نوزدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران
دانشکده مهندسی کامپیوتر، دانشگاه شهید بهشتی، تهران، ۱۵-۱۳ اسفندماه ۱۳۹۲

پیش‌گویی نوع پیوند در شبکه‌های اجتماعی با استفاده از روش یادگیری با نظارت

زهرا زمانی علویجه[۱]، احمد براآنی دستجردی[۲]

[۱] دانشجوی کارشناسی ارشد، گروه کامپیوتر، دانشکده فنی مهندسی، دانشگاه اصفهان
z.zamani@eng.ui.ac.ir

[۲] دانشیار، گروه کامپیوتر، دانشکده فنی مهندسی، دانشگاه اصفهان
ahmadb@eng.ui.ac.ir

چکیده

تحلیل شبکه‌های اجتماعی در سال‌های اخیر بسیار مورد توجه قرار گرفته‌است. بسیاری از شبکه‌های اجتماعی به شکل شبکه‌های پیچیده و ناهمگن مدل می‌شوند. ظهور شبکه‌های ناهمگن در فضای وب، تنوع در انواع موجودیت‌های شبکه و پیوندهای مختلف میان موجودیت‌ها را به همراه داشته‌است. بنابراین چالش جدیدی به نام پیش‌گویی نوع پیوند[۱]، مطرح گردیده‌است که هدف آن تشخیص نوع رابطه‌ی موجود، میان موجودیت‌هاست. با توجه به تاثیری که این موضوع در مدیریت بهینه‌ی اطلاعات کاربران دارد، در این مقاله ابتدا به بیان اهمیت پیش‌گویی نوع پیوند در شبکه‌های اجتماعی و مزایای آن پرداخته‌ایم. سپس برای حل این مسئله، الگوریتم جدیدی ارائه شده که براساس روش یادگیری با نظارت و قوانین احتمال، به منظور پیش‌بینی نوع پیوند در یک شبکه اجتماعی عمل می‌کند. این الگوریتم با کلاس‌بندی[۲] نوع روابط براساس ویژگی‌های کاربران درگیر در یک رابطه، سعی در پیش‌بینی نوع روابط ناشناخته دارد. این روش به کمک مجموعه داده Epinions[۳] مورد ارزیابی قرار گرفته‌است ونتایج آن از نظر دقت و فراخوان با الگوریتم کلاس‌بندی بیز مقایسه شده‌است.

کلمات کلیدی

پیش‌گویی نوع پیوند، تحلیل شبکه‌های اجتماعی، یادگیری بانظارت

۱- مقدمه

امروزه کاربران وب علاقه‌ی زیادی به پیوستن به سایت‌های شبکه‌ی اجتماعی و استفاده از آن‌ها، به عنوان یک ابزار ارتباط جمعی، دارند. از جمله‌ی این سایت‌ها می‌توان به فیس بوک، گوگل پلاس، تویتر و غیره اشاره کرد. این سایت‌ها به یکی از رسانه‌های پر نفوذ به خصوص در میان جوانان تبدیل شده‌اند و بخش مهمی از فعالیت‌های روی خط وب را تشکیل می‌دهند. اطلاعات موجود در این سایت‌ها اصولاً افراد و تعاملات آن‌ها و حتی علاقه‌ها و عقایدشان را بر اساس داده‌های انتشار یافته از سوی فرد، توصیف می‌نماید. در حالی که تعداد کاربران شبکه‌های اجتماعی روز به روز درحال افزایش است، نیاز روز افزونی در برخورد با چالش‌هایی هم‌چون، تسهیل ارتباطات میان کاربران، ساخت یک سیستم مدیریت اطلاعات شخصی برای سازمان‌دهی بهتر داده‌ها، ذخیره و بازیابی و تحلیل تعاملات اجتماعی برای مدیریت بهتر روابط، احساس می‌شود[۱]. از طرف دیگر یک سرویس شبکه اجتماعی می‌تواند از دو طریق کاربردهای شبکه و جامعه‌ی کاربری آن بر رقبا پیروز شود. هر جامعه‌ای که خدمات بهتری ارائه دهد، موفق‌تر است.

