معرفی یک الگوریتم حریصانه جهت یافتن گروهی از خبرگان در شبکه‌های اجتماعی

ندا شیرویه[۱]، آسیه قنبرپور لیمویی[۲]، حسن نادری[۳]

[۱] دانشکده کامپیوتر، دانشگاه علم و صنعت ایران، تهران
Neda.shirooyeh@gmail.com

[۲] دانشکده کامپیوتر، دانشگاه علم و صنعت ایران، تهران
Ghanbarpour@iust.ac.ir

[۳] استادیار دانشکده کامپیوتر، دانشگاه علم و صنعت ایران، تهران،
Naderi@iust.ac.ir

چکیده

مسئله یافتن گروه خبرگان، یکی از مسائل تعریف شده در حوزه شبکه‌های اجتماعی است. در این مسئله، هدف انتساب گروهی از خبرگان که هر کدام دارای مهارت‌های خاصی هستند به مجموعه‌ای از کارها است که هر کار برای انجام نیازمند تخصص‌های ویژه‌ای است. این مسئله از جمله مسائل NP-Hard به شمار می‌رود و راه‌حل دقیق آن از مرتبه نمایی است. به همین دلیل ما در این مقاله با تکیه بر دو معیار پوشش‌دهی و خاص بودن، روشی حریصانه با مرتبه زمانی چندجمله‌ای را برای انتخاب مجموعه خبرگان پیشنهاد داده‌ایم. نتیجه اجرای این الگوریتم روی مجموعه داده‌های استاندارد نشان می‌دهد، الگوریتم پیشنهادی قادر به بازیابی پاسخ‌های بهینه در زمانی مناسب خواهد بود. علاوه بر این، روش پیشنهادی با چندین روش حریصانه که در سال‌های اخیر پیشنهاد شده‌اند، از لحاظ زمانی، میزان پوشش‌دهی و مقدار اطمینان مورد مقایسه قرار گرفته است که تمامی آزمایشات بیانگر برتری روش پیشنهادی از لحاظ کارایی و موثر بودن هستند.

کلمات کلیدی

شبکه اجتماعی، تشکیل گروه، افراد خبره، مهارت

Archive of SID

www.SID.ir

۱- مقدمه

موفقیت هر پروژه‌ای وابسته به مهارت افرادی است که آن پروژه را انجام می‌دهند [17]. هر اندازه مهارت افرادی که در انجام یک پروژه شرکت دارند مرتبط‌تر به موضوع پروژه بوده و میزان اشراف افراد بر جنبه‌ها متفاوت آن پروژه بیشتر باشد، امکان موفقیت بالاتری وجود دارد [19]. مسئله یافتن گروه خبرگان، یکی از مسائل تعریف شده در حوزه شبکه‌های اجتماعی است که هدف آن انتساب گروهی از خبرگان برای انجام یک کار[1] خاص است. این مسئله می‌تواند در زمینه‌های مختلفی چون یافتن افراد مناسب برای انجام پروژه‌ها، یافتن داوران برای داوری مقالات، پیدا کردن مکان مناسب جهت ساخت کارخانه و کارگاه‌هایی که باید دور از مکان‌های مسکونی باشند، یافتن مکان احداث ایستگاه‌های آتش‌نشانی، بیمارستان و سایر ارگان‌هایی که باید حتی‌الامکان نزدیک به مشتریان خود باشند، مورد استفاده قرار گیرد [12, 13] . در حالت کلی هر گاه نیاز به یافتن گروهی از اشیاء با ویژگی‌های مختلف جهت انتساب به گروه دیگری از اشیاء که آنها نیز ویژگی‌های مشخص خود را دارند وجود داشته باشد، می‌توان از راه حل یافتن گروه خبرگان با مهارت‌های مشخص استفاده کرد [3, 4, 8, 16, 10, 14, 25]. به عنوان مثال برای یافتن افراد مناسب جهت داوری مقالات، لیستی از داوران با توانایی‌های مختلف که گاهاً مشترک هستند و نیز لیستی از مقالاتی که موضوعات خاصی را مورد نگارش قرار داده و باید داوری شوند، ورودی‌های مسئله خواهند بود[3, 4, 8, 11, 12, 13, 16]. هدف انتساب بهترین گروه داوری به مقالات است به طوری که مجموعه دانش داوران کلیه موضوعات مورد بحث مقالات را پوشش دهند. در حل این مسئله باید توجه کرد که اگر هر داور از مهارت‌های بیشتر مورد نیاز داوری مقاله برخوردار باشد، توان داوری وی نسبت به زمانی که داور فقط یک یا تعداد محدودتری از مهارت‌های مورد نیاز مقاله را داراست، بیشتر خواهد بود [6, 9, 10].

امروزه با گسترش روزافزون شبکه‌های اجتماعی و اتصال گروه‌های علمی و فنی–مهارتی خبرگان به یکدیگر، شبکه‌های اجتماعی مکان مناسبی جهت یافتن افراد مناسب با مهارت‌های مطلوب هستند. افراد خبره با قراردادن پروفایل و سابقه عملکرد خود در محیط‌های اجتماعی امکان دیده شدن خود را افزایش داده و شبکه اجتماعی را به عنوان ابزاری جهت یافتن افراد خبره با ویژگی‌های مورد نظر می‌گرداند [7, 9, 14, 23].

