ارائه یک روش جدید برای پیش‌بینی لینک در شبکه‌های اجتماعی براساس الگوریتم‌های فراابتکاری

نویسندگان: یزدان پیلتن- piltan.yazdan@gmail.com – ارشد کامپیوتر موسسه آموزش عالی لیان بوشهر
موسی مجرد –m.mojarad@iauf.ac.ir
حسن ارفعی نیا–msc.harfaeinia@gmail.com

چکیده: امروزه تحلیل شبکه‌های اجتماعی یکی از مهمترین زمینه‌های تحقیقاتی در داده‌کاوی است. هدف از تحلیل این شبکه-های استخراج دانش نهفته در مجموعه داده‌ها و یادگیری رفتار کاربران در محیط شبکه‌های اجتماعی می‌باشد. یکی از جذاب‌ترین و اصلی‌ترین کاربردهای تحلیل شبکه‌های اجتماعی، مسئله پیش‌بینی لینک است. هدف از پیش‌بینی لینک در شبکه‌های اجتماعی، شناسایی اطلاعات گم شده و ناشناخته از کاربران و یا پیش‌بینی ایجاد لینک بین دو کاربر در آینده است. در سال‌های اخیر، الگوریتم‌های هوش مصنوعی مختلفی بعنوان یکی از مهمترین ابزارهای حل مسائل پیش‌بینی لینک و داده‌های حجیم معرفی شده‌اند. در این تحقیق راهکاری مبتنی بر یک الگوریتم فراابتکاری برای بهبود پیش‌بینی لینک در شبکه‌های اجتماعی استفاده شده است. روش پیشنهادی بر اساس خصوصیات شبکه‌های اجتماعی علامتدار ارائه شده و مسئله پیش‌بینی لینک را به یک مسئله دسته‌بندی دوکلاسه تبدیل می‌کند. سپس از قابلیت الگوریتم ازدحام ذرات و ویژگی‌های توپولوژیکی گراف شبکه اجتماعی برای ایجاد یک پایگاه داده با دو کلاس استفاده کرده که کلاس اول به وجود ارتباط بین کاربران و کلاس دوم به عدم وجود این ارتباط اشاره دارد. با ایجاد پایگاه داده از مدل ماشین بردار پشتیبان برای کار دسته‌بندی و از معیار شباهت کلاسیک کاتز برای پیشنهادهای نهایی کاربران استفاده می‌شود. برای مقایسه و ارزیابی روش پیشنهادی از اطلاعات شبکه اجتماعی توییتر بهره گرفته شده است. نتایج آزمایش‌ها برتری ۰/۲۳، ۹۹/ و ۶/۳۲ روش پیشنهادی را به ترتیب نسبت به الگوریتم‌های Meta-Path، Katz و CN در معیار اندازه‌گیری F نشان می‌دهد.

کلید واژه: شبکه‌هایی اجتماعی، پیش‌بینی لینک، الگوریتم‌های فراابتکاری، معیارهای شباهت.

۱- مقدمه

شبکه‌های اجتماعی به عنوان ساختارهای اجتماعی از اشخاص تعریف می‌شوند که بصورت مستقیم یا غیرمستقیم و برمبنای یکسری روابط مثل دوستی، اشتراک فایل، ارسال ایمیل و غیره با هم ارتباط دارند. فیسبوک و توییتر نمونه-های مهمی از این شبکه‌ها هستند. شبکه‌های اجتماعی یک راه محبوب برای مدل‌سازی تعاملات میان افراد در یک گروه یا جامعه می‌باشد. این شبکه‌ها حاصل ترکیبی از مفاهیم علوم مختلف بخصوص علوم اجتماعی و علوم کامپیوتری است[۱]. با گسترش جوامع اجتماعی مجازی، تحلیل شبکه و استخراج جوامع نیز تکامل یافته‌اند.

شبکه‌های اجتماعی می‌توانند به صورت گراف تجسم شوند، جایی که یک گره مربوط به یک فرد در یک گروه است و یک لبه نشان دهنده نوعی ارتباط بین افراد مرتبط است[۲]. شبکه‌های اجتماعی بسیار پویا هستند، زیرا لبه‌ها و گره‌های جدید در طول زمان به گراف اضافه می‌شوند[۳-۵]. درک پویایی‌هایی که تکامل شبکه اجتماعی را به وجود می‌آورند به دلیل تعداد زیادی از پارامترهای متغیر یک مشکل

پیچیده است. اما یک مشکل نسبتاً ساده‌تر این است که ارتباط بین دو گره خاص را درک کنید. به عنوان مثال، برخی از سوالات جالب که می‌توانند مطرح شوند عبارتند از: چگونه الگوهای ارتباطی در طول زمان تغییر می‌کنند؟ عواملی که انجمن را مدیریت می‌کنند چیست؟ ارتباط بین دو گره تحت تاثیر گره‌های دیگر چطور است؟

در واقع مسئله اصلی که در این پژوهش مورد بررسی قرار می‌گیرد، پیش‌بینی احتمال اتصال بین دو گره در آینده است با این فرض که هیچ ارتباطی بین دوگره در وضعیت فعلی گراف وجود ندارد. این مشکل معمولاً به عنوان پیش بینی لینک[۶] شناخته می‌شود. اگر شبکه اجتماعی با یک گراف G=(V,E) که V مجموعه ای از گره‌ها و E مجموعه ای از یال‌های گراف باشد، نمایش دهیم، آنگاه مسئله پیش بینی لینک، شناسایی مجموعه یال E’ است که در حال حاضر در گراف G وجود ندارند ولی ممکن است در یک بازه زمانی در آینده به گراف G اضافه شوند[۷]. این یال‌ها یک گراف پیچیده با حداکثر تعداد یال ممکن را مطابق رابطه (۱) تشکیل می‌دهند.

$$U = \frac{|V| \times (|V| - 1)}{2} \qquad (۱)$$

در چنین چیدمانی، مجموعه یال‌های موجود E و مجموعه یال‌هایی که ممکن است در آینده اضافه شوند (E’) با یکدیگر اشتراکی ندارند ($E \cap E' = \emptyset$). پیش‌بینی لینک، یال‌های مجموعه E’ را با استفاده از خواص چهارچوب گراف در شبکه‌های اجتماعی مشخص می‌کند. همچنین $E + E' = U$ و پیش‌بینی لینک سعی در یافتن مجموعه E’ دارد برای مثال $E' = U - E$.

در ادامه این تحقیق؛ معیارهای شباهت کلاسیک در بخش دوم ارائه می‌شوند. در بخش سوم به بررسی برخی از جدیدترین کارهای انجام شده در زمینه پیش‌بینی لینک در شبکه‌های اجتماعی می‌پردازیم. روش پیشنهادی در بخش چهارم مطرح شده و نتایج حاصل از آن در بخش پنجم ارائه می‌گردد. در نهایت در بخش ششم نتیجه‌گیری و پیشنهادات مطرح می‌شود.

۲- معیارهای شباهت

در این بخش برخی از معروف‌ترین معیارهایی که به طور تجربی در پیش‌بینی لینک، استفاده شده‌اند را مورد بررسی قرار می‌دهیم. از معیارهای شباهت جهت محاسبه احتمال برقراری لینک (ارتباط) در آینده بین دو کاربر استفاده می‌شود. این معیارها مبتنی بر توپولوژی گراف هستند.