مدیریت اطلاعات تماس کاربران که از آن با عنوان مدیریت روابط یاد می‌شود، یکی از جنبه‌های مهم در سایت‌های شبکه‌ی اجتماعی است و این سایت‌ها در جهت بهبود و تسهیل آن تلاش بسیاری می‌کنند[۲]. یکی از فعالیت‌هایی که در زمینه‌ی مدیریت روابط مطرح است و اخیراً جایگاه ویژه‌ای میان محققان پیدا کرده‌است، تعیین خودکار نوع روابط اجتماعی میان یک کاربر و افراد در ارتباط با اوست. با توجه با این‌که کاربران شبکه‌های اجتماعی با انواع روابط مختلف مانند دوستی، خویشاوندی، هم‌کلاسی وغیره می‌توانند به یکدیگر متصل شوند، تعیین نوع رابطه‌ی اجتماعی به عنوان یکی از جذاب‌ترین موضوعات مورد بررسی در این زمینه، مطرح شده است.

تعیین نوع روابط در شبکه‌های اجتماعی می‌تواند اطلاعات مفیدی در اختیار قرار دهد که برای بهبود مدیریت اطلاعات کاربران بسیار موثر است. این روش کمک می‌کند تا اشتراک‌گذاری یا فیلترگذاری مطالب انتشار یافته از سوی کاربر، تسهیل یابد و در بهبود حفظ حریم شخصی کاربران، تأثیر به سزایی دارد. همچنین با تعیین نوع روابط میان کاربران، می‌توان سیستم‌های پیشنهاد دوست دقیق‌تری داشت.

اکثر مطالعات انجام شده در زمینه‌ی تعاملات اجتماعی، روی مسئله‌ی وجود پیوند میان دو موجودیت تمرکز کرده‌اند[۳]. تحقیقات محدودی در خصوص تشخیص نوع پیوند موجود، میان دو موجودیت‌های به هم متصل، انجام شده‌است. ماتسو و همکاران در[۴] ، یک سیستم کاوش شبکه‌ی اجتماعی به نام Polyphonet را ارائه کرده‌اند. این سیستم می‌تواند چهار نوع رابطه‌ی هم‌نویسندگی[۴]، هم‌آزمایشگاهی[۵]، هم‌پروژه ای[۶] و هم‌کنفرانسی[۷] را استخراج نماید.

در[۵]، مسئله‌ی استنتاج روابط اجتماعی در شبکه‌های بزرگ مورد بررسی قرار گرفته‌است. به منظور تشخیص نوع رابطه‌ی اجتماعی یک مدل گراف نیمه برچسب‌گذاری شده، معرفی شده‌است. این گراف نشان دهنده روابط اجتماعی میان افراد است که نوع بعضی از این روابط، به عنوان برچسب گراف، مشخص شده‌است. هدف تشخیص نوع روابطی است که نامعلوم هستند. در این کار روی کاوش روابط معنایی تمرکز شده‌است. برای تعیین نوع رابطه‌ی آن‌ها، اطلاعات پروفایل کاربر، اطلاعات پیوندها و شرایط عمومی شبکه در نظر گرفته شده‌اند. وانگ و سان یک مدل گراف احتمالی وابسته به زمان در نظر گرفته‌اند که آن را TPFG نامیده اند[۶]. هدف این الگوریتم کشف روابط دانشجو-استاد راهنما در مجموعه‌داده DBLP است. برای حل مسئله تابع هدفی براساس شباهت تعریف شده‌است. در ابتدا یک گراف همگن تشکیل شده است که یک یال بین نویسندگان مقاله و مقاله انتشار یافته رسم شده‌است. سپس براساس آن تابع شباهت و محاسبه هم‌بستگی میان دونویسنده یک مقاله، تصمیم‌گیری می‌نماید که آیا بین آن‌ها رابطه‌ی دانشجو- استاد راهنما برقرار است.

هدری و همکاران یک روش برای کلاس‌بندی شبکه‌های اجتماعی براساس نوع پیوند ارائه داده‌اند. در این الگوریتم از نوع پیوند میان گره‌ها برای کلاس‌بندی شبکه استفاده شده است. این روش که مبتنی بر احتمالات بیز است، و در ادامه مقاله، با افزودن یک ضریب و در نظر گرفتن اهمیت برخی از پیوندها نسبت به بقیه، این احتمال را تصحیح کرده‌اند. برای ارزیابی الگوریتم از مجموعه داده‌ی IMDB استفاده کرده‌اند[۷].

اخیراً یک روش خوشه‌بندی برای پیش‌بینی علامت[۸] در شبکه‌های با پیوندهای مثبت و منفی، توسط آقایان جواری و جلیلی ارائه شده است[۸]. در این روش ابتدا شبکه را به تعدادی خوشه تقسیم می‌کنند و سپس روش فیلترگذاری مشارکتی را روی آن اعمال می‌کنند. سعی بر این است که پیوندهای منفی میان اعضای یک خوشه و پیوندهای مثبت میان دو خوشه مختلف، حداقل باشند. دقت حاصل از این روش در بهترین حالت، بسته به تعداد خوشه‌ها بین۸۲ تا ۸۵ درصد متغیر است. البته این روش تنها برای شبکه‌هایی با پیوند مثبت و منفی قابل اعمال است و برای یک شبکه با چندین نوع پیوند مفید نیست.

