده است.

الگوریتم ۲ : الگوریتم حداکثر ماژولاریتی[۳]

ورودی: P خوشه p=1,2,3,….

خروجی: جوامع برچسب گذاری شده

۱. مقادیر اولیه ΔQ_{ij}^{p} و a_i بر طبق فرمول۸ و ۹ محاسبه می شود.

۲. گام های ۳ تا ۷ ، تا زمانی که تعداد جوامع جاری بزرگتر از یک است اجرا می شود.

۴-۲-۴- خوشه بندی کاربران

هدف از این بخش، به دست آوردن جوامع از طریق تقسیم بندی هر خوشه به ناحیه هایی براساس میزان ارتباط آن هاست. کاربران در هر خوشه اغلب با قدرت های متفاوتی با هم در ارتباط هستند. کاربرانی که ارتباط بیشتری با هم دارند، باعث به وجود آمدن اتصال قوی تری بین آن ها خواهد شد، همچنین اتصالات ضعیف تر مربوط به کاربرانی است که ارتباطات کمتری با یکدیگر دارند. در این تحقیق از روش حداکثر ماژولاریتی، جهت تجزیه و تحلیل ساختار گراف ارتباطی و مشخص کردن جوامع در هر خوشه استفاده شده است. ساختار این جوامع طوری است که گره های داخل هر جامعه دارای اتصال قوی تری نسبت به اتصال گره های خارج از آن جامعه می باشند.الگوریتم ۲، الگوریتم این روش ، که برای هر خوشه مورد استفاده قرار گرفته است در زیر توصیف شده است.

۳. برای هر ردیف از ماتریس ΔQ^p، گام ۴ اجرا می شود.
۴. H با بزرگترین عناصر ایجاد می شود.
۵. بزرگترین ΔQ_{ij}^p، از H انتخاب می شود.
۶. جوامع مشابه با هم یکی می شوند.
۷. ماتریس ΔQ^p، به روز رسانی می شود.
۸. پایان

۳-۴ پیدا کردن متخصص ترین افراد

پس از انجام مرحله قبل، جوامع برچسب گذاری شده استخراج شده، که این برچسب ها بیان کننده موضوع اشتراکی بین اعضاء می باشد. پیدا کردن متخصصین در یک زمینه خاص اغلب مورد توجه افراد در بسیاری از حوزه های مختلف مانند دانشگاه و صنعت می باشد. با استفاده از جوامع استخراج شده از مراحل قبل به راحتی می توانیم متخصص ترین افراد را در یک زمینه خاص جستجو کنیم. برای این منظور پس از استخراج جوامع، از بینابینی و درجه و قدرت نود برای بررسی متخصص ترین افراد در هر گروه استفاده شده است[۱۳]. چندین عضو با درجه بالا یعنی با افراد بیشتری در گروه در ارتباط است و چندین عضو با قدرت بالا یعنی این افراد، اعضای مهمی در آن جامعه هستند و چندین عضو با بینابینی بالا یعنی نقش مهمی برای ارتباط برقرار کردن با متخصصان دیگر در گروه دارند. اعضاء با درجه و قدرت و بینابینی بالا یعنی افراد خیلی مهمی در گروه خواهد بود.

در این مرحله جوامعی که دارای یک موضوع هستند استخراج می شوند و سپس در هر جامعه، برای تمامی نودها مقادیر بینابینی، درجه و قدرت محاسبه می شود و با توجه به این سه ویژگی برای تعیین متخصص ترین افراد در هر جامعه، از الگوریتم Topsis[۲۶] استفاده شده است. در نهایت، الگوریتم ۳، روش پیشنهادی برای کشف متخصص ترین افراد در شبکه های همکاری را نمایش می دهد.

الگوریتم ۳: الگوریتم روش پیشنهادی

ورودی: مجموعه داده شبکه اجتماعی

خروجی: متخصص ترین افراد در هر جامعه برچسب گذاری شده ، جوامع برچسب گذاری شده

۱. خزنده صفحه وب، برای استخراج اطلاعات مورد نظر از صفحات وب اجرا می شود.
۲. عملیات پاک سازی بر روی مجموعه داده ورودی اعمال می شود.
۳. داده های شبکه اجتماعی مدل سازی می شود.
۴. ماتریس اشیاء اجتماعی ایجاد می شود.
۵. الگوریتم EWKM، برای خوشه بندی اشیاء اجتماعی اجرا می شود.