۲-۱ معیار کاتز

شاخص کاتز از مجموع تمام مسیرهای بین جفت گره‌ها در گراف G بصورت نمایی توسط طول مسیرهای کوتاه برای محاسبه شباهت استفاده می‌کند. این معیار، شباهت بین گره‌های u و v را با رابطه (۲) محاسبه می‌کند [۸].

$$Score(u,v) = \sum_{l=1}^{\infty} \beta^l . |paths_{u,v}^l| \qquad (۲)$$

در این رابطه، برای کاهش اثر مسیرهای طولانی در تعیین امتیاز یک لینک، از ضریب ثابت β استفاده شده است. استفاده از β^l در این رابطه باعث می‌شود که مقدار محاسبه شده برای پیش‌بینی لینک در مسیرهایی با طول کوتاه‌تر بیشتر شود. $|paths_{u,v}^l|$ نیز تعداد مسیرهایی به طول l بین دو کاربر u و v است.

۲-۲ معیار دوست یک دوست

معیار دوست یک دوست به تعداد دوستانی که در گره‌های u و v مشترک هستند بستگی دارد، که در رابطه (۳) نشان داده شده است[۹].

$$Score(u,v) = \Gamma(u) \cap \Gamma(v) \qquad (۳)$$

در این رابطه $\Gamma(u)$ و $\Gamma(v)$ مجموع همسایه‌های u و v هستند.

۲-۳ معیار همسایه مشترک

شاخص همسایگان مشترک، تعداد همسایگان مشترک بین u و v را اندازه‌گیری می‌کند. ثابت شده که این معیار احتمال وقوع لینک‌های آینده در یک شبکه همکاری با تعداد همسایه مشترک را به خوبی محاسبه می‌کند. رابطه (۴) نحوه محاسبه این معیار را نشان می‌دهد[۱۰].

$$Score(u,v) = |\Gamma(u) \cap \Gamma(v)| \qquad (۴)$$

۲-۴ معیار پیوند دوست

معیار پیوند دوست با توجه به فرکانس و طول مسیر بین دو گره u و v معیار شباهت را اندازه‌گیری می‌کند. رابطه (۵) معیار پیوند دوست را نشان می‌دهد[۱۱].

$$Score(u,v)=\sum_{i=2}^{l}\frac{1}{i-1}.\frac{|paths_{u,v}^{i}|}{\prod_{j=2}^{i}(n-j)} \quad (۵)$$

جایی که n تعداد گره‌ها؛ l حداکثر طول یک مسیر در نظر گرفته شده بین گره‌های u و v می‌باشد. ترم $1/(i-1)$ یک فاکتور تضعیف کننده است که وزن مسیرها را با توجه به طولl تعیین می‌کند. $|path_{vi,vj}^{i}|$ تعداد تمام مسیرهای با طول l بین u و v است. $\sum_{j=2}^{i}(n-j)$ تعداد کل مسیرهای امکان‌پذیر بین u و v در صورتی است که هر گره از گراف با تمام گره‌های دیگر لینک داشته باشد.

۲-۵ معیار جاکارد

شاخص جاکارد، احتمال اینکه یک همسایه از u و v، همسایه هر دو گره u و v باشد را اندازه‌گیری می‌کند. این اندازه‌گیری، روش تعیین خصوصیات محتوای به اشتراک گذاشته شده است و در بازیابی اطلاعات معنی‌دار می‌باشد. رابطه (۶) این معیار را نشان می‌دهد[۱۲].

$$Score(u,v)=\frac{|\Gamma(u)\cap\Gamma(v)|}{|\Gamma(u)\cup\Gamma(v)|} \quad (۵)$$

۳- پیشینه تحقیق

بلیس و همکاران[۱۰]، استراتژی تکامل انطباق ماتریس کوواریانس(CMA-ES) را برای پیش‌بینی پیوندها در شبکه اجتماعی توییتر پیشنهاد کرده‌اند. ماتریس امتیاز با در نظر گرفتن ۱۶ معیار تشابه متفاوت ساخته شده و در آن از روش الگوریتم تکاملی برای یافتن ضرایب بهینه استفاده شده است. آل‌حسن و همکاران[۱۳]، نشان دادند که استفاده از داده‌های خارجی خارج از محدوده توپولوژی گراف می‌تواند به‌طور قابل توجهی نتایج پیش‌بینی لینک در شبکه‌های اجتماعی را بهبود دهد. همچنین آنها از معیارهای شباهت متفاوتی به عنوان ویژگی در یک برنامه یادگیری با نظارت استفاده کردند که در آن مسئله پیش بینی لینک به عنوان یک مسئله دسته‌بندی دودویی مطرح شده است. از آن به بعد، روش طبقه‌بندی با نظارت در پیش بینی‌های لینک محبوبیت بسیاری یافت[۱۴].

دوپا و همکاران [۱۵]، یک الگوریتم یادگیری برای پیش بینی پیوند براساس محدودیت‌های احتمالی پیشنهاد دادند. آنها یادگیری را با مجموعه بردارهای ویژگی محاسبه شده از داده‌های آموزشی انجام داده سپس از این مدل برای پیش بینی لینک‌های آینده استفاده کردند. لی و همکاران[۱۶]، یک روش برای پیش‌بینی لینک در شبکه‌های ناهمگون براساس شبکه عصبی پس انتشار ارائه داده‌اند. آنها در روش خود از ویژگی‌های فرا-مسیر استفاده کرده‌اند. در این تحقیق پس از استخراج این ویژگی‌ها یک مدل پیش‌بینی لینک نظارت شده با استفاده از شبکه‌ی عصبی پس‌انتشار سه لایه‌ای ایجاد شده است.

میبدی و مرادآبادی[۱۷]، یک روش براساس یادگیری آتوماتا برای پیش‌بینی ایجاد لینک بین دو گره در آینده ارائه دادند. مطالعه آنها متمرکز بر روی استراتژی یادگیری آتوماتا در شبکه‌های اجتماعی وزن‌دار بوده است. هدف آنها برآورد وزن هر لینک تستی با استفاده از از اطلاعات سایر وزن‌ها در شبکه است. ترابی و همکاران[۱۸]، یک روش طبقه‌بندی ترکیبی برای پیش‌بینی لینک در شبکه‌های اجتماعی مبتنی بر مکان ارائه داده‌اند. آنها گره‌های شبکه را با توجه به موقعیت خود در ۸ گروه تقسیم کرده و سپس به این نتیجه رسیدند که میانگین گره‌ها در هر گروهی که بالاتر از ۶۰۰ گره باشد، آن گروه تراکم بالایی است.

سریلاتا و ماجولا[۱۹]، یک روش مبتنی بر الگوریتم کرم شب‌تاب برای پیش‌بینی لینک در شبکه‌های اجتماعی ارائه کرده‌اند. آنها روش کاوش‌گرانه کرم‌های شب‌تاب را برای پیش‌بینی لینک بکار بردند. امتیاز شباهت بین جفت گره در انتهای تکرارها محاسبه می‌شود و جفت گره با بالاترین امتیاز احتمال تشکیل لینک در آینده را دارند. در تحقیقی دیگر سان و همکاران[۲۰]، یک رویکرد منظم اجتماعی حاوی اطلاعات شبکه‌های اجتماعی برای ایجاد یک سیستم توصیه‌گر ارائه دادند. در این تحقیق از سوابق امتیازات و دوستی‌های بین کاربران جهت پیش‌بینی لینک‌های از دست رفته در ماتریس کاربر-آیتم استفاده شده است.