در برخی دیگر از مقالاتی که در این زمینه منتشر شده‌اند، نوع رابطه‌ی میان افراد را در یک مجموعه از تصاویر تخمین می‌زنند. از مجموعه تصاویر خانوادگی، یک درخت رابطه ساخته می‌شود و نوع رابطه‌ی موجود مانند خانوادگی، دوستی و غیره استنتاج می‌شود. روش ساخت درخت، مبتنی بر الگوریتم تشخیص چهره، الگوریتم خوشه‌بندی و یک الگوریتم سنجش شباهت میان دو چهره است. در این تحقیق از جنسیت و سن افراد و موقعیت آن‌ها در تصویر برای سنجش نوع رابطه استفاده می‌شود[۹].

برای این‌که بتوان از مزایای وجود نوع پیوند در یک شبکه بهره برد، نیاز به سیستمی است که به طور خودکار با توجه به پیوندهای دارای نوع، بتواند نوع روابط ناشناخته را پیش‌بینی نماید. هر چند در اکثر موارد از الگوریتم‌های پایه کلاس بندی مانند درخت تصمیم یا الگوریتم بیز برای حل این‌گونه مسائل بهره می‌برند اما این الگوریتم‌ها برای این‌گونه مسائل دقت کافی ندارند. در این مقاله الگوریتمی مبتنی بر احتمالات، برای پیش‌بینی نوع پیوند برای رابطه‌هایی که نوع آن‌ها مشخص نیست، ارائه شده است. این الگوریتم از ویژگی‌های گره‌ها[۹] و ویژگی‌های ساختار شبکه برای پیش‌بینی دقیق بهره برده است. نتایج حاصل از آن با الگوریتم بیز و الگوریتم[۷] مقایسه شده و نشان داده‌شده که الگوریتم ارائه شده، دقت بهتری در پیش‌گویی نوع پیوند، نسبت به دو روش دیگر دارد.

۲- اهمیت و بیان مسئله

تعیین نوع روابط در شبکه‌های مختلف می‌تواند اطلاعات مفیدی در اختیار قرار دهد که برای بهبود عملکرد دیگر قسمت‌ها قطعاً موثر است. به عنوان مثال در شبکه‌ی صفحات وب، اگر نوع رابطه‌ی میان صفحات وب را بتوان تعیین کرد و دانستن این‌که پیوند برقرار شده میان دو صفحه یک پیوند تبلیغاتی است یا یک پیوند مفید محتوایی، کمک شایانی در بازیابی اطلاعات به موتورهای جستجوگر می‌کند[۲]. در سیستم‌های پیشنهاد دهنده که در تجارت الکترونیک به طور گسترده استفاده می‌شوند، اگر بتوان برچسب رتبه بندی کالای جدید توسط کاربر را بر اساس رتبه بندی های قبلی کاربر، پیش‌بینی نمود، می‌توان پیشنهادهای مفیدی برای کاربران داشت.

تعیین نوع رابطه، در شبکه‌های اجتماعی نیز می‌تواند مزایای زیادی به همراه داشته باشد که عبارتند از:

- **مدیریت لیست تماس**: زمانی که کاربر با تعداد زیادی از افراد تعامل دارد و یک لیست طولانی از ارتباطاتش نگهداری می‌کند، مدیریت بهینه‌ی این لیست امری طاقت فرسا و خسته‌کننده است. یک راه حل برای این چالش، در تعیین خودکار نوع روابط اجتماعی نهفته است. در واقع سازمان‌دهی روابط بر اساس نوع رابطه، می‌تواند در بسیاری از موارد از جمله به اشتراک گذاشن مطالب یا فیلتر گذاری بعضی از محتویات مفید باشد[۲].

- **بهبود اشتراک گذاری مطالب:** هر چند امروزه شبکه‌های اجتماعی به کاربران اجازه می‌دهند که تنظیماتی شخصی برای کنترل و مدیریت اطلاعات پروفایل و مطالبی که به اشتراک می‌گذارند، داشته‌باشند اما اصولاً این کار با استفاده از گروه‌بندی لیست تماس کاربر امکان‌پذیر می‌باشدولی این کار برای بسیاری از کاربران خسته کننده و وقت‌گیر است. بنابراین اگر نوع روابط اجتماعی به صورت خودکار تعیین شوند و گروه‌ها را براساس نوع رابطه سازمان‌دهی شوند، در این صورت گام بزرگی در این راستا برداشته‌شده‌است. به این ترتیب کاربر برای مثال می‌تواند بعضی از عکس‌ها را برای گروه همکاران فیلتر کند و یا برخی مطالب را تنها با هم‌کلاسی‌های خود به اشتراک گذارد.