انتساب افراد خبره برای انجام یک کار/عملیات بر اساس مهارت‌های مورد نیاز پروژه یک مسئله NP-Hard بوده و راه حل دقیق آن دارای مرتبه زمانی نمایی است[9]. ما در این مقاله روشی حریصانه ارائه داده‌ایم که قادر به بازیابی پاسخ بهینه با مرتبه زمانی چندجمله‌ای است. توجه به خاص بودن مهارت‌ها موضوع جدیدی است که در این مقاله مورد توجه قرار گرفته و در منابع مورد مطالعه توجهی به آن نگردیده است. منظور از مهارت خاص، توانایی کمیابی است که در حداقل داوران یافت گردد. داوری که از چنین مهارتی برخوردار است باید در اولویت انتخاب به گونه‌ای قرار گیرد که مقالات خاص ابتدا به وی نسبت داده شوند. این درحالی است که ممکن است در صورت عدم توجه به خاص بودن مهارت این داور، ظرفیت وی با ارجاع سایر مقالات معمولی تکمیل شده و داوری مقالاتی که نیازمند تخصص وی است، در صف انتظار برای داوری باقی بمانند. در حالی که این امکان وجود دارد که مقالات معمولی‌تر به سایر داوران ارجاع شده و مقالات نیازمند مهارت خاص به داوران دارای مهارت خاص ارجاع گردد. این امر مسئله را از صرف یک انتساب حریصانه بر اساس تعداد مهارت‌های بیشتر خارج کرده و در نظر گرفتن دو اولویت انتخاب داوران با تعداد مهارت‌های بیشتر و انتساب داوران با تخصص خاص‌تر به مقالات را به صورت توامان می‌طلبد.

۲- کارهای مرتبط

برخی محققان در مقالات خود [1,7, 9, 10, 14, 15, 18, 23] بیشتر به وجه شبکه اجتماعی مسئله، نوع ورودی اطلاعات و ماهیت گرافی مسئله توجه داشته‌اند. هانگزی [7] با در نظر گرفتن ویژگی‌های اساسی شبکه‌های اجتماعی مانند نوع ارتباطات، رابطه دوستی و غیره به اندازه‌گیری معیارهایی چون تاثیر اجتماعی گره‌ها بر یکدیگر پرداخته و روشی برای محاسبه آن پیشنهاد داده است. بالوگ و مک‌دونالد [1, 15] به جستجوی افراد خبره در سازمان‌ها و ارگان‌های بزرگ پرداختند. کنستان [20] با استفاده از پرسش و پاسخ‌هایی که در فرم‌ها مطرح گردیده به جستجوی افراد خبره پرداخت. سان [22] بدون توجه به میزان پوشش‌دهی و قابلیت اطمینان مهارت‌های مقالات و فقط با توجه به عنوان مقالات به انتساب مقالات به داوران پرداخت. لاپاس [14] کار T را با مجموعه مهارت‌های مختلف X در نظر گرفته و در یک شبکه اجتماعی مانند G به حل مسئله یافتن خبرگان پرداخته است. کارگر [10] به جستجوی k گروه برتر از خبرگان در شبکه اجتماعی با/بدون داشتن راهنما پرداخته است. منظور از راهنما فردی است که نظاره‌گر و راهبر پروژه بوده و هزینه ارتباطی خاصی را در پروژه متحمل می‌گردد. هدف در مقاله وی یافتن یک مجموعه از خبرگانی است که همه مهارت‌های مورد نظر را پوشش داده و از کمترین هزینه ارتباطی با یکدیگر برخوردار باشند.

گروه دیگری از محققان توجه خود را به یافتن الگوریتم مناسب جهت پردازش اطلاعات ورودی بدون در نظر گرفتن ارتباط بین افراد و صرف یافتن راه حل مناسب معطوف داشته‌اند [10, 11, 12, 13, 16, 17, 18, 19]. یافتن k خبره برتر از بین کاندیداها، با استفاده از مدل بازیابی اطلاعات استاندارد و امتیازدهی افراد توسط کریم‌زادگان در سال 2012 مطرح شد، به صورت حریصانه عمل می‌کند. در این مدل ابتدا مهارت‌های هر خبره لیست شده و امتیازدهی می‌گردد. سپس خبرگان را بر اساس میزان امتیاز اکتسابی آنها مرتب‌سازی کرده و لیست می‌کنند [11, 12, 13]. در ادامه بر اساس لیست مهارت‌های مورد نیاز جهت انجام کار T، از خبره‌ای که بیشترین امتیاز یعنی بیشترین تعداد مهارت‌ها را داراست، شروع کرده و فرد خبره را به لیست منتخب می‌افزایند. سپس به سراغ خبره دوم رفته و بر اساس سایر مهارت‌هایی که هنوز پوشش داده نشده‌اند، خبره دوم را انتخاب می‌کنند. این کار تا پوشش تمامی خبرگان به صورت حریصانه ادامه پیدا خواهد کرد [12] . بالوگ و ازوپاردی نیز روشی با یافتن خبره با بیشترین امتیاز الگوریتم حریصانه‌ای را تا یافتن آخرین داور بر روی مجموعه داده‌ای اعمال کرده‌اند[1,2].