استفان و همکاران[۲۱]، یک تحقیق مقایسه‌ای بین روش-های محاسبه شباهت در سیستم پیشنهاد دهنده مبتنی بر فیلترینگ مشترک با حافظه ارائه دادند. این مطالعه به منظور افزایش درک علاقه کاربر از دامنه گسترده به منظور ارائه توصیه‌های بهتر و همچنین برای کاهش داده‌های

اسپارتی می‌باشد. در پژوهشی مشابه اسپرلی و همکاران[۲۲]، به ارائه پژوهش در زمینه سیستم پیشنهاد دهنده به منظور بهبود رویکرد شبکه‌های اجتماعی پرداختند. در این تحقیق، یک سیستم پیشنهاد دهنده برای برنامه های داده بزرگ طراحی شده برای ارائه توصیه های مفیدی در شبکه های اجتماعی آنلاین استفاده می‌شود. تکنیک پیشنهادی، یک رویکرد مشارکتی و کاربر محور است که سوءاستفاده از تعامل بین کاربران و ایجاد محتوای چندرسانه‌ای در یک یا چند شبکه اجتماعی را در یک روش جدید و مؤثر ایجاد می‌کند.

مدحالی و همکاران[۲۳]، برای پیش‌بینی پیوند در شبکه-های اجتماعی، از تجزیه و تحلیل ویژگی‌های گرافیکی(پروفایل) کاربران استفاده کردند. آنها بیان کرده‌اند که به طور کلی سه روش برای حل مسئله پیش‌بینی پیوند وجود دارد: مدل‌های مبتنی بر ویژگی، مدل‌های احتمالاتی بیزی و مدل‌های رابطه‌ای احتمالی. در روش‌های مبتنی بر ویژگی، ویژگی‌های گرافیکی استخراج شده و برای طبقه-بندی مورد استفاده قرار می‌گیرند. معمولاً این ویژگی‌ها به سه گروه مبتنی بر همسایگی، مبتنی بر مسیر و مبتنی بر گره تقسیم‌بندی می‌شوند. این تحقیق بر استخراج اطلاعات توپولوژیکی مبتنی بر ویژگی تأیید می‌کند که در مطالعات قبلی به آن توجه نشده است. در نهایت نتایج مقایسه ویژگی‌های استخراج شده حاکی از مؤثر بودن ضریب خوشه‌ای و کوتاه‌ترین مسیر در پیش‌بینی لینک با معیار درستی ۰/۹۷ است.

آقابزرگی و خیامامشی[۲۴]، یک معیار شباهت جدید برای پیش‌بینی پیوند بر اساس ساختارهای محلی در شبکه‌های اجتماعی پیشنهاد دادند. این معیار شباهت از طریق یک مدل یادگیری با ناظر در دو مرحله ارائه می‌شود. در مرحله اول، یک پایگاه داده بزرگ شامل اطلاعات یالهای شبکه اجتماعی با اندازه‌گیری شباهت رأس‌های مبدأ و مقصد ایجاد می‌شود. این پایگاه داده علاوه براین حاوی اطلاعات ساختاری مرتبط با رأس‌ها نیز می‌باشد. در مرحله دوم یک مدل یادگیری با ناظر بر مبنای پیش‌بینی پیوند ارائه می-شود. دقت این روش Highschool حدود ۹۳٪ گزارش شده است. پروازه و همکاران[۲۵]، از نظریه گراف و مشخصات کاربران برای دوست‌یابی در شبکه های اجتماعی استفاده کردند. در مرحله اول سیستم پیشنهادی، تمام کاربران با توجه به شباهت ساختاری و پروفایل و با استفاده از روش K_Medoids خوشه‌بندی می‌شوند. در مرحله دوم، الگوریتم FriendLink به کاربران هر خوشه اعمال شده و میزان تشابه بین کاربران محاسبه می‌شود. نتایج حاصل از آزمایش‌ها، معیار F را در پیش‌بینی لینک برای شبکه اجتماعی +Google حدود ۰/۳ و برای شبکه اجتماعی Last.fm حدود ۰/۲۵ نشان می‌دهد.

چن و چن[۲۶]، با استفاده از الگوریتم بهینه‌سازی کلونی مورچگان و شاخص شباهت همسایگان مشترک برای پیش بینی لینک‌های گم شده در شبکه اجتماعی استفاده کردند. شرکات و همکاران[۲۷]، الگوریتمی برای پیش‌بینی لینک مبتنی بر بهینه‌سازی کلونی مورچگان(ACOLP) ارائه دادند. در این روش ساختارهای مثلثی با استفاده از بهینه سازی کلونی مورچگان پیش‌بینی شده و سپس زیر ساختارهایی از گراف‌ها که ممکن است در آینده لینک‌ها را ایجاد کنند، پیش‌بینی می‌شود.

جلیلی و همکاران[۲۸]، الگوریتم Meta-Path را برای پیش‌بینی لینک در شبکه‌های اجتماعی چندگانه مبتنی بر مدل‌های دسته‌بندی ارائه دادند. این الگوریتم مبتنی بر متا-مسیر است و از سه روش NB، SVM و KNN برای کار دسته‌بندی استفاده می‌شود. در تعریف متا-مسیر، u و v دو کاربر در شبکه اجتماعی چندگانه می‌باشند. یک متا-مسیر با طول ۲ از u در شبکه نظیر توییتر به v در شبکه نظیر چهارضلعی بصورت $u \xrightarrow{follows} w \xrightarrow{friends} v$ نشان داده شده که برای ارتباط بین u و v در نظر گرفته می‌شود؛ u به دنبال w در توییتر است و w با v در چهارضلعی دوست است. به منظور تعریف ویژگی‌های مرتبط با مسئله پیش-بینی لینک در الگوریتم دسته‌بندی، در اینجا طول و نوع ارتباطات در متا-مسیر بین دو کاربر تعیین می‌شود.

۴- روش پیشنهادی

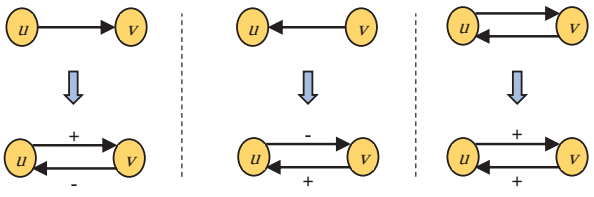
این تحقیق یک روش بهینه جهت پیش‌بینی لینک در شبکه‌های اجتماعی با رویکرد فراابتکاری ارائه می‌دهد. روش پیشنهادی برگرفته از استخراج ویژگی‌های توپولوژیک گراف بر مبنای شبکه‌های اجتماعی خودمحور(مبتنی بر EGO) ارائه می‌شود. شبکه‌های اجتماعی خودمحور نوعی از شبکه‌های اجتماعی می‌باشند که ارتباطات بین کاربران را از طریق شبکه‌هایی با نام EGO سنجیده و به تصویر می‌کشد. در تجسم گراف چنین شبکه‌هایی، کاربر EGO معمولا در مرکز گراف قرار دارد.

