روشی نوین برای پیش‌بینی ارتباط در شبکه‌های اجتماعی ناهمگن

سعیده رضاوندی شعاعی و هادی زارع

چکیده: با گسترش روزافزون شبکه‌های اجتماعی، علوم شبکه مورد توجه بسیاری از پژوهشگران در زمینه‌های مختلف قرار گرفته است. علاوه بر آن بسیاری از مسایل کاربردی مهندسی با استفاده از ابزار شبکه‌های اجتماعی مدل‌سازی شده‌اند. پیش‌بینی تغییر و تحول در ساختار شبکه‌های اجتماعی یکی از مسایل اساسی در تحلیل شبکه‌های اجتماعی است که با عنوان مسأله پیش‌بینی ارتباط در علوم شبکه شناخته می‌شود. امروزه با گسترش استفاده از شبکه‌های اجتماعی، فعالیت افراد در قالب چندین شبکه با عنوان شبکه‌های اجتماعی ناهمگن رواج پیدا کرده است. پیش‌بینی ارتباط در شبکه‌های اجتماعی ناهمگن را می‌توان بر اساس اطلاعات اضافی موجود نسبت به روش‌های قبلی مورد بهبود قرار داد. در رویکرد پیشنهادی این مقاله، ابتدا یک معیار شباهت جدید برای کاربران در شبکه‌های ناهمگن بر اساس توسعه روش‌های مطرح پیشین و با در نظر گرفتن ارتباط بین لایه‌های مختلف معرفی می‌شود، سپس با استفاده از رویکرد یادگیری باناظر و بهره‌گیری از ویژگی‌های تولیدشده بر مبنای معیار شباهت معرفی‌شده، الگوریتم پیشنهادی مورد تشریح قرار می‌گیرد. برای ارزیابی روش پیشنهادی از معیارهای استاندارد همانند دقت و صحت بهره گرفته‌ایم. مقایسه روش پیشنهادی با روش‌های شناخته‌شده پیشین بر روی مجموعه داده‌های مختلف نشان می‌دهد که روش پیشنهادی ما برای پیش‌بینی ارتباط از عملکرد بهتر و مطلوب‌تری برخوردار است به طوری که از نظر صحت تا ۲۰ درصد موجب بهبود عملکرد شده است.

کلیدواژه: **اندازه شباهت، پیش‌بینی ارتباط، شبکه‌های اجتماعی، شبکه‌های اجتماعی ناهمگن، کاوش ارتباطات، یادگیری با ناظر.**

۱- مقدمه

گسترش شبکه‌های اجتماعی، استقبال زیاد افراد مختلف از آنها و پدیدآمدن حجم انبوهی از اطلاعات، موجب پیدایش علم نوینی به نام شبکه در قرن حاضر گردید که به تحلیل و مدل‌سازی پدیده‌های مختلفی که بتوان آنها را در قالب شبکه نمایش داد می‌پردازد [۱]. فراگیری، استفاده روزافزون و مدل‌سازی پدیده‌های مختلف در مهندسی و علوم کاربردی با استفاده از ابزارها و دانش سامانه‌های شبکه‌ای در میان پژوهشگران در حوزه‌های مختلف فناوری اطلاعات، ارتباطات، ریاضیات، کامپیوتر و فیزیک برای مدل‌سازی پدیده‌های مختلف بر اهمیت آن افزوده است. در حقیقت می‌توان گفت که در دنیای امروز، افراد از شبکه‌های مختلفی به صورت آشکار یا پنهان بهره می‌گیرند که کاوش اطلاعات کاربران، نحوه تعاملات و رفتار آنها در این شبکه‌ها می‌تواند دارای کاربردهای زیادی باشد. این شبکه‌ها می‌توانند دربرگیرنده شبکه‌های اجتماعی، شبکه‌های زیستی، شبکه‌های حمل و نقل، شبکه دوستی و همکاری بین افراد و انواع دیگری از شبکه‌ها باشد. از طرف دیگر بحث چگونگی تحول شبکه‌ها و پیش‌بینی رفتار آنها در آینده از مباحث بسیار مهمی است که مورد توجه بسیاری از پژوهشگران قرار دارد.