باسو [3] چکیده مقالاتی را که برخی داوران مقالات نوشته بودند، از طریق وب یافته و از TF-IDF برای مرتب‌سازی و امتیازدهی آنها استفاده کرد. بیسواس و حسن [4] از عناوین مقالات و نام نویسندگان آنها استفاده کرده و پس از وزن‌دهی با TF-IDF به مرتب‌سازی اطلاعات حاصله پرداختند. میمو و مک‌کالم [16] یک مدل APT[2] پیشنهاد کردند که در آن پس از استخراج عناوین مقالات هر نویسنده، مهارت‌های مربوط به آن

Archive of SID

www.SID.ir

استخراج می‌گردید. رودریگز و بولن [21] روشی را ارائه دادند که با جستجوی شبکه به یافتن گروه نویسندگان می‌پرداخت. وی این کار را با بررسی مقالات ارسالی نویسندگان انجام داد. کریم‌زادگان [13] مجموعه مهارت‌های مورد نیاز مقالات و نویسندگانی را در نظر گرفته و بر اساس تطبیق مهارت‌ها با یکدیگر عمل کرد. نشاطی [17] چهار فاکتور و سه مسئله را مطرح کرد و بر اساس چهار فاکتور مطرح شده به حل سه مسئله و یافتن پاسخ برای آنها پرداخت. در این مسئله k داور مرتبط (یک گروه از خبرگان با k عضو) باید به هر مقاله منتسب شود به گونه‌ای که چهار شرط زیر در نظر گرفته شود. اول آنکه مهارت‌های مورد نیاز برای مطالعه مقالات توسط برخی کلمات کلیدی و یا با استخراج از چکیده یا بدنه مقالات بدست می‌آید. دوم آنکه محدوده مهارت‌های هر خبره یا داور از کلمات کلیدی و یا منتشرات قبلی بدست می‌آید. سوم آنکه در یک نگاشت ایده‌آل گروه منتسب برای یک مقاله باید تمامی مهارت‌های مورد نیاز مقاله را پوشش دهند. چهارم اینکه ترجیح می‌دهیم هر داور منتسب به یک مقاله بیشترین تعداد ممکن از مهارت‌های مورد نیاز مقاله را پوشش دهند. آخر اینکه هر فرد خبره فقط درگیر تعداد محدودی از مقالات باشند.

تمامی موارد ذکر شده از روش حریصانه بالا به پایین استفاده می‌کنند. مشکلی که در این روش‌ها رخ می‌دهد عدم توجه به خاص بودن مهارت‌ها است. چرا که مرتب‌سازی بر اساس تعداد بیشتر مهارت‌های هر داور بوده و توجهی به تعداد مهارت‌هایی که از یک نوع وجود دارد (خاص بودن مهارت) نشده است. منظور از خاص بودن مهارت، وجود آن در تعداد محدودی داور است. حالتی را در نظر بگیرید که مقاله M_i برای داوری به مهارت T_kای نیازمند است که فقط در داور D_j وجود دارد. اگر داور D_j از تعداد زیاد مهارت‌های دیگری نیز برخوردار باشد، ممکن است قبل از انتساب داور به مقاله M_i، برای داوری مقالات دیگری انتخاب شده و ظرفیت وی تمام شود. در این صورت مقاله M_i بدون داور رها می‌شود. در حالی که این امکان وجود داشت، مقالات غیر خاص توسط سایر داوران داوری شده و داور D_j به M_i انتساب یابد.

۳- تعریف مسئله

شبکه اجتماعی با گراف G را در نظر بگیرید. n گره تشکیل دهنده گراف، افراد خبره $X=\{x_1,\ldots,x_n\}$ با مهارت‌های مختلف هستند به گونه‌ای که مجموعه مهارتهای فرد xi، را با A_i به صورت $A_i=\{a_1,\ldots a_m\}$ نشان می‌دهیم. فرض بر آن است که مهارت‌های هر فرد به صورت اطلاعات محلی[۳] در گره مربوط به آن فرد ذخیره شده است. برای انجام کار T، نیازمند بکارگیری زیرمجموعه‌ای از افراد خبره (X') با مجموع مهارت‌های A' هستیم به‌صورتی که $(T \subseteq A')$ باشد[9]. مجموعه A' به صورت زیر به دست می آید:

$$A' = \bigcup A_i \ \text{Such that} \ x_i \in X' \qquad (1)$$

این مسئله قابل تبدیل به مسئله مجموعه پوششی[۴] است که یک مسئله NP-Hard می‌باشد. لذا می‌بایست از روش‌های غیر دقیق و تقریبی برای حل آن استفاده کرده و در نهایت با مقایسه نتایج حاصل از حل مسئله با نتایج بدست آمده از طریق الگوریتم‌های دیگری که صحت آنها ثابت شده و یا با سوال از افراد خبره (نیروی انسانی) نتایج حاصل را بررسی نمود. روش‌هایی که برای حل مسئله مورد استفاده قرار می‌گیرد می‌تواند به صورت حریصانه یا تقریبی باشد[9]