روش پیشنهادی براساس خصوصیات شبکه‌های اجتماعی علامتدار ارائه شده و مسئله پیش‌بینی لینک را به یک مسئله دسته‌بندی دوکلاسه تبدیل می‌کند. سپس از قابلیت الگوریتم ازدحام ذرات و ویژگی‌های توپولوژیکی گراف شبکه اجتماعی برای ایجاد یک پایگاه داده با دو کلاس استفاده کرده که کلاس اول به وجود ارتباط بین کاربران و کلاس دوم به عدم وجود این ارتباط اشاره دارد. در اینجا برای کار دسته‌بندی از الگوریتم ماشین بردار پشتیبان (SVM) استفاده می‌شود. ایجاد پایگاه داده بر مبنای هر جفت کاربر انجام می‌شود. به این منظور هر جفت کاربر دارای یک نمونه در پایگاه داده خواهند بود. کلاس این نمونه بر مبنای اطلاعات شبکه اجتماعی تعیین می‌گردد. اگر بین این دو کاربر لینک وجود داشته باشد، کلاس «ارتباط» و در غیراین‌صورت کلاس «عدم ارتباط» تخصیص داده می‌شود. شبکه‌های اجتماعی واقعی، معمولا بسیار پراکنده می‌باشند و این مورد باعث کاهش کارایی الگوریتم‌های دسته‌بندی می‌شود. چون در یک شبکه اجتماعی تعداد کاربران بسیار زیاد است و یک کاربر در مقیاس با تعداد کل کاربران فقط با تعداد اندکی از آنها ارتباط دارد. این حالت موجب بی-نظمی بسیار بالا داده‌ها در پایگاه داده می‌شود. به طور کلی این بی‌نظمی به دلیل افزایش نمونه‌هایی با کلاس «عدم ارتباط» در مقایسه با کلاس «ارتباط» در پایگاه داده می-باشد. این مورد یک چالش و مشکل اساسی برای الگوریتم پیشنهادی است که در اینجا تلاش است تأثیر منفی آن از طریق انتخاب بخشی از نمونه‌های کلاس «عدم ارتباط»، توسط الگوریتم ازدحام ذرات کاهش داده شود. روش فراابتکاری پیشنهادی با ادغام نمونه‌های مشابه و نزدیک به هم از کلاس عدم ارتباط قبل از اعمال الگوریتم ازدحام ذرات بهبود داده می‌شود.

با ایجاد پایگاه داده از مدل ماشین بردار پشتیبان برای کار دسته‌بندی و از معیار شباهت کلاسیک کاتز برای پیشنهادهای نهایی کاربران استفاده می‌شود.

۴-۱ پیش‌پردازش و علامت‌دار کردن شبکه اجتماعی

در این بخش کاربران بخش‌های آموزش(E^T) و آزمایش(E^P) تعیین می‌شوند. در اینجا تقسیم‌بندی کاربران برمبنای اعتبارسنجی 10-Fold انجام می‌شود. در مرحله بعد ساختار شبکه اجتماعی جهتدار با تخصیص علامت اصلاح می‌شود. بین هر دو کاربری که حداقل یکی، دیگری را دنبال کرده باشد، یک علامت مثبت در جهت دنبال‌کننده و علامت منفی در جهت مخالف تخصیص داده می‌شود. شکل۱ چگونگی علامتدار کردن شبکه اجتماعی را نشان می‌دهد.



شکل ۱ : تخصیص علامت به لینک‌های شبکه اجتماعی

۴-۲ استخراج ویژگی

برای ساخت پایگاه داده چندین ویژگی‌های توپولوژیکی از گراف شبکه اجتماعی استخراج می‌شود. در اینجا برای هر دو کاربر u و v در کاربران آموزشی ۶ ویژگی استخراج می-شود.

ویژگی اعتبار برای کاربر u : این ویژگی در [۲۸] ارائه شده و به محبوبیت کاربر u در یک شبکه اجتماعی علامت‌دار مرتبط است. $\boldsymbol{RP(u)}$ ویژگی اعتبار نرمال شده کاربر u است و به صورت رابطه (۶) تعریف می‌شود.

$$\boldsymbol{RP(u)} = \frac{\boldsymbol{d^+_{in}(u) - d^-_{in}(u)}}{\boldsymbol{d^+_{in}(u) + d^-_{in}(u)}} \qquad (۶)$$

در این رابطه، $d_{in}^{+}(u)$ و $d_{in}^{-}(u)$ به ترتیب تعداد لینک-های ورودی با علامت «مثبت» و «منفی» به کاربر u می-باشد.

ویژگی اعتبار برای کاربر v : این ویژگی مشابه ویژگی قبل برای کاربر v محاسبه می‌شود. $RP(v)$ ویژگی اعتبار نرمال شده کاربر v است.

ویژگی خوش‌بینی برای کاربر u : خوش‌بینی یکی دیگر از ویژگی‌های ساختاری از توپولوژی گراف است که در [۲۸] ارائه شده است. $OP(u)$ ویژگی خوش‌بینی نرمال شده کاربر u است و به صورت رابطه (۷) تعریف می‌شود.

$$OP(u) = \frac{d_{out}^{+}(u) - d_{out}^{-}(u)}{d_{out}^{+}(u) + d_{out}^{-}(u)} \qquad (۷)$$

در این رابطه، $d_{out}^{+}(u)$ و $d_{out}^{-}(u)$ به ترتیب تعداد لینک‌های خروجی با علامت «مثبت» و «منفی» از u می-باشد.

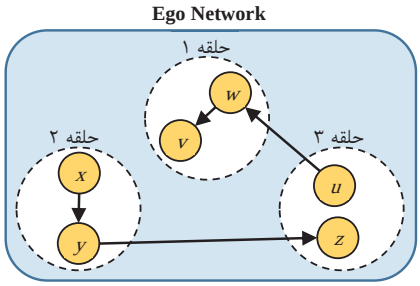
ویژگی خوش‌بینی برای کاربر v : مقدار $OP(v)$ نیز مشابه ویژگی خوش‌بینی برای کاربر u تعریف می‌گردد.

ویژگی توییت‌های مشترک : این ویژگی با توجه به تعداد کلمات کلیدی مشابه استفاده شده در توییت‌های دو کاربر محاسبه می‌شود. رابطه (۸) مقدار این ویژگی را برای دو کاربر u و v تعریف می‌کند.

$$CT(u,v) = |feat_u \cap feat_v| \qquad (۸)$$

در این رابطه، $feat_u$ و $feat_v$ به ترتیب لیست کلمات کلیدی استفاده شده توییت‌های کاربران u و v می‌باشد.

ویژگی حلقه‌های مشترک : در یک شبکه اجتماعی خودمحور کاربران از طریق مسیرهایی که از حلقه‌های مختلف شبکه عبور می‌کنند، با همدیگر ارتباط دارند. ارزش ویژگی حلقه‌های مشترک برای دو کاربر u و v با توجه به تعداد مسیرهای خودمحور با حلقه‌های مشترک بین این دو کاربر محاسبه می‌گردد. مسیر خود محور $u \xrightarrow{follows} w \xrightarrow{friends} v$ با طول ۲ تعریف شده، جائی که کاربر w بین دو کاربر u و v ارتباط برقرار می‌کند. یک مسیر خودمحور می‌تواند با طول‌های متفاوتی ارائه شود، به طوری که کاربر مشابه w باید حداقل با یکی از کاربران u یا v در حلقه مشترکی از یک شبکه EGO باشند. ویژگی حلقه‌های مشترک بین کاربران u و v به صورت $CC(u,v)$ نشان داده شده و برای هر طول مسیر به صورت تعداد کاربران مشابه w تعریف می‌شود که حداقل با یکی از کاربران u یا v در حلقه مشترکی قرار داشته باشد. طول‌های درنظر گرفته شده ۲، ۳ و ۴ می‌باشد. شکل ۲ مثالی از این ویژگی را نشان می‌دهد.