- **حفظ حریم شخصی**: یکی از مهمترین دغدغه‌های کاربران شبکه‌های اجتماعی، حفظ حریم شخصی است. افراد مختلف موجود در لیست تماس یک کاربر می‌توانند از طریق توضیحاتی[۱۰] که روی صفحه‌ی او ارسال می‌کنند با یکدیگر به تبادل اطلاعات بپردازند. یک کاربر ممکن است تمایل داشته باشد که توضیحاتی که دوستانش برای او می‌نویسند توسط خویشاوندانش قابل رؤیت نباشد و یا حتی لیست دوستانش از دید خویشاوندانش مخفی باشند. با دسته بندی روابط بر اساس نوع رابطه و تعیین یک سری گزینه در جهت انجام این تنظیمات می‌توان به این مهم دست یافت[۹].

- **افزایش دقت سیستم‌های پیشنهاد پیوند**: امروزه صدها سرویس شبکه‌ی اجتماعی، مانند سرویس‌های پیشنهاد دوست، به کاربران خود کمک می‌کنند که روابط جدید ایجاد کنند و دوستان بیشتری را به لیست خود اضافه نماید[۳]. با استفاده از الگوریتم تعیین نوع پیوند می‌توان عملکرد سیستم‌های پیشنهاد پیوند را بهبود بخشید.

سایت‌های اجتماعی محبوبی مانند فیس‌بوک این امکان را برای کاربران فراهم می‌کنند که نوع رابطه‌اشان را با افراد موجود در لیست تماسشان تعین نمایند. البته این یک مورد اختیاری است و برخی کاربران از آن بهره نمی‌برند. سایت گوگل پلاس جدیداً امکان ایجاد ارتباط جدید را به کاربرانش، تنها در صورتی می‌دهد که نوع آن رابطه را مشخص نمایند. این قضیه اهمیت مسئله نوع پیوند را در راستای مدیریت بهتر شبکه مشخص می‌کند. با توجه به این‌که این سایت اجتماعی قبلاً تأکیدی بر تعیین نوع رابطه نداشت، بنابراین شبکه اجتماعی این سایت شامل روابطی است که نوع برخی از آن‌ها مشخص است و نوع برخی دیگر خیر.

در این مقاله با فرض در دست داشتن یک شبکه نیمه برچسب‌گذاری شده، سعی در ارائه راه‌حلی مبتنی بر یادگیری با نظارت شده است که بتواند نوع روابط نامشخص را پیش بینی نماید.

۳- راه‌حل پیشنهادی

الگوریتم ارائه شده دارای دو مرحله است که مرحله اول شامل محاسبه احتمالات می‌باشد که بسیار شبیه روش بیز است و مرحله دوم به منظور حل مسئله تضاد گره‌ها در پیش‌بینی نوع پیوند میان آن‌هاست.

۳-۱ مرحله اول الگوریتم

فرض کنید یک شبکه‌ی اجتماعی با تعداد زیادی گره و روابط میان آن‌ها موجود است. هر رابطه میان یک جفت گره برقرار است و هر رابطه دارای برچسب "نوع رابطه" است. این برچسب برای برخی از پیوندها مشخص و برای برخی نامعین است. تعداد m نوع رابطه به صورت مجموعه $R=\{r_1,r_2,\ldots,r_m\}$ در این شبکه موجود است. هر گره در شبکه دارای مجموعه ویژگی‌های $A=\{a_1,a_2,\ldots,a_n\}$ است.

برای محاسبه نوع یک رابطه، با توجه به درگیر بودن یک جفت گره در این رابطه، با استفاده از فرمول (۱)، قانون بیز ، احتمال این‌که نوع این رابطه هر یک از مقادیر R باشد، براساس ویژگی های هر گره محاسبه می‌شود.

$$P(r_i \mid a_1,a_2,\ldots,a_n) = \frac{P(a_1,a_2,\ldots,a_n \mid r_i)\,P(r_i)}{P(a_1,a_2,\ldots,a_n)} \quad (۱)$$