داده های ورودی این مسئله لیست مقالات مورد داوری و عناوین مربوط به هر مقاله، به همراه لیستی از مهارت‌های داوران مقالات می‌باشد. این اطلاعات را می توان به دو شکل صریح و ضمنی به دست آورد. در حالت صریح، داده های موردنیاز آماده و مستقیماً قابل دسترسی است. به عنوان مثال این داده ها قبلا گردآوری شده و در یک ماتریس صفر و یک یا یک آرایه، موجود و قابل استفاده است. در حالت ضمنی، داده های موردنیاز مستقیما در دسترس نیستند و باید آنها را از طریق کاوش منابع مرتبط مثل داده های متنی، اسناد XML و داده های گرافی به دست آورد. به عنوان مثال در مسئله داوری مقالات، می توان موضوعات مورد بررسی هر مقاله را از روی عنوان مقاله، چکیده مقاله و متن مقاله به دست آورد. مهارت داوران نیز از مطالعه پروفایل داوران، انتشارات قبلی و یا مقالاتی که قبلاً داوری کرده‌اند، قابل حصول است.

در این مقاله ما فرض کرده ایم اطلاعات موردنیاز توسط یکی از روش‌های موجود، از شبکه اجتماعی استخراج شده و به صورت دو فایل متنی در دسترس است. فایل اول حاوی اطلاعات مربوط به افراد خبره است بطوریکه در هر سطر از این فایل مهارت های مربوط به یک خبره ذخیره شده است. فایل متنی دوم شامل اطلاعات مربوط به کارهای موردنیاز است. در هر سطر از این فایل مجموعه مهارت های لازم برای انجام هر کدام از کارها خاص ذخیره شده است. هدف یافتن راه حل و الگوریتمی است که در حداقل زمان، بهترین گزینه‌های را برای انجام هر کار پیشنهاد دهد.

٤- الگوریتم پیشنهادی

در این بخش یک روش حریصانه پیشنهاد شده است که بر اساس آن می‌توان با سرعت بالا، جواب‌هایی بسیار نزدیک به بهینه را یافت. بدین معنا که اگر در بهترین حالت و با بررسی تمامی حالات ممکن بتوان d داور را برای داوری یک مقاله یافت، الگوریتم بهینه نیز جوابی بسیار نزدیک به d داور (و یا دربسیاری موارد برابر d داور) را خواهد یافت.

الگوریتم پیشنهادی دو مجموعه مهارت های خبرگان و مهارت های موردنیاز برای کار موردنظر(T) را از ورودی گرفته و یک مجموعه خبرگان پوشاننده مهارت های موردنیاز را به عنوان خروجی تولید می کند. این الگوریتم به صورت تکراری عمل کرده و در هر مرحله امتیاز داوران مشخص شده و داوری با امتیا بیشتر به مجموعه پاسخ اضافه می شود. این مراحل تا انتخاب مجموعه داورانی که تمامی مهارت های مورنیاز را پوشش دهند ادامه می یابد. بنابراین در این الگوریتم تعیین امتیاز داوران در هر مرحله نقش اساسی را در تعیین مجموعه پاسخ نزدیک به بهینه ایفا خواهد کرد. ما برای امتیاز دهی به داوران دو معیار پوشش دهی و خاص بودن داوران را مد نظر قرار داده‌ایم. به این معنا که در هر مرحله باید داوری انتخاب شود که بتواند بیشترین تعداد مهارت‌های باقیمانده را پوشش دهد. اکثر کارهای گذشته این پارامتر را در نظر دارند. علاوه بر این ممکن است در بین مجموعه مهارت‌های خواسته شده، مهارتی موجود باشد که تنها یکی از خبرگان موجود تخصص انجام آن را داشته باشد. مطئناً برای پوشش تمامی مهارت‌های لازم، این خبره در مجموعه جواب حضور خواهد داشت بنابراین باید امتیاز بالایی به این خبره داده شود. با انتخاب یک داور در هر مرحله از اجرای الگوریتم تعدادی از

Archive of SID

www.SID.ir

مهارت‌های مورد نیاز پوشش داده می‌شوند. بنابراین امتیاز داوران در مرحله بعد به روز خواهد شد. این الگوریتم به صورت دقیق‌تر در ادامه آمده است.

فرض کنید دو ماتریس مجموعه مهارت‌های خبرگان و مجموعه مهارت‌های موردنیاز به الگوریتم داده شده است. ماتریس داوران (D) یک ماتریس با مولفه‌های صفر و یک است که هر سطر آن نشان دهنده یک خبره است. ستون‌های ماتریس D، مهارت‌های خبرگان را نشان می‌دهد. در صورتی که داور j مهارت k را داشته باشد در سطر jام از ماتریس D و ستون kام آن عدد ۱ درج شده و در غیر صورت عدد صفر وجود دارد. ما در یک مرحله پیش پردازش مهارت‌های مورد نیاز مقالات که احیاناً در لیست مهارت داوران وجود ندارد را به طور کامل حذف می‌کنند. چون مطمئنا برای پوشش کامل این مهارت‌ها نمی‌توان پاسخی را یافت.