۶. کاربران براساس خوشه بندی اشیاء اجتماعی از مرحله ۵ و روابط بین کاربران از مرحله ۳، در خوشه های مجزایی قرار می گیرند.
۷. الگوریتمModularity، برای هر خوشه از مرحله ۶ اجرا می شود.
۸. جوامع با برچسب برای تعیین متخصص ترین افراد از لیست جوامع انتخاب می شود.
۹. الگوریتم Topsis برای تعیین متخصص ترین افراد در هر گروه اجرا می شود.
۱۰. پایان

- پیچیدگی زمانی الگوریتم پیشنهادی

پیچیدگی زمانی الگوریتم پیشنهادی به الگوریتم خزنده صفحه وب، EWKM، الگوریتم ماژولاریتی و الگوریتم Topsis وابسته است. پیچیدگی زمانی هر یک از این الگوریتم ها در جدول زیر بیان شده است:

جدول ۱ پیچیدگی زمانی الگوریتم های استفاده شده در روش پیشنهادی

نام الگوریتم	پیچیدگی زمانی	توضیحات
خزنده صفحه وب	$O(M)$	M: تعداد اشیاء اجتماعی
EWKM	$O(M \times N \times K \times T)$	M: تعداد اشیاء اجتماعی، N: تعداد کلمات کلیدی در همه اشیاء اجتماعی، K: تعداد خوشه های در نظر گرفته برای خوشه بندی اشیاء اجتماعی، T: تعداد تکرار
الگوریتم Modularity	$O(hmnk$	K: تعداد خوشه های در نظر گرفته برای خوشه بندی اشیاء اجتماعی ، N_k: تعداد کاربران در جامعه k ام.
Topsis	$O(F \times N_k)$	F ، تعداد ویژگی های در نظر گرفته برای هر کاربر، N_k: تعداد کاربران در جامعه k ام.

پیچیدگی زمانی الگوریتم خزنده صفحه وب و Topsis نسبت به دیگر الگوریتم ها قابل چشم پوشی است. بنابراین پیچیدگی زمانی الگوریتم پیشنهادی برای تشخیص متخصص ترین افراد در هر جامعه برابر $O(MNKT + K\mathrm{N}_k^2 \log N_k)$ می باشد.

۵-ارزیابی روش پیشنهادی

۵-۱- مجموعه داده مورد آزمایش

در ابتدا از مجموعه داده DBLP ، به عنوان مجموعه داده پایه استفاده شده است.[۲۸]DBLP، یک وب سایت مرجع کامپیوتر می باشد که توسط دانشگاه ترير درکشور آلمان ایجاد شده است. DBLP یک پایگاه داده مرجع برای رشته کامپیوتر می باشد و از حدود سال ۱۹۸۰ وجود دارد. این پایگاه داده تا تاریخ اکتبر ۲۰۱۳ بیش از ۲,۳ میلیون مقاله رشته کامپیوتر را لیست کرده است. البته این پایگاه داده معمولا مقالات چاپ شده در کنفرانس ها و مجلات معتبر خاصی را در نظر می گیرد. مثلا ممکن است شخصی ۳۰ مقاله داشته باشد ولی فقط ۱۰ مقاله از آن شخص در این پایگاه داده قرار گرفته باشد. به هر حال این پایگاه داده می تواند مرجع مناسب و معتبری برای تشخیص اعتبار کنفرانس ها، مجلات و یا حتی سطح علمی افراد باشد. این مجموعه داده حاوی Article, Inproceedings, Proceedings,Book و Thesis می باشد. در این مقاله تنها مجموعه ی Articleها در نظر گرفته شده است. اطلاعات موجود در این مجموعه داده عنوان مقاله، نام نویسندگان، نام مجله، سال انتشار و آدرس مقاله می باشد. با توجه به اطلاعات موجود در مجموعه داده و نداشتن چکیده مقالات و کلمات کلیدی، با

استفاده از بخش خزنده اطلاعات مورد نظر استخراج شده و در پایگاه داده، قرار داده شده است. در این مجموعه داده، مقالات، اشیاء اجتماعی و نویسندگان کاربران درگیر در اشیاء اجتماعی هستند.

- معیار ارزیابی

برای بررسی صحت و کارایی، از معیار بازیابی اطلاعاتP@n استفاده شده است[۲۳].