در اینجا، $P(r_i)$ احتمال وجود رابطه‌ای با نوع r_i در مجموعه داده است. $P(a_1,a_2,\ldots,a_n)$ احتمال وجود گره‌هایی است که دارای مجموعه ویژگی $A=\{a_1,a_2,\ldots,a_n\}$ باشند.
همچنین مقدار احتمال $P(a_1,a_2,\ldots,a_n \mid r_i)$ از فرمول (۲) قابل محاسبه است.

$$P(a_1,a_2,\ldots,a_n \mid r_i) = \prod P(a_i \mid r_i) \quad (۲)$$

اگر فرض کنیم شبکه‌ای مانند شکل(۱) وجود دارد که ۳ نوع رابطه r_1,r_2,r_3 میان گره‌ها برقرار است و هدف ما در اینجا تعیین نوع پیوند میان دو گره X و Y است، در مرحله اول باید برای گره X براساس مجموعه‌ی ویژگی‌های $\{a_1,a_2,\ldots,a_n\}$ که متعلق به گره X است، مقدار احتمالات زیر را محاسبه کنیم:

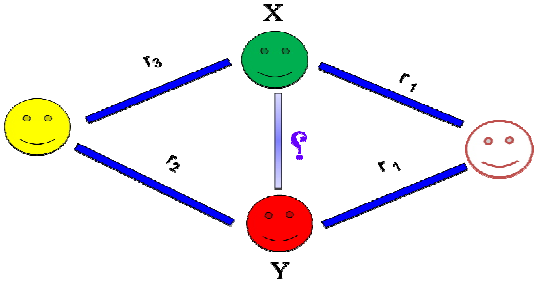
$$P(r_1 \mid a_1,a_2,\ldots,a_n) = \frac{P(a_1,a_2,\ldots,a_n \mid r_1)\,P(r_1)}{P(a_1,a_2,\ldots,a_n)}$$

$$P(r_2 \mid a_1,a_2,\ldots,a_n) = \frac{P(a_1,a_2,\ldots,a_n \mid r_2)\,P(r_2)}{P(a_1,a_2,\ldots,a_n)}$$

$$P(r_3 \mid a_1,a_2,\ldots,a_n) = \frac{P(a_1,a_2,\ldots,a_n \mid r_3)\,P(r_3)}{P(a_1,a_2,\ldots,a_n)}$$

سپس هر کدام از مقادیر احتمال که بزرگتر بود را به عنوان پاسخ گره X برای تعیین نوع رابطه‌اش با گره Y بر می‌گزینیم.

برای گره‌ی Y نیز مقادیر احتمال را مشابه آن‌چه برای گره X ذکر شد، براساس مجموعه ویژگی‌هایش محاسبه نموده و مقدار احتمال بیشتر را به عنوان پاسخ گره Y به عنوان نوع رابطه انتخاب می‌نماییم. پس در این روش ابتدا برای هر دو گره درگیر در رابطه، مقادیر احتمالات را حساب می‌نماییم و برای هر کدام، نوع رابطه‌ای که بیشترین احتمال را دارد برمی‌گزینیم.



شکل (۱) : نمونه‌ای از یک شبکه اجتماعی با نوع پیوند در رابطه‌ها

تا به اینجا بسیار شبیه الگوریتم بیز بوده با این تفاوت که برای هر گره به طور مجزا مقادیر احتمالات را محاسبه می‌نماییم.

مجموعه ویژگی‌های $A=\{a_1,a_2,\ldots,a_n\}$ که برای هر گره در شبکه ذکر شده‌است، می‌تواند ویژگی‌های توصیفی مربوط به گره، مانند سن، تحصیلات، محل سکونت و غیره باشد و یا ویژگی‌های ساختاری شبکه مانند درجه گره، تعداد دوستان مشترک میان دو گره باشد و یا ترکیبی از این دو نوع ویژگی. این مورد با توجه به اطلاعات شبکه‌ای که در دسترس است انتخاب می‌گردد.

۳-۲ مرحله دوم الگوریتم

هرچند تصمیم گیری بر اساس ویژگی های هرگره می‌تواند مفید باشد اما روابط مشترک میان دو گره، اطلاعات مفیدی در اختیار قرار می‌دهد که می‌تواند برای این مساله خاص راه‌گشا باشد. در اینجا باید سه‌تایی‌هایی تعیین شوند که در آن‌ها هر دوگره درگیر در پیوند قرار دارند. به این ترتیب دوستان مشترک میان دو گره به دست می‌آید. البته باید توجه داشت سه‌تایی‌هایی تعیین شوند که در آن‌ها نوع پیوند روابط یکسان است. به عبارتی دوستان مشترکی که با هر دو گره رابطه‌ی یکسانی دارند.