جمع هر ستون ماتریس D نشان‌دهنده تعداد خبرگان متخصص در آن مهارت را نشان می‌دهند. بنابراین با استفاده از این مجموع می‌توان مهارت‌هایی که تنها یک داور در آنها تخصص دارند را یافت. به عنوان مثال اگر ستون مربوط به مهارت x از ماتریس D فقط یک عدد ۱ داشته باشد یعنی مهارت x بسیار خاص است و باید در انتخاب داوران، داوری که این مهارت را دارد، حتماً انتخاب گردد. بر این اساس به هر مهارت امتیاز α^{-n} نسبت داده می‌شود. در نتایج تجربی α در α^{-n} برابر ۱۰ حاصل شده است و n بیانگر تعداد تکرار مهارت x است. آنگاه اگر مهارت x در تخصص دو داور مختلف باشد، وزن 0.01 و اگر در تخصص سه داور باشد وزن 0.001 به آن اختصاص خواهد یافت. مقدار مربوط به خاص بودن مهارت fام در مولفه fام بردار specialty ذخیره خواهد شد. طول این بردار برابر تعداد کل مهارت های موجود(m) است.

علاوه بر این جمع افقی هر سطر مبین تعداد مهارت‌هایی یک خبره است. طبق معیار حداکثر پوشش مورد نظر ما، هر اندازه مقدار مجموع عناصر یک سطر بزرگتر باشد به این معناست که آن داور مهارت‌های بیشتری را پوشش داده و برای انتخاب اولویت بیشتری دارد. بنابر این جمع سطری عناصر ماتریس D نیز در تعیین امتیاز داوران موثر است. با توجه به موارد گفته شده امتیاز خبره D_k که مشخصات او در سطر k ماتریس D موجود است، در هر مرحله با استفاده از فرمول (۲) تعیین می شود.

$$Score(D_k) = \sum_{j=1}^{m} d_{kj} \times Speciality[j] \quad (۲)$$

در این فرمول d_{kj} نشان دهنده مولفه سطر k ستون g ماتریس D بوده و $Speciality[j]$ نشان دهنده مقدار خاص بودن مهارت jام است. برای روشن شدن این مطلب مثال زیر را در نظر بگیرید: فرض کنید ماتریس شکل (۱) با داوران D1 تا D6 و مهارت‌های x، y، z، w، s و m داده شده است. هدف یافتن مجموعه داورانی است که کلیه مهارت‌های x,y,z,w,s,m را پوشش دهد.

	x	y	z	w	s	m
D1	۱	۰	۰	۰	۰	۱
D2	۰	۱	۰	۰	۱	۰
D3	۰	۰	۱	۰	۰	۰
D4	۱	۰	۰	۱	۰	۰
D5	۰	۱	۰	۰	۰	۰
D6	۱	۰	۰	۰	۰	۱

شکل(۱): ماتریس ورودی داوران

در اینصورت با فرض $\alpha = 10^{-n}$ امتیاز نهایی داوران در این مرحله به صورت زیر تعیین خواهد شد:

	D1	D2	D3	D4	D5	D6
خاص بودن مهارت	۰,۰۱	۰,۱	۰,۱	۰,۱	۰,۰۱	۰,۰۱
میزان پوشش	۲	۲	۱	۲	۱	۲
وزن نهایی	۰,۰۲	۰,۲	۰,۱	۰,۲	۰,۰۱	۰,۰۲

شکل(۲): وزن دهی به داوران

پس از وزن‌دهی به مهارت‌های خبرگان، فردی با بیشترین امتیاز انتخاب شده و به مجموعه پاسخ اضافه خواهد شد. پس از یک مرحله انتخاب، امتیاز سایر خبرگان بهنگام سازی می‌شود. به عنوان مثال در شکل – داور D2 به مجموعه پاسخ اضافه می‌شود. سپس از آنجاییکه این داور با داور D5 در مهارت y اشتراک دارند و مهارت y اکنون توسط D2 پوشش داده شده است، امتیاز داور D5 در مرحله بعد صفر خواهد شد.

تکرار مراحل این الگوریتم تا پوشش کامل همه مهارت‌ها ادامه خواهد یافت. از آنجایی‌که ما در ابتدای کار مجموعه مهارت‌هایی را که هیچ داوری آنها را پوشش نمی‌دهد حذف کرده‌ایم، این الگوریتم پایان پذیر خواهد بود. شبه کد مربوط به الگوریتم پیشنهادی در شکل (۳) آمده است.

Greedy Algorithm
Begin
1. Input a file of exports skills and transform it to a 0-1 Matrix
2. Remove Skills that not be in Exports
3. Weight each export based on two measures specialty and number of covered skills.
4. Select an expert with maximum weight
5. Remove the selected expert from matrix
6. Update the weight of remained experts
7. Continue from step 2 until the selected experts cover all task's skills.
8. Output result set
END

شکل(۳): شبه کد سطح بالای الگوریتم پیشنهادی

الگوریتم پیشنهادی در عمل از نظر زمانی بسیار سریع عمل کرده و قادر به یافتن جواب های بهینه است. از طرفی پیچیدگی زمانی این الگوریتم از مرتبه چندجمله‌ای است.