شکل ۲ : مثالی از ویژگی حلقه‌های مشترک

این مثال شامل سه حلقه در یک شبکه EGO و دو مسیر خودمحور با طول ۲ می‌باشد. مسیر اول به صورت $u \rightarrow w \rightarrow v$ است که کاربر w در حلقه مشترکی (حلقه ۱) با کاربر v قرار دارد. مسیر دوم نیز به صورت $x \rightarrow y \rightarrow z$ است که کاربر y بعنوان ارتباط دهنده با کاربر x در حلقه مشترکی (حلقه ۲) هستند.

۴-۳ ایجاد پایگاه داده

ویژگی‌های توپولوژیکی استخراج شده برای ایجاد پایگاه داده با دو کلاس استفاده می‌شود، جایی که کلاس اول به وجود ارتباط بین کاربران و کلاس دوم به عدم وجود این ارتباط اشاره دارد. نمونه‌های پایگاه داده بر مبنای هر جفت کاربر ایجاد می‌شوند. بدین منظور هر جفت کاربر دارای یک نمونه در پایگاه داده خواهند بود. کلاس این نمونه بر مبنای اطلاعات شبکه اجتماعی تعیین می‌گردد. اگر بین این دو کاربر لینک وجود داشته باشد، کلاس «ارتباط» و درغیر این صورت کلاس «عدم ارتباط» تخصیص داده می‌شود.

با در نظر گرفتن N کاربر، برای هر کاربر حدود N^2 نمونه ایجاد خواهد شد. این تعداد نمونه با توجه به حجم بالای تعداد کاربران منجر به ایجاد یک پایگاه داده بزرگ می‌کند. از طرفی تعداد نمونه‌هایی با کلاس «عدم ارتباط» نسبت به کلاس «ارتباط» بسیار زیاد می‌باشد. برای رفع این دو

چالش از فرایند ادغام نمونه‌های مشابه و همچنین کاهش پراکندگی داده‌ها استفاده شده است.

۴-۴ ادغام نمونه‌های مشابه

در پایگاه داده ایجاد شده بسیاری از نمونه‌ها با تشابه بسیار زیاد دیده می‌شوند. از این رو، نمونه‌هایی از پایگاه داده که فاصله‌ای کمتر از مقدار حدآستانه θ دارند، ادغام می‌شوند. $D(i,j)$ فاصله بین دو نمونه i و j را نشان می‌دهد و با رابطه (۹) تعریف می‌گردد.

$$D(i,j)=\frac{1}{M}\sum_{k=1}^{M}|f_k(i)-f_k(j)| \qquad (۹)$$

در این رابطه، $f_k(i)$ و $f_k(j)$ به ترتیب مقدار ویژگی k-ام در نمونه i و j است و M تعداد نمونه‌های استخراج شده است. با توجه به این فاصله، ادغام دو نمونه تنها در صورت مشترک بودن برچسب کلاس آنها امکان‌پذیر است. ادغام دو نمونه بصورت مقدار میانگین برای هر ویژگی انجام می‌شود.

۴-۵ کاهش پراکندگی پایگاه داده

شبکه‌های اجتماعی واقعی، معمولاً بسیار پراکنده می‌باشند و این مورد باعث کاهش کارایی الگوریتم‌های دسته‌بندی می‌شود. چون در یک شبکه اجتماعی تعداد کاربران بسیار زیاد است و یک کاربر در مقیاس با همه کاربران شبکه اجتماعی تنها با تعداد کمی از آنها لینک دارد. برای رفع این مشکل از یک الگوریتم ازدحام ذرات در قالب یک الگوریتم فراابتکاری برای انتخاب و حذف نمونه‌هایی از کلاس «عدم ارتباط» به منظور کاهش پراکندگی داده‌ها استفاده می‌شود. مراحل این الگوریتم به صورت زیر است:

۱. ساختار ذرات : ساختار موقعیت هر ذره در فضای جستجو با توجه به هدف مسئله به صورت شکل ۳ است.

s_1	s_2	⋯	s_i	⋯	s_w

شکل ۳ : ساختار موقعیت ذرات در الگوریتم ازدحام ذرات

در این ساختار s_i بیانگر i-مین نمونه از پایگاه داده در کلاس عدم ارتباط است و به صورت باینری تعریف می‌گردد. w تعداد کل نمونه‌هایی با کلاس عدم ارتباط است. با توجه به این ساختار، هر راه‌حل به صورت یک رشته باینری تعریف شده که ۰ در آن به مفهوم حذف نمونه از پایگاه داده و ۱ بیانگر استفاده از آن نمونه است.

۲. جمعیت اولیه : با توجه به ساختار ذره جمعیت اولیه به صورت تصادفی ایجاد می‌گردد.

۳. تابع هدف : تابع هدف مسئله براساس آنتروپی تعریف می‌شود. آنتروپی در یک پایگاه داده میزان بی‌نظمی (پراکندگی) را نشان می‌دهد. این شاخص برای یک پایگاه داده با دو کلاس، زمانی حداقل (مقدار صفر) می‌شود که نیمی از نمونه‌ها دارای کلاس ۱ و نیمی دیگر دارای کلاس ۲ باشند. از طرفی زمانی که همه نمونه‌ها تنها متعلق به یک کلاس باشند، آنتروپی حداکثر (مقدار یک) می‌شود. بنابراین آنتروپی کم نشان دهنده پراکندگی کمتری در پایگاه داده است و این امر شرایط مناسب‌تری را برای مدل‌های دسته‌بندی مهیا می‌سازد. از اینرو، تابع هدف برمبنای آنتروپی و همچنین دقت مدل دسته‌بندی SVM بصورت رابطه (۱۰) محاسبه می‌شود.

$$F(P_i)=\frac{ACC(S,P_i)}{I(S,P_i)} \qquad (۱۰)$$

در این رابطه $F(P_i)$ برازندگی ذره i-ام، $I(S,P_i)$ و $ACC(S,P_i)$ به ترتیب آنتروپی و دقت دسته‌بندی SVM روی پایگاه داده S با توجه به نمونه‌های انتخاب شده در P_i می‌باشد.

۴. پارامترها : بردار سرعت (V) برای هر ذره مطابق با ابعاد بردار موقعیت در ابتدا به صورت تصادفی تعیین می‌گردد. علاوه بر این، پارامتر pBest به بهترین موقعیت ذره که موفق به رسیدن به آن شده، اشاره دارد و gBest به بهترین موقعیتی اشاره دارد که تاکنون توسط جمعیت ذرات کشف شده است.

۵. به‌روزرسانی سرعت و موقعیت ذرات : در هر تکرار الگوریتم، سرعت و موقعیت هر ذره برمبنای پارامترهای pBest و gBest مطابق روابط (۱۱) و (۱۲) بروز می‌شود.

$$V_{k+1} = V_k + c_1 \times r_1 \times (pBest_k - P_k) + c_2 \times r_2 \times (gBest - P_k) \quad (۱۱)$$

$$P_{k+1} = \langle P_k + V_k \rangle \quad (۱۲)$$

در این روابط، V_k و P_k به ترتیب سرعت و موقعیت فعلی ذره k-ام، c_1 و c_2 پارامترهای ثابت یادگیری و r_1 و r_2 اعداد تصادفی بین [0, 1] هستند. $pBest_k$ بهترین موقعیت ذره k-ام و $gBest$ بهترین موقعیت سراسری در جمعیت است. در این تحقیق به دلیل محیط گسسته از ترم $\langle P_k + V_k \rangle$ برای بروزرسانی موقعیت ذرات استفاده می‌شود. این ترم به صورت عدد گرد برای هر نمونه در نظر گرفته می‌شود. حداقل و حداکثر محدوده سرعت و موقعیت پس از بروزرسانی بررسی می‌شود. حداقل و حداکثر سرعت در محدوده [0, 1] تنظیم شده و بردار موقعیت همواره دارای یکی از دو مقدار ۰ یا ۱ است.