تعریف نسبت یک رابطه: نسبیت یک رابطه از نوع r_k برای گره X برابر است با تعداد روابط از نوع r_k گره X با همسایگان مجاورش به کل روابط گره X و با استفاده از فرمول (۳) قابل محاسبه است.

$$\text{Proportion}(X, r_k) = \frac{\text{تعداد روابط از نوع } r_k \text{ گره } X}{\text{تعداد کل روابط گره } X} \quad (۳)$$

این مقدار در واقع نشان می‌دهد چه درصدی از روابط گره X از نوع r_k است.

دانستن تعداد دوستان مشترک از هر نوع پیوند، هرچند مفید است اما اطلاعات کاملی در اختیار ما قرار نمی‌دهد، زیرا برخی از گره‌ها وجود دارند که با تعداد زیادی از گره‌های دیگر ارتباط بدون هدف دارند. برای مثال در برخی از شبکه‌های اجتماعی افرادی هستند که هدفشان تنها اضافه کردن به لیست دوستانشان است در حالی که قبلاً این افراد را نمی‌شناخته‌اند. بنابراین اطلاعاتی که این افراد به عنوان دوست مشترک دوگره در اختیار قرار می‌دهد قطعاً مفید نیستند چون بر مبنای واقعیت نیستند و نباید مبنای تصمیم گیری باشند. بنابراین نیاز است تا به گونه‌ای تاثیر این افراد را کم کنیم و در مقابل افرادی که روابط محدودتری دارند را بیشتر در تصمیم‌گیری خود دخیل نماییم. برای رسیدن به این هدف از فرمول (۴) بهره می‌بریم. در این فرمول با استفاده از اعمال تابع Log توانسته‌ایم، تاثیر گره‌های کم اهمیت را کاهش دهیم.

$$W(s , t , r_k) = Proportion(S, r_k) * \sum_{n \in N(s) \cap N(t) \cap R = rk} \frac{1}{\log(|N(n)|)} \quad (۴)$$

در فرمول(۴)، s و t دوگره اصلی هستند که می‌خواهیم نوع پیوند میان آن‌ها را تعیین کنیم و r_k نوع رابطه‌ای است که فعلاً برای آن دو حدس زده‌ایم. N(s) و N(t) به ترتیب تعداد همسایگان گره‌های s و t هستند.

مقدار $n \in N(s) \cap N(t) \cap R = rk$ گره‌هایی را شامل می‌شود که از همسایگان مشترک s و t هستند و با آن‌ها رابطه‌ی r_k دارد و N(n) تعداد همسایگان گره n است. با اعمال $\frac{1}{\log(|N(n)|)}$ تاثیر گره‌هایی که روابط بیهوده ایجاد می‌کنند را کاهش می‌دهیم.

مقدار به دست آمده از فرمول (۴) را به عنوان یک وزن در نظر می‌گیریم و در فرمول (۱) اعمال می‌کنیم. بنابراین فرمول اصلی برابر است با فرمول (۵).

$$P(r_i \mid a_1,a_2,\ldots,a_n) = W(b_i , b_j , r_k) \frac{P(a_1,a_2,\ldots,a_n \mid r_i)\, P(r_i)}{P(a_1,a_2,\ldots,a_n)} \quad (۵)$$

پس از محاسبه فرمول‌های (۵) برای هر دو گره‌ی درگیر در رابطه، نوع رابطه‌ای پیروز است که در مجموع بیشترین احتمال را از آن خودکند.

۴- ارزیابی روش پیشنهادی

برای ارزیابی روش ارائه‌شده، از مجموعه داده Epinions استفاده شده‌است. Epinions یک سایت تجارت الکترونیک است که کاربران می‌توانند در آن در مورد کالاها نظرات خود را ثبت نمایند. همچنین کاربران سایت می‌توانند با کاربران دیگر ارتباط برقرار نمایند و دو نوع رابطه‌ی دوستی و یا خصومت را با یکدیگر داشته باشند. پس از انجام مراحل پاک‌سازی داده، این مجموعه شامل ۹۵۳۱۴ گره و ۸۴۱۳۶۹ رابطه است. برای پیاده‌سازی روش، از نرم افزار Matlab استفاده شده‌است. انتخاب مجموعه تست با استفاده از روش Hold out است. بنابراین 2/3 از داده ها برای یادگیری و 1/3 برای تست، در نظر گرفته شده‌است. این مجموعه داده با روش ارائه شده در این مقاله و الگوریتم بیز و الگوریتم[۷]، برای پیش‌بینی نوع پیوندها، مورد ارزیابی قرار گرفته‌است. البته برای الگوریتم بیز نیاز بود که از روی روابط کلاس‌بندی انجام شود که مجموعه داده بر این اساس آماده شده‌است، تا بتواند با این الگوریتم سازگار باشد.