۵- نتایج تجربی

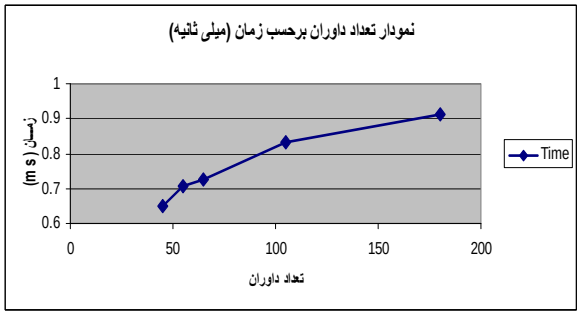
ما برای ارزیابی روش پیشنهادی از مجموعه داده‌ای معروف و استانداردی استفاده می‌کنیم که محققان بسیاری در مقالات خود از آن استفاده کرده‌اند [12, 13, 18, 24]. این مجموعه داده که از کاوش بر روی چکیده مقالات ACM SIGIR در سال‌های ۱۹۷۱ تا ۲۰۰۶ بدست آمده است [11, 12, 13, 17, 18, 19]. شامل ۷۳ مقاله و ۱۸۹ داور (هر یک با حداقل ۲ مهارت) بوده که مهارت‌های هر گروه با بررسی پروفایل نویسندگان و چکیده مقالات، مشخص گردیده است. هر مقاله حداکثر ۲۵ عنوان داشته و داوران نیز حداکثر در ۲۵ مهارت مختلف تبحر دارند.

این مجموعه داده‌ای در زمان نگارش مقاله از آدرس http://timan.cs.uiuc.edu/data/review.html قابل دسترسی است. الگوریتم پیشنهادی نیز با زبان جاوا پیاده سازی شده است. نمودار شکل(۴)

Archive of SID

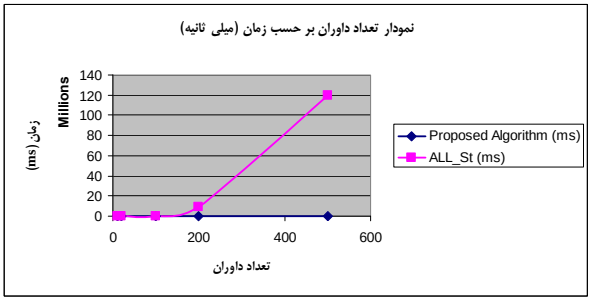
www.SID.ir

خروجی مربوط بـه زمـان یـافتن بهتـرین داوران (بـا رعایـت شـروط بهتـرین پوشش‌دهی و خاص بودن) برای یک مقاله ۲۵ عنوانی با تعداد متغیر داوران را نشان می‌دهد:

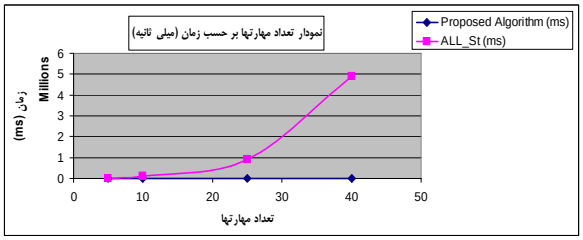


شکل(۴): نمودار زمان یافتن بهترین گروه خبرگان برای کاری با ۲۵ مهارت

همچنین برای تعیین صحت پاسخ‌های یافت شده، یک الگـوریتم دقیـق نیاز پیاده سازی شده است که با تولید تمامی جایگشت‌هـای داوران، مجموعـه داوران کمینه با پوشش کامل مجموعـه مهـارت‌هـای مـورد نیـاز را اسـتخراج می‌کند. با اجرای دو الگوریتم دقیق و الگوریتم پیشنهادی در ایـن مقالـه روی مجموعه داده ورودی، روش پیشنهادی توانست پاسخ‌هایی یکسان با الگـوریتم دقیق را در تمام حالات تست شده به دست آورد. با این تفاوت که زمان اجرای الگوریتم پیشنهادی خیلی کمتـر از الگـوریتم Brute force اسـت. مـا در دو حالت زمان‌های این دو الگوریتم را مورد مقایسه قرار داده‌ایم. شکل شماره (۵) زمان اجرای این دو روش را در شبکه‌هایی با تعداد داوران مختلـف نـشان داده است. در تمامی این شبکه ها هدف یافتن مجموعه داورانی است که مقالـه‌ای با ۲۵ موضوع را پوشش دهند. شکل شماره (۶) زمان اجرای الگوریتم دقیـق و الگوریتم پیـشنهادی را در حـالتی کـه تعـداد داوران موجـود ثابـت امـا تعـداد مهارت‌های موردنیاز متغیر است نشان می دهد.