۶. شرط خاتمه : در این تحقیق شرط توقف الگوریتم، تکرار به تعداد نسل ثابت است.

۴-۶ فرایند پیش‌بینی لینک

فرایند پیش‌بینی لینک برای هر کاربر آزمایش به صورت مجزا با توجه به مدل دسته‌بندی انجام می‌شود. برای هر کاربر آزمایش، تعداد Top-K کاربر از بخش آموزش به او پیشنهاد داده می‌شود. فرایند پیش‌بینی لینک در ۵ مرحله انجام می‌شود. ۱- ویژگی‌های کاربر هدف با تمام کاربران آموزش استخراج می‌شود. ۲- یک پایگاه داده کوچک برای کاربر هدف با توجه به ویژگی‌های استخراج شده ایجاد می‌شود. ۳- نمونه‌های مشابه از کلاس عدم ارتباط ادغام می‌شوند. ۴- مدل دسته‌بندی SVM روی پایگاه داده اعمال شده و کاربران آموزشی با کلاس پیش‌بینی شده «ارتباط» انتخاب می‌شوند. ۵- رتبه‌بندی کاربران انتخاب شده با معیار شباهت کاتز انجام شده و Top-K کاربر با بالاترین رتبه به کاربر هدف پیشنهاد می‌گردد.

۴- نتایج و بحث

در این تحقیق آزمایشات گسترده‌ای را برای نشان داده برتری الگوریتم پیشنهادی ارائه شده است. نتایج در مقابل معیارهای شباهت کلاسیک Katz (کاتز)، CN (همسایه مشترک) و همچنین الگوریتم Meta-Path که به تازگی ارائه شده، مورد مقایسه قرار می‌گیرد. برای انجام شبیه‌سازی از نرم افزار MATLAB ورژن 2017a استفاده شده که امکانات و جعبه افزارهای آن اجازه تجزیه و تحلیل دقیق‌تر نتایج روش پیشنهادی را می‌دهد. شبیه‌سازی با استفاده از یک کامپیوتر شخصی با پیکربندی پردازنده اینتل ۵ هسته و فرکانس ۲,۴ گیگاهرتز، حافظه ۱۶ گیگابایت و سیستم عامل ویندوز ۱۰ انجام شده است.

برای حصول اطمینان، در همه مقایسه‌ها از شبکه اجتماعی و پارامترهای یکسانی استفاده شده و به منظور ارائه دقیق‌تر، نتایج به صورت میانگین ۱۰ اجرای متمایز گزارش می‌شوند. علاوه بر این، برای اعتبارسنجی روش 10-Fold استفاده شده است. الگوریتم ازدحام ذرات پیشنهادی با تعداد جمعیت ۳۵ و حداکثر تعداد تکرار ۲۰۰ پیکربندی شده است. سایر پارامترهای استفاده شده در شبیه‌سازی بصورت زیر است:

$$\theta = 0.15, \beta = 0.05, c_1 = 0.6, c_2 = 0.4$$

برای شبیه‌سازی از مجموعه داده شبکه اجتماعی خودمحور توییتر استفاده شده که یک اطلاعات ان جهتدار می‌باشد. این مجموعه داده از وبسایت SNAP قابل دسترسی آزاد است. مجموعه داده توییتر از تعدادی شبکه EGO تشکیل شده که هر شبکه EGO دارای یک کاربر EGO و تعدادی «حلقه» است. حلقه‌ها لیستی از کاربران را نشان می‌دهند که حول کاربر EGO پخش شده‌اند. با توجه به حجیم بودن داده‌های شبکه اجتماعی توییتر و بالا بودن تعداد شبکه‌های EGO، کاربران و ارتباطات بین آنها، در این پژوهش تنها از بخشی از این اطلاعات برای کار مقایسه و ارزیابی استفاده شده است. در اینجا از ۲۰ شبکه EGO ابتدایی با ۲۱۰۱ کاربر و ۳۴۰۲۳ ارتباط استفاده شده است.

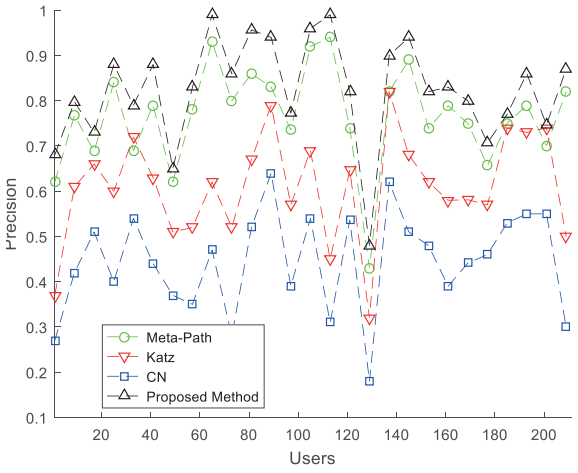
اطلاعات دردست از هر شبکه EGO شامل «لینک‌های جهت‌دار بین کاربران»، «حلقه‌های بین کاربران» و «توییت‌های استفاده شده کاربران» می‌باشد.

در اینجا از معیارهای مختلف و استاندارد Precision (صحت)، Recall (یادآور) و F-Measure (اندازه‌گیری F)، برای ارزیابی و مقایسه نتایج استفاده می‌شود. این معیارها

برای هر کاربر و با توجه به لیست کاربران مرتبط واقعی و لیست کاربران مرتبط پیش‌بینی شده محاسبه می‌شود. معیار Precision نسبت تعداد کاربران مرتبط از لیست Top-K را به تعداد کاربران Top-K محاسبه می‌کند. معیار Recall نسبت تعداد کاربران مرتبط از لیست Top-K را به تعداد به تعداد کل کاربران مرتبط محاسبه می‌کند و معیار F-Measure میانگین هارمونیک نرمال شده معیارهای Precision و Recall است.

مقایسه‌های انجام شده با توجه به دو لیست «کاربران دوست واقعی» و «کاربران دوست پیشنهادی» انجام می‌شود. علاوه بر این، آزمایش‌ها برای هر معیار در دو حالت «ارزیابی برمبنای کاربر» و «ارزیابی برمبنای تعداد پیشنهادات» انجام می‌شود.

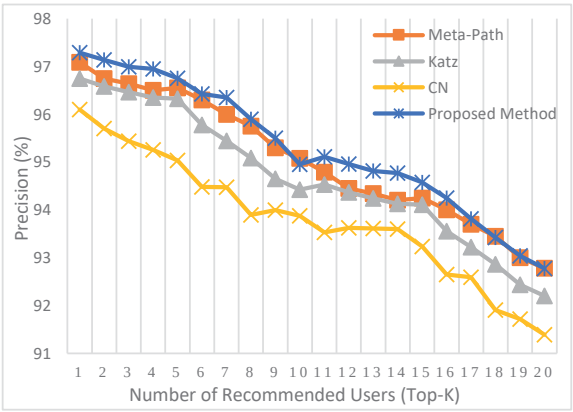
ارزیابی برمبنای کاربر، به این مفهوم است که هر یک از معیارهای ارزیابی براساس اندازه لیست Top-K برای هر یک از کاربران آزمایش محاسبه و ارائه می‌شود. همچنین ارزیابی برمبنای تعداد پیشنهادات، به مفهوم محاسبه هر یک از معیارهای ارزیابی برای تمام کاربران آموزش با اندازه لیست Top-K مشخص و گزارش هر معیار بصورت میانگین است. در اولین آزمایش Top-K را برابر ۱۰ در نظر گرفته و معیار Precision برای هر کاربر آزمایش در شکل ۱ گزارش شده است.