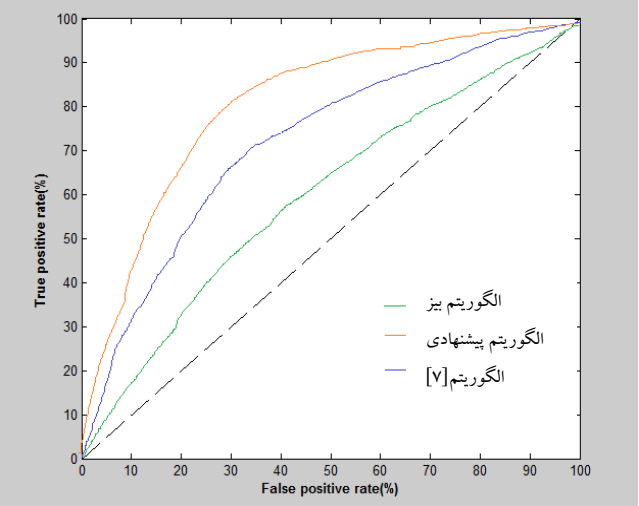
جدول (1) معیارهای دقت[۱۱]، فراخوان[۱۲] و نسبتF1[۱۳] را برای الگوریتم ارائه‌شده، روش کلاس‌بندی بیز و الگوریتم[۷]، نشان می‌دهد. همان طور که مشاهده می‌شود، الگوریتم ارائه شده کارایی بیشتری دارد.

جدول(۱): مقایسه معیارهای مختلف الگوریتم ارائه شده با روش بیز

معیار / الگوریتم	دقت	فراخوان	نسبت F1
بیز	۶۳.۹	۶۲.۱	۶۲.۸
الگوریتم[۷]	۷۸.۶	۷۴.۲	۷۶.۳
الگوریتم پیشنهادی	۸۳.۰	۸۵.۸	۸۲.۹

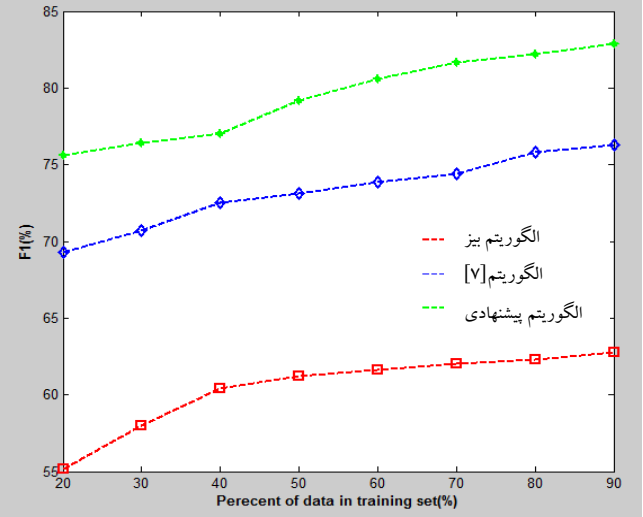
علاوه بر معیارهای ذکر شده در جدول(۱)، پاسخ هایی که به درست مثبت[۱۴] و یا به اشتباه مثبت[۱۵] ارزیابی شده‌اند، معیارهای مهمی در جهت ارزیابی کارایی یک الگوریتم هستند. نمودار ROC ابزار گرافیکی مناسبی برای بررسی این دو معیار می‌باشد.

شکل (۲) نمودار ROC را برای این دو روش نشان می‌دهد. خط اوریب در وسط نمودار نشان دهنده حالت حدس تصادفی برای تعیین نوع پیوند داده‌های تست است. دو منحنی دیگر الگوریتم بیز و روش ارائه شده در مقاله را نشان می‌دهند. هر منحنی که به خط حدس تصادفی نزدیک‌تر باشد، دقت کمتری دارد. به طور واضح مشخص است که روش ارائه شده در این مقاله دقت بیشتری در پیش‌بینی‌ها دارد.



شکل(۲): نمودار Roc مربوط به الگوریتم ارائه شده و روش بیز

در شکل (۳) مقایسه‌ای بین مقدار F1 الگوریتم ارائه شده، روش بیز و الگوریتم [۷] برحسب اندازه مجموعه یادگیری دارد. در واقع همان‌طور که انتظار می‌رود با افزایش حجم مجموعه‌ی یادگیری دقت هر سه روش افزایش می‌یابد. البته در روش بیز و الگوریتم [۷] یک جهش در افزایش دقت تا فاصله ۴۰ درصدی داریم و سپس حرکت کندتر انجام می‌شود، اما در روش ارائه شده این جهش تا فاصله ۵۰ درصد رخ می‌دهد و سپس شیب کندتر می‌شود.