شکل(۵): مقایسه زمان اجرای الگوریتم پیشنهادی با الگوریتم دقیق بر حسب تعداد داوران



شکل(۶): مقایسه زمان اجرای الگوریتم پیشنهادی با الگوریتم دقیق بر حسب تعداد مهارت ها

در مقالات مختلف از سه معیار میزان پوشش‌دهـی، اطمینـان و متوسـط اطمینان برای تعیین کارایی روش استفاده شده است. این سه مفهوم در ادامـه توضیح داده شده است:

میزان پوشش‌دهی[۵]: اگر مهارت‌های مورد نیاز یک مقاله با n_A مهارت را به صورت $A_1, \ldots, A_{nA}$ نشان دهیم و n_r تعداد مهارت‌های باشد که n داور انتخابی پوشش می‌دهند، میزان پوشش دهی از فرمول (۳) بدست می‌آید [12, 17]:

$$Coverage = \frac{n_r}{n_A} \qquad (۳)$$

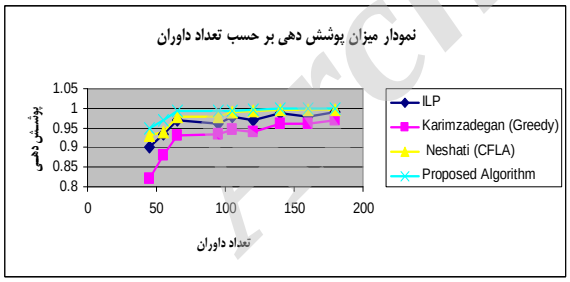
قابلیت اطمینان[۶]: منظور از قابلیـت اطمینـان، محاسـبه میـزان افزونگـی موجود در مهارت‌های داوران است. بدین معنا کـه چـه مقـدار از مهـارت‌هـای داوران انتخابی با یکدیگر مشترک بوده و در داوران مختلف تکراری است. اگر $A'1, \ldots, A'_{nr}$ را مهارت‌های n_r مهارتی در نظر بگیریم کـه توسـط n داور انتخابی پوشش داده می‌شود و $n_{A'i}$ تعداد داورانـی باشـد کـه مهـارت A'_i را شامل می‌شود، مقدار اطمینان از فرمول (۴) بدست می‌آید [12, 17]:

$$Confidence = \frac{\sum_{i=1}^{n_r} \frac{n_{A'_i}}{n}}{n_r} \qquad (۴)$$

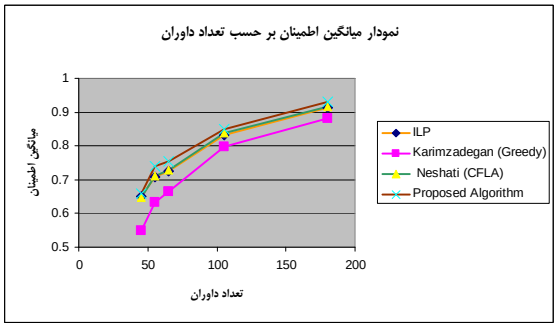
یک انتساب خوب باید هر دو مقدار پوشش‌دهی و قابلیـت اطمینـان بـالا برخوردار باشد. یک راه برای ترکیب پوشش‌دهی و قابلیـت اطمینـان نرمـالیزه کردن قابلیت اطمینان بر اساس تمامی مهارت‌ها و نه صرفاً مهارتهای پوشـش داده شده است. برای این منظور قابلیت اطمینان میانگین تعریف شده است که از طریق فرمول (۵) قابل حصول است [12, 17].

$$Average\ Confidence = \frac{\sum_{i=1}^{n_A} \frac{n_{A'_i}}{n}}{n_A} \qquad (۵)$$

در ادامه مقدار پوشش‌دهـی، اطمینـان و میـانگین اطمینـان پاسـخ‌هـای بازیابی شده توسط الگوریتم پیشنهادی روی مجموعه داده‌هایی با تعداد داوران مختلف محاسبه شده است. علاوه بر این مقدار این سه معیـار بدسـت آمـده از الگوریتم آقای نشاطی [17] و الگوریتم خانم کریم زادگان که دو روش جدیـد در این حوزه به شمار می‌آیند نیز در نمودار شکل های(۷) و(۸) نشان داده شده است. همانطور که در این نمودارها مشخص اسـت، روش پیـشنهادی، بهتـر از هر دو روش قبلی عمل می‌کند.



شکل(۷): میزان پوشش‌دهی چهار الگوریتم مختلف براساس تعداد داوران برای یک مقاله



شکل(۸): میزان متوسط اطمینان چهار الگوریتم مختلف براساس تعداد داوران برای یک مقاله

Archive of SID

www.SID.ir

٦- نتیجه گیری

ما در این مقاله روشی حریصانه برای حل مسئله یافتن گروهی از خبرگان برای انجام کاری با مهارت‌های خاص پیشنهاد داده‌ایم. این روش با در نظر گرفتن دو ویژگی حداکثرسازی میزان پوشش و اولویت‌دهی به داورانی با مهارت‌های خاص قادر به بازیابی نتایجی خیلی نزدیک به الگوریتم دقیق می‌باشد. همچنین نتایج اجرای این روش روی مجموعه دادگان در حالات مختلف مورد آزمایش و بررسی قرار گرفته است. بر اساس نتایج حاصل از این آزمایشات و مقایسه نتایج آن با نتایج حاصل از الگوریتم‌های دیگر، می‌توان گفت الگوریتم پیشنهادی از نظر میزان پوشش و قابلیت اطمینان بهتر از روش‌های دیگر عمل کرده و نتایج مناسبی را در زمان کوتاه بازیابی می‌کند. این الگوریتم از نظر مرتبه زمانی در رده الگوریتم‌های چندجمله‌ای قرار می‌گیرد.