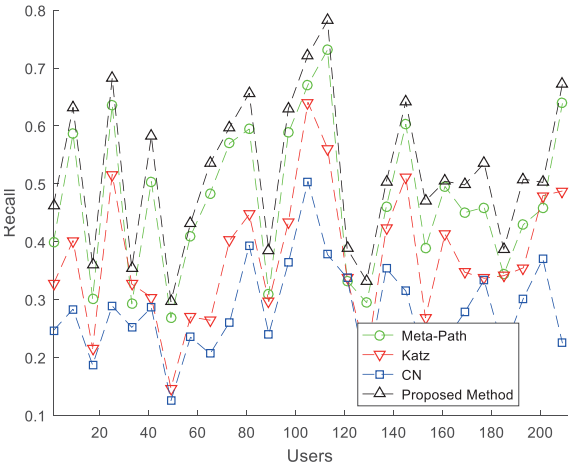
شکل ۱ : نتایج معیار Precision برای هر کاربر

نتایج این مقایسه برتری روش پیشنهادی را برای اغلب کاربران نشان می‌دهد. پس از روش پیشنهادی، الگوریتم‌های Meta-Path، Katz و CN در رتبه‌های بعدی قرار دارند. در آزمایش دیگری مقدار Top-K از ۱ تا ۲۰ بصورت میانگین این معیار برای همه کاربران آزمایش ارائه شده است. اینکار عملکرد بهتری را برای مقایسه از معیار Precision ارائه می‌دهد. شکل ۲ نتایج این مقایسه را نشان می‌دهد.

نتایج برتری محسوس روش پیشنهادی در اغلب Top-Kها را نشان می‌دهد. بجز در تعداد ۱۰، ۱۸، ۱۹ و ۲۰، در سایر موارد روش پیشنهادی از الگوریتم Meta-Path عملکرد بهتری دارد. در بهترین حالت معیار Precision برای روش پیشنهادی ۹۷/۳ درصد است و این دقت با Top-K=1 حاصل شده است. در آزمایش بعدی نتایج مقایسه معیار Recall در شکل ۳ برای هر کاربر آزمایش و Top-K = 10 ارائه شده است.

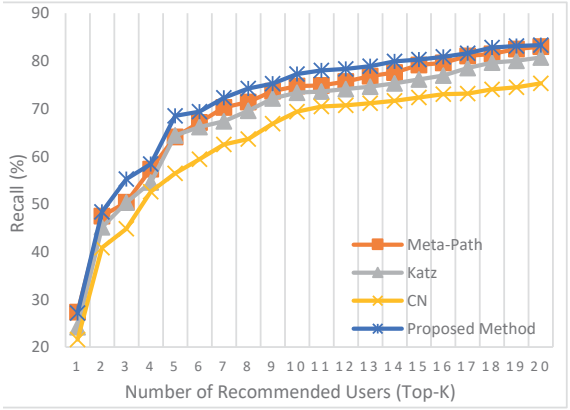


شکل ۲ : نتایج معیار Precision با تعداد کاربران پیشنهادی مختلف



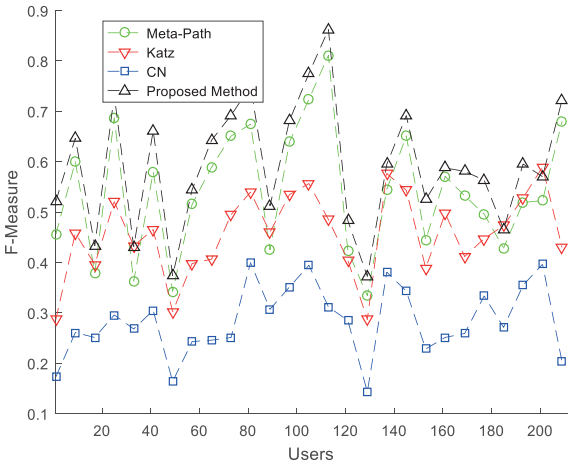
شکل ۳ : نتایج مقایسه معیار Recall برای هر کاربر

مشابه معیار Precision، در این معیار نیز اندازه Top-K بر روی دقت حاصل شده تأثیر دارد. از اینرو نتایج این معیار برای روش‌های مختلف و Top-Kهای ۱ تا ۲۰ مورد بررسی قرار داده شده است. شکل ۴ نتایج این مقایسه را نشان می‌دهد.



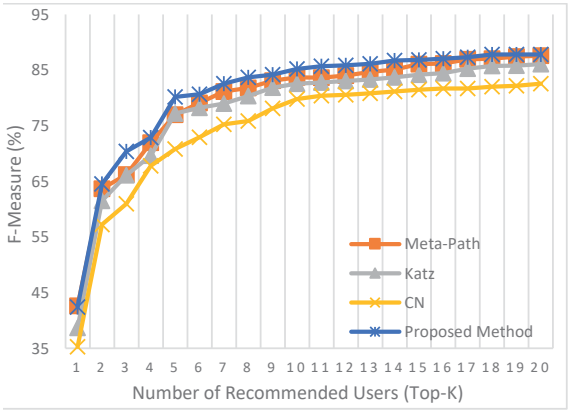
شکل ۴ : نتایج معیار Recall با تعداد کاربران پیشنهادی مختلف

نتایج این مقایسه نشان می‌دهد که در همه تعداد مختلف از Top-K به جزء ۲ و ۲۰، روش پیشنهادی نسبت به همه روش‌ها برتر است. پس از روش پیشنهادی، الگوریتم‌های Meta-Path، Katz و CN به ترتیب در رتبه‌های بعدی قرار دارند. در حالت میانگین معیار Recall در روش پیشنهادی با تعداد کاربران پیشنهادی ۲۰ برابر ۸۳/۳۱ درصد است. این در حالی است که در بهترین حالت گاهی دقت روش پیشنهادی به حدود ۸۸ درصد نیز می‌رسد. در نهایت شکل ۵ نتایج مقایسه معیار F-Measure را برای روش پیشنهادی و روش‌های Meta-Path، Katz و CN نشان می‌دهد. این آزمایش برای هر کاربر آزمایش و براساس Top-K = 10 انجام شده است.



شکل ۵ : نتایج معیار F-Measure برای هر کاربر

در آزمایش دیگری به منظور بررسی دقیق‌تر، نتایج معیار F-Measure برای تعداد کاربران پیشنهاد شده مختلف بررسی شده است. در اینجا تعداد تعداد کاربران پیشنهاد شده از ۱ تا ۲۰ در شکل ۶ مورد بررسی قرار گرفته است.



شکل ۶ : نتایج معیار F-Measure با تعداد کاربران پیشنهادی مختلف

نتایج این مقایسه برتری روش پیشنهادی را به وضوح نشان می‌دهد. بهترین عملکرد روش پیشنهادی در این معیار با Top-K = 18 و دقت ۸۷/۷۹ درصد حاصل شده است. پس از روش پیشنهادی Meta-Path با Top-K = 20 و دقت ۸۷/۵۹ درصد در رتبه دوم قرار دارد. الگوریتم‌های Katz و CN نیز به ترتیب با ۸۶/۰۷ و ۸۲/۵۷ درصد در رتبه‌های بعدی قرار دارند. به طور کلی نتایج برتری روش پیشنهادی را در اغلب آزمایش‌ها و به ازای برخی معیارها نشان می‌دهد. در سایر موارد نیز روش پیشنهادی نتایج قابل قبولی را گزارش داده است.