P@n: یکی از معیارهای ارزیابی است که دقت را در n رتبه اول ارزیابی می کند و به صورت زیر محاسبه می گردد.

$$P@n = \frac{\text{کاندیدهای مرتبط در } n \text{ نتیجه بالا}}{n} \quad (۱۴)$$

۵-۳- نتایج آزمایش

پس از استخراج داده های مورد نظر و آماده سازی اشیاء اجتماعی در مرحله خوشه بندی اشیاء اجتماعی وزن کلمات کلیدی محاسبه شده است و در نهایت همه اشیاء اجتماعی بوسیله ماتریس $M_{m\times n}$، مشخص می شوند که m تعداد اشیاء اجتماعی و n تعداد کل کلمات کلیدی است. این ماتریس، به الگوریتم EWKM[۲۴] ارسال می شود مقدار K=7، در نظر گرفته شده است[۳]. بخشی از نتایج در جداول (۲) و (۳) بیان شده است.

جدول۲ برچسب های تخصیص داده شده به خوشه ۴ پس از اجرای الگوریتم EWKM

شماره خوشه	برچسب
۴	Semanticweb, Communication, Mine, Associative rule, vision,Face Recognition,Image,Algorithm classification,clustering, Learning

جدول(۳) برچسب های اختصاص داده شده به دو جامعه نهایی مربوط به موضوع Machine vision، را نشان می دهد.

جدول ۳ برچسب های اختصاص داده شده به جوامع استخراج شده

شماره جامعه	برچسب	موضوع
۱۲۰۰	Face image, Face recognition, Pattern analysis, machine intelligence	Machine vision
۱۰۱۱	Image feature, Clustering algorithm, Image analysis, Image retrieval	Machine vision

در مرحله نهایی برای تعیین متخصص ترین افراد، جوامعی که دارای یک موضوع هستند استخراج می شوند و سپس در هر جامعه، با استفاده از الگوریتم Topsis، براساس معیارهای درجه، قدرت و بینابینی به هر کاربر در جامعه رتبه ای اختصاص داده شد و در پایان کاربران با بالاترین رتبه ها، به عنوان متخصصین در آن زمینه معرفی شدند.

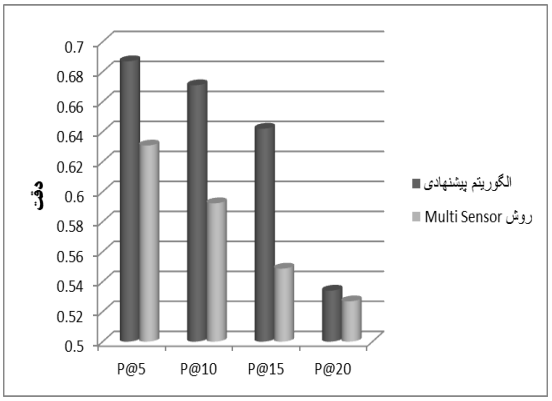
۵-۴- مقایسه و ارزیابی

برای ارزیابی صحت تحقیق انجام شده، نیاز به مجموعه ای درست از متخصصین در هر موضوع داریم. مجموعه داده استفاده شده شامل ۱۳ موضوع از دامنه علوم کامپیوتر است که بوسیله افراد حاضر در کنفرانس های مهم ساخته شده است[۲۳]. جدول (۴)، یک Benchmark ازتعداد متخصصان مربوط به هر موضوع را نشان می دهد.

جدول ۴ مشخصات مجموعه داده استفاده شده برای ارزیابی صحت روش پیشنهادی

موضوع	تعداد نویسندگان	موضوع	تعداد نویسندگان
Boosting (B)	۴۶	Natural Language (NL)	۴۱
Computer Vision (CV)	۱۷۶	Neural Networks (NN)	۱۰۳
Cryptography (C)	۱۴۸	Ontology (O)	۴۷
DataMining (DM)	۳۱۸	Planning (P)	۲۳
Information Extraction (IE)	۲۰	SemanticWeb (SW)	۳۲۶
Intelligent Agents (IA)	۳۰	Support VectorMachines (SVM)	۸۵
Machine Learning	۳۴		

نتیجه مقایسه روش پیشنهادی با الگوریتم استفاده شده در شکل (۷) آورده شده است.