مراجع

[1] Balog, K., Azzopardi, L., & De Rijke, M., A language modeling framework for expert finding, Information Processing and Management, 45(1), 1-19. , 2009

[2] Balog, K., Azzopardi, L., de Rijke, M. Formal models for expert finding in enterprise corpora, Proceedings of 29th International Conference on Research and Development in Information Retrieval, 43-50, 2006

[3] Basu, C., Hirsh, H., Cohen, W., Nevill-Manning, C., Recommending papers by mining the web, Proceeding of 16th International Joint Conferences on Artificial Intelligence, , 1999

[4] Biswas, HK., Hasan, MM., Using publications and domain knowledge to build research profiles: An application in automatic reviewer assignment, Proceedings of International Conference on Information and Communication Technology, p. 82-86, 2007

[5] Charikar, M., & Guha, S., Improved combinatorial algorithm for location and k-median problems, In Processing of the 40th annual Symposium on Foundations of Computer Science, FOCS '99 (378-385), Washington, DC, USA:IEEE Computer Society, 1999

[6] Deng, H., King, I., & Lyu, M. R. Enhanced models for expertise retrieval using community-aware strategies, IEEE Transactions on Systems, Man, and ybernetics, Part B (Cybernetics), 42(1), 93–106 (ISSN 1083-4419) , 2012

[7] Goyal, A., Bonchi, F., Lakshmanan, L.V., Learning influence probabilities in social networks, In: WSDM, 241-250, 2010

[8] Hashemi, S. H., Neshati, M., & Beigy, H., Expertise retrieval in bibliographic network: A topic dominance learning approach, In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM 2013) (in press), 2013

[9] Yin, H., Cui, B., Huang, Y., Finding a Wise Group of Experts in Social Networks, Data Mining And Applications, 1(1), 381-394, 2011

[10] Kargar, M., & An, A, Discovering top-k teams of experts with/without a leader in social networks, In Processing of the 20th ACM international confrance of Information and Knowledge management, CIKM '11 (985-994), New York, NY, USA:ACM (ISBN 978-1-4503-0717-8), 2011

[11] Karimzadehgan, M.; Zhai, C., Constraind multi-aspect expertise matching for committee review assignment, Proceeding of the 18th ACM conference on Information and Knowledge Management, 1697-1700, 2009

[12] Karimzadehgan, M., & Zhai, C., Integer linear programming for constrained multi-aspect committee review assignment, Information Processing and Management, 48(4), 725-740 (ISSN 0306-4573), 2012

[13] Karimzadehgan, M., Zhai, C., Belford, G. Multi-aspect expertise matching for review assignment. Proceedings of ACM 17th Conference on Information and Knowledge Management, 1113-1122, 2008

[14] Lappas, T., Liu, K., & Terzi, E., Finding a team of experts in social networks. In Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '09 (pp. 467–476). New York, NY, USA: ACM (ISBN 978-1-60558-495-9), 2009

[15] Macdonald, C., The Voting Model for People Search, Ph.D. thesis, Department of Computing Science. University of Glasgow, 2009

[16] Mimno, D., McCallum, A., Expertise modeling for matching papers with reviewers. Proceeding of 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 500-509, 2007

[17] Neshati, M., Beigy, H., Hiemstra, D., Expert group formation using facility location analysis, In Processing of the seventeenth Australasian, Information Processing & Management, 50(42), 361-383, 2014

[18] Neshati, M., Asgari, E., Hiemstra, D., & Beigy, H., A joint classification method to integrate scientific and social networks. In Proceedings of the IR research, 35th European conference on Advances in information retrieval, ECIR'2013 (122–133). Berlin, Heidelberg: Springer-Verlag, 2013

[19] Neshati, M., Beigy, H., & Hiemstra, D., Multi-aspect group formation using facility location analysis, In Processing of the seventeenth Australasian, Document Computing Symposium, ADCS '12 (62-71), New York, NY, USA:ACM (ISBN: 978-1-4503-1411-4), 2012

[20] Pal, A., & Konstan, J. A., Expert identification in community question answering: Exploring question selection bias. In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10 (pp. 1505–1508). New York, NY, USA: ACM (ISBN 978-1-4503-0099-5), 2010

[21] Rodriguez, MA., Bollen, J., An algorithm to determine peer-reviewers. Proceedings of ACM 17th Conference on Information and Knowledge Management, 319-328, 2008

[22] Sun YH, Ma J, Fan ZP, Wang J., A hybrid knowledge and model approach for reviewer assignment. Proceedings of Expert Systems with Applications: An International Journal 34:817–824., 2008

[23] Smirnova, E., A model for expert finding in social networks, In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, SIGIR '11 (pp. 1191–1192). New York, NY, USA: ACM (ISBN 978-1-4503-0757-4), 2011

[24] Tang, W., Tang, J., Lei, T., Tan, C., Gao, B., & Li, T., On optimization of expertise matching with various constraints, Neurocomputing, 76(1), 71–83 (ISSN 0925-2312), 2012

[25] Wi, H., Oh, S., Mun, J., & Jung, M., A team formation model based on knowledge and collaboration, Expert Systems with Applications, 36(5), 9121-9134, 2009

زیرنویس‌ها

Archive of SID

www.SID.ir

[1] Task
[2] Author-Persona-Topic
[3] Local Information
[4] Set Cover Problem
[5] Coverage
[6] Confidence

Archive of SID

www.SID.ir