نتایج مقایسه‌ها نشان می‌دهد که در اغلب مواقع دو روش CN و Katz عملکرد ضعیف‌تری در مقابل روش‌های پیشنهادی و MetaPath دارند. دلیل این امر خصوصیت این الگوریتم است. در CN فقط به تعداد همسایگان و در Katz به فرکانس و طول مسیر توجه می‌شود و اهمیتی به ویژگی‌های شبکه‌های اجتماعی خودمحور داده نمی‌شود. می‌توان نتیجه گرفت که این دو الگوریتم که تنها بر اساس خصوصیت گراف محلی کار می‌کنند چندان برای شبکه‌های اجتماعی خودمحور مناسب نیستند.

۵- نتیجه‌گیری

در این تحقیق یک روش ترکیبی مبتنی بر الگوریتم‌های فراابتکاری برای بهبود مسئله پیش‌بینی لینک در شبکه اجتماعی توییتر ارائه شد. الگوریتم ازدحام ذرات برای ایجاد یک پایگاه داده دو کلاسه از ویژگی‌های توپولوژیکی گراف شبکه اجتماعی و الگوریتم دسته‌بندی ماشین بردار پشتیبان برای مدل کردن این پایگاه داده استفاده شد. به طور کلی استفاده از ویژگی‌های محلی و توپولوژیکی مختلف استخراج شده یکی از عوامل برتری روش پیشنهادی است. از دیگر عوامل این برتری می‌توان به ساختار الگوریتم ازدحام ذرات در کاهش پراکندگی داده‌های پایگاه داده اشاره کرد. همچنین در نظر گرفتن توییت‌های مشترک کاربران در افزایش دقت روش پیشنهادی مؤثر بوده است. به طور کلی، برتری روش پیشنهادی در معیار F-Measure و در مقابل روش‌های Meta-Path، Katz و CN به ترتیب ۰/۲۳، ۱/۹۹ و ۶/۳۲ درصد می‌باشد.

مراجع

[1] Acquisti, A., & Gross, R. (2009). Predicting social security numbers from public data. Proceedings of the National academy of sciences, 106(27), 10975-10980.
[2] Barabâsi, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. Physica A: Statistical mechanics and its applications, 311(3), 590-614.
[3] Papadimitriou, A., Symeonidis, P., & Manolopoulos, Y. (2012). Fast and accurate link prediction in social networking systems. Journal of Systems and Software, 85(9) 2119-2132.
[4] Jalili, M., Orouskhani, Y., Asgari, M., Alipourfard, N., & Perc, M. (2017). Link prediction in multiplex online social networks. Royal Society Open Science, 4(2), 160-863.
[5] Szczepański, P. L., Barcz, A. S., Michalak, T. P., & Rahwan, T. (2015). The game-theoretic interaction index on social networks with applications to link prediction and community detection. In Twenty-Fourth International Joint Conference on Artificial Intelligence.
[6] Liu, F., Liu, B., Sun, C., Liu, M., & Wang, X. (2015). Deep belief network-based approaches for link prediction in signed social networks. Entropy, 17(4), 2140-2169.

[7] Katz, L. (1953). A new status index derived from sociometric analysis. Psychometrika, 18(1), 39-43.
[8] Al Hasan, M. and Mohammed J. Z. (2011). A survey of link prediction in social networks. In Social network data analytics, Springer US. 243-275.
[9] Bliss, C. A., Frank, M. R., Danforth, C. M., Dodds, P. S. (2014). An evolutionary algorithm approach to link prediction in dynamicsocial networks. Journal of Computational Science, 5(5), 750-764.
[10] Papadimitriou, A., Panagiotis s., & Yannis M. (2014). Fast and accurate link prediction in social networking systems. Journal of Systems and Software, 85(9), 2119-2132.
[11] Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013, March). Using of Jaccard coefficient for keywords similarity. In Proceedings of the International MultiConference of Engineers and Computer Scientists, 1(6).
[12] Al Hasan, M., Chaoji, V., Salem, S., & Zaki, M. (2006, April). Link prediction using supervised learning. In SDM06: workshop on link analysis, counter-terrorism and security.
[13] Bilgic, M., Namata, G. M., & Getoor, L. (2007, October). Combining collective classification and link prediction. In Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007) (pp. 381-386). IEEE.
[14] Doppa, J. R., Yu, J., Tadepalli, P., & Getoor, L. (2010, September). Learning algorithms for link prediction based on chance constraints. In Joint European conference on machine learning and knowledge discovery in databases (pp. 344-360). Springer, Berlin, Heidelberg.
[15] Li, J. C., Zhao, D. L., Ge, B. F., Yang, K. W., & Chen, Y. W. (2018). A link prediction method for heterogeneous networks based on BP neural network. Physica A: Statistical Mechanics and its Applications, 495, 1-17.
[16] Moradabadi, B., & Meybodi, M. R. (2018). Link prediction in weighted social networks using learning automata. Engineering Applications of Artificial Intelligence, 70, 16-24.
[17] Torabi, N., Shakibian, H., & Charkari, N. M. (2016, May). An ensemble classifier for link prediction in location based social network. In 2016 24th Iranian Conference on Electrical Engineering (ICEE) (pp. 529-532). IEEE.
[18] Srilatha, P., & Manjula, R. (2017, August). Structural similarity based link prediction in social networks using firefly algorithm. In 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon) (pp. 560-564). IEEE.
[19] Sun, Z., Han, L., Huang, W., Wang, X., Zeng, X., Wang, M., & Yan, H. (2015). Recommender systems based on social networks. Journal of Systems and Software, 99, 109-119.

[20] Stephen, S. C., Xie, H., & Rai, S. (2017, July). Measures of Similarity in Memory-Based Collaborative Filtering Recommender System: A Comparison. In Proceedings of the 4th Multidisciplinary International Social Networks Conference on ZZZ (p. 32). ACM.

[21] Sperlì, G., Amato, F., Mercorio, F., Mezzanzanica, M., Moscato, V., & Picariello, A. (2018). A Social Media Recommender System. International Journal of Multimedia Data Engineering and Management (IJMDEM), 9(1), 36-50.
[22] Madahali, L., Najjar, L., & Hall, M. (2019, March). Exploratory Factor Analysis of Graphical Features for Link Prediction in Social Networks. In International Workshop on Complex Networks (pp. 17-31). Springer, Cham.
[23] Aghabozorgi, F., & Khayyambashi, M. R. (2018). A new similarity measure for link prediction based on local structures in social networks. Physica A: Statistical Mechanics and its Applications, 501, 12-23.
[24] Parvazeh, F., Harounabadi, A., & Naizari, M. A. (2016). A Recommender System for Making Friendship in Social Networks Using Graph Theory and users profile. Journal of Current Research in Science, (1), 535.

[25] Chen, B., & Chen, L. (2014). A link prediction algorithm based on ant colony optimization. Applied Intelligence, 41(3), 694-708.

[26] Sherkat, E., Maseud R., and Masoud A.(2015) "Structural link prediction based on ant colony approach in social networks." Physica A: Statistical Mechanics and its Applications 419 : 80-94.