شکل ۷ مقایسه روش پیشنهادی و روش حسگر چندگانه

بر طبق نتایج حاصل، الگوریتم پیشنهادی در مقایسه با روش استفاده از چند سنسور از دقت و کارایی خوبی در استخراج افراد متخصص برخوردار است .

۶- جمع بندی و کارهای آینده

در این تحقیق به معرفی شبکه های اجتماعی، جوامع در شبکه های اجتماعی و تعیین متخصصین در هر جامعه برچسب دار پرداخته شد. در این راستا، برای استخراج جوامع از روشی استفاده شد که علاوه بر در نظر گرفتن ساختار موجود در شبکه، موضوعات به اشتراک گذاشته شده بین افراد را در نظر می گیرد، برای این منظور برای خوشه بندی اشیاء اجتماعی از الگوریتم EWKM استفاده شد این الگوریتم نسخه ای از الگوریتم K-means است با این تفاوت که علاوه بر خوشه بندی، به هر یک از بعدها وزنی اختصاص می دهد که از ابعاد با بیشترین وزن برای برچسب گذاری این خوشه ها استفاده شد، و سپس براساس این خوشه ها، کاربران درگیر در این اشیاء اجتماعی را در خوشه های مجزایی قرار دادیم به منظور دسترسی به جوامع دقیق تر یک تجزیه و تحلیل ساختاری بر روی خوشه ها با استفاده از الگوریتم ماژولاریتی انجام شد و همچنین بعدها با بیشترین وزن در گیر در این خوشه ها به عنوان برچسب جوامع در نظر گرفته شد و در نهایت برای دسترسی به متخصص ترین افراد در هر جامعه با استفاده از الگوریتم Topsis، براساس معیارهای درجه، قدرت و بینابینی به هر کاربر در شبکه رتبه ای اختصاص داده شد و در پایان کاربران با بالاترین رتبه ها، به عنوان متخصصین در آن زمینه معرفی شدند. از جمله موارد برای بهبود کیفیت در نتایج پیشنهادی می توان به موارد زیر اشاره نمود:

استفاده از بانک های اطلاعاتی مرتبط با هستی شناسی، جهت حذف کلمات هم معنی در محتوای استخراج شده از خروجی خزنده صفحه وب و عنوان مقاله.

استفاده از اطلاعات کاربران درگیر در شبکه، از دیگر شبکه هایی که در آن فعالیت دارند. به عنوان مثال استخراج زمینه کاری نویسنده از شبکه هایی که در آن فعالیت دارد، اما ممکن است نویسنده در هر شبکه ای روی زمینه ی خاصی فعال باشد و نتایج مختلفی برای یک نویسنده حاصل شود که این موضوع خود جای کار دارد و یا در مورد موقعیت مکانی نویسندگان، که می توان درصد نویسندگان متخصص در هر موضوع را در هر منطقه مشخص کنیم.

استفاده از برخی اطلاعات نهفته دیگر در مجموعه داده. به عنوان مثال مدت زمان فعالیت یک نویسنده در یک موضوع می تواند به عنوان یک معیار دیگر در تعیین متخصص بودن نویسندگان استفاده شود.

نتایج حاصل از تجزیه و تحلیل شبکه های اجتماعی و تعیین جوامع و متخصصین در هر جامعه می تواند نقش مهمی را در رسیدن افراد به اهدافشان ایفا کند. با توجه به مطالب بیان شده در این مقاله می توانیم در بسیاری از برنامه های کاربردی از این روش استفاده کنیم.

مراجع و منابع

[1] L. Tang and H. Liu, *Community detection and mining in social media*, vol. 2, no. 1. Morgan & Claypool Publishers, 2010, pp. 1–137.

[2] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.

[3] Z. Zhao, S. Feng, Q. Wang, J. Z. Huang, G. J. Williams, and J. Fan, "Topic oriented community detection through social objects and link analysis in social networks," *Knowledge-Based Systems*, vol. 26, pp. 164–173, Feb. 2012.

[4] M. Steyvers and P. Smyth, "Probabilistic author-topic models for information discovery," ... *discovery and data mining*, no. 1990, pp. 1–10, 2004.

[5] T. Hofmann, "Probabilistic latent semantic indexing," *Proceedings of the 22nd annual international ACM ...*, pp. 50–57, 1999.

[6] S. Jarukasemratana, "Community Detection Algorithm based on Centrality and Node Distance in Scale-Free Networks," no. May, pp. 258–262, 2013.