شکل (۳) مقایسه مقدار F1 الگوریتم ارائه شده، روش بیز و الگوریتم [۷] برحسب اندازه مجموعه یادگیری

۵- نتیجه

در سال‌های اخیر تحلیل شبکه‌های اجتماعی مورد توجه بسیاری از محققان قرار گرفته‌است. در رابطه با شبکه‌های اجتماعی و بهبود عملکرد آن‌ها مسائل زیادی از دیر باز مطرح گردیده است و به مرور زمان مسائل جدیدی در حال گسترش است. یکی از مواردی که در تحلیل شبکه‌های اجتماعی در سال‌های اخیر مورد توجه قرار گرفته است، پیش‌بینی نوع پیوند در شبکه‌های ناهمگن است. در این مقاله یک راه حل برای مسئله پیش‌بینی نوع پیوند در شبکه‌های اجتماعی ناهمگن ارائه گردید. راه حل ارائه شده یک روش یادگیری بانظارت است که بر مبنای اصول کلاس‌بندی استوار است. در واقع روش ارائه شده توسعه‌ای بر مبنای الگوریتم بیز بود که موجب بهبود عملکرد و دقت الگوریتم بیز، در جهت حل مسئله پیش‌بینی نوع پیوند، گردید. روش ارائه شده به طور کلی در همه‌ی شبکه‌های اجتماعی قابل توسعه است. البته در این روش تمام موجودیت‌ها یکسان و از یک نوع در نظر گرفته شده‌اند اما برخی شبکه‌های اجتماعی شامل چندین نوع موجودیت هستند، مانند شبکه‌های فهرست مقالات[۱۶] که هر گره ممکن است از نوع نویسنده، مقاله و یا کنفرانس باشد[۴]. در این گونه شبکه‌ها نیاز است که ابتدا نوع موجودیت‌های یک رابطه شناسایی شود و سپس بررسی شود میان آن دو چند نوع پیوند می‌تواند برقرار گردد. این زمینه‌ی تحقیقاتی می‌تواند شروعی برای تحقیقات آینده در این زمینه باشد.

مراجع

[1] D.M. Boyd and N.B. Ellison, *Social Network Sites: Definition, History, and Scholarship,* Journal of Computer-Mediated Communication, Vol.13, pp. 210–230, 2008.

[2] E. Raad, R. Chbeir, and A. Dipanda, *Discovering relationship types between users using profiles and shared photos in a social network,* Journal of Multimedia Tools and Applications, pp.1–30, 2011.

[3] M. Roth, A. David, D. Deutscher, G. Flysher, I. Horn, A. Leichtberg, N. Leiser,Y. Matias, and R. Merom, *Suggesting friends using the implicit social graph,* In Proc. of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 233–242, 2010.

[4] Y. Matsuo, J. Mori, M. Hamasaki, T. Nishimura, H. Takeda, K. Hasida, and M. Ishizuka, *Polyphonet: An advanced social network extraction system from the web*, Journal of Web Semant.,vol. 5, pp. 262–278, December 2007.

[5] W. Tang, H. Zhuang, and J. Tang, *Learning to infer social ties in large networks,* In Proc. of the ECML/PKDD 2011.

[6] H.Wang and L. Sun, *Trust-involved access control in collaborative open social networks,* In Proc. of the 2010 Fourth International Conference on Network and System Security, ser. NSS '10. Washington, DC, USA: IEEE Computer Society, pp. 239–246, 2010.

[7] R. Heatherly, M. Kantarcioglu, and B. M. Thuraisingham, *Social network classification incorporating link type values,* ISI, pp. 19-24, IEEE, 2009.

[8] A. Javari and M. Jalili, *Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links,* ACM Transaction on Intelligent Systems and Technology, 2013.

[9] T. Zhang, H. Chao, C. Willis, and D. Tretter, *Consumer image retrieval by estimating relation tree from family photo collections,* In Proc. of the ACM International Conference on Image and Video Retrieval, 2010.

زیرنویس‌ها

۱ Link type prediction
۲ Classification
۳ www.epinions.com
۴ Co-author
۵ Co-lab
۶ Co-project
۷ Co-conference
۸ Sign
۹ Nodes
۱۰ Comment
۱۱ Precision
۱۲ Recall
۱۳ F1-score
۱۴ True positive
۱۵ False positive
۱۶ Bibliographic network