[7] M. Newman, "Fast algorithm for detecting community structure in networks," *Physical review E*, no. 2, pp. 1–5, 2004.

[8] X. Zhu, F. Wang, H. Wang, C. Liang, R. Tang, X. Sun, and J. Li, "TOPSIS method for quality credit evaluation: A case of air-conditioning market in China," *Journal of Computational Science*, pp. 1–7, Feb. 2013.

[9] H. Jiawei, K. Micheline, and P. Jian, *Data Mining concepts and techniques*. 2012, pp. 452–454.

[10] C. C.Aggarwal, *Social Network Data Analytics*. 2011, pp. 1–502.

[11] J. Ruan and W. Zhang, “An Efficient Spectral Algorithm for Network Community Discovery and Its Applications to Biological and Social Networks,” *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pp. 643–648, Oct. 2007.

[12] Y. Kim, S. Son, and H. Jeong, “LinkRank: Finding communities in directed networks,” pp. 1–9, 2009.

[13] J. H. Jin, S. C. Park, and C. U. Pyon, “Finding research trend of convergence technology based on Korean R&D network,” *Expert Systems with Applications*, vol. 38, no. 12, pp. 15159–15171, Nov. 2011.

[14] S. White and P. Smyth, “A Spectral Clustering Approach To Finding Communities in Graph.,” in *SDM*, 2005, vol. 5, pp. 76–84.

[15] Y. Zhou, H. Cheng, and J. X. Yu, “Clustering Large Attributed Graphs: An Efficient Incremental Approach,” *2010 IEEE International Conference on Data Mining*, pp. 689–698, Dec. 2010.

[16] C. C. Aggarwal and H. Wang, *Managing and Mining Graph Data*, vol. 40. Boston, MA: Springer US, 2010.

[17] Y. Xu, X. Guo, J. Hao, J. Ma, R. Y. K. Lau, and W. Xu, “Combining social network and semantic concept analysis for personalized academic researcher recommendation,” *Decision Support Systems*, vol. 54, no. 1, pp. 564–573, Dec. 2012.

[18] F. Breu, S. Guggenbichler, and J. Wollmann, “A Study of Expert Finding on DBLP Bibliography Data,” *Vasa*, 2008.

[19] H. Deng, I. King, and M. R. Lyu, “Formal models for expert finding on DBLP bibliography data,” in *Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on*, 2008, pp. 163–172.

[20] J. Zeng, W. K. Cheung, C. Li, and J. Liu, “Coauthor Network Topic Models with Application to Expert Finding,” *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 366–373, Aug. 2010.

[21] K. Balog, L. Azzopardi, and M. De Rijke, “Formal models for expert finding in enterprise corpora,” *... of the 29th annual international ACM ...*, pp. 43–50, 2006.

[22] G. A. Wang, J. Jiao, A. S. Abrahams, W. Fan, and Z. Zhang, “ExpertRank: A topic-aware expert finding algorithm for online knowledge communities,” *Decision Support Systems*, vol. 54, no. 3, pp. 1442–1451, Feb. 2013.

[23] C. Moreira and A. Wichert, “Finding Academic Experts on a MultiSensor Approach using Shannon’s Entropy,” *Expert Systems with Applications*, pp. 1–29, 2013.

[24] L. Jing, M. Ng, and J. Huang, “An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data,” *Knowledge and Data Engineering, ...*, vol. 19, no. 8, pp. 1026–1041, Aug. 2007.

[25] M. El Agha and W. M. Ashour, “Efficient and Fast Initialization Algorithm for K-means Clustering,” *International Journal of Intelligent Systems and Applications*, vol. 4, no. 1, pp. 21–31, Feb. 2012.

[26] Z.-X. Wang and Y.-Y. Wang, “Evaluation of the provincial competitiveness of the Chinese high-tech industry using an improved TOPSIS method,” *Expert Systems with Applications*, vol. 41, no. 6, pp. 2824–2831, May 2014.

[27] Z. Gao and N. Jin, “Detecting community structure in complex networks based on K-means clustering and data field theory,” *Control and Decision Conference, 2008. CCDC ...*, pp. 4411–4416, 2008.

[28] “dblp.uni-trier.de.” [Online]. Available: http://dblp.uni-trier.de/xml/. [Accessed: 13-Oct-2013].