به منظور ارائه یک تجزیه و تحلیل صحیح از شبکه، باید ماهیت آن به خوبی شناخته شود. شناخت ماهیت شبکه‌ها وابسته به درک تعاملات و ارتباطات مختلف موجود در میان گره‌ها به عنوان موجودیت‌های شبکه است. از طرفی داده‌های موجود در شبکه‌های اجتماعی به دلیل دارابودن خصوصیاتی از قبیل پویایی، تخصیص ویژگی‌های منحصر به فرد به هر گره و حجم انبوه داده‌ها، مورد توجه بسیاری از پژوهشگران در زمینه‌های متفاوت قرار گفته است.

پژوهشگران علوم مختلف نظیر روانشناسی، پزشکی، علوم کامپیوتر، فیزیک، اقتصاد و علوم زیستی از پیش‌بینی ارتباط در علوم شبکه در کاربردهای گوناگون بهره گرفته‌اند. از جمله می‌توان استفاده در سیستم‌های توصیه‌گر به منظور ارائه پیشنهادات بهتر به کاربر [۲]، شبکه‌های جنایی به منظور شناخت روابط بین افراد و شناسایی افراد کلیدی در این شبکه‌ها [۳]، شبکه‌های اجتماعی با هدف پیشنهاد دوستی و یا همکاری بین افرادی که شباهت زیادی بین آنها وجود دارد [۴]، شناسایی تعامل بین ژن‌ها در شبکه‌های زیستی [۵] و شناسایی و تشخیص بیماری‌ها [۶] را نام برد. از دیگر کاربردهای پیش‌بینی ارتباط می‌توان به پیش‌بینی اطلاعات کل شبکه با استفاده از اطلاعات بخشی از شبکه، درک بهتر تکامل شبکه، پیش‌بینی ارتباطات در شبکه‌های پویا و ناهمگن اشاره کرد [۷] و [۸].

مسئله پیش‌بینی ارتباط شامل پیچیدگی‌ها و دشواری‌های متعددی می‌باشد که از جمله آنها می‌توان به امکان‌پذیربودن دسترسی به سطحی از ارتباطات در شبکه با توجه به حجم عظیم اطلاعات در دسترس و وابستگی به زمان و پویایی شبکه‌ها اشاره کرد [۹]. پیش‌بینی ارتباطات از بین رفته- ارتباطاتی که واقعاً در شبکه وجود دارند اما داده‌های موجود فاقد این ارتباطات است- و ارتباطاتی که ممکن است در زمان‌های آینده در شبکه ایجاد یا حذف شوند، از مسایل مطرح در زمینه تحلیل شبکه‌های اجتماعی است که به عنوان مسأله پیش‌بینی ارتباط شناخته شده و در سال‌های اخیر حجم زیادی از تحقیقات در این زمینه صورت گرفته است [۱۰] تا [۱۲].

اکثر فعالیت‌های انجام‌شده در زمینه پیش‌بینی ارتباط برای شبکه‌های همگن، شبکه‌هایی با یک نوع ارتباط و یک نوع گره بوده است [۱۳] در حالی که اکثر شبکه‌های واقعی ماهیتی ناهمگن دارند. ناهمگن‌بودن شبکه‌ها می‌تواند از نظر نوع گره‌ها یا از نظر تعداد ارتباطات بین گره‌ها باشد [۱۴]. در سال‌های اخیر پیش‌بینی ارتباط در شبکه‌های ناهمگن مورد توجه پژوهشگران مختلف قرار گرفته است.

اکثر الگوریتم‌های ارائه‌شده در این زمینه، ابتدا درجه نزدیکی یا مشابهت بین دو گره را تعیین کرده و سپس با استفاده از این درجه شباهت به پیش‌بینی ارتباط می‌پردازند. درجه مشابهت بین دو گره می‌تواند تنها بر اساس ویژگی‌های خود گره‌ها محاسبه شود یا علاوه بر ویژگی‌های گره‌ها می‌تواند از اطلاعات ساختاری شبکه نظیر مسیرهای موجود بین دو

این مقاله در تاریخ ۳۱ مرداد ماه ۱۳۹۵ دریافت و در تاریخ ۲ دی ماه ۱۳۹۶ بازنگری شد.

سعیده رضاوندی شعاعی، دانشکده علوم و فنون نوین، دانشگاه تهران، تهران، ایران، (email: s.rezavandi@ut.ac.ir).

هادی زارع (نویسنده مسئول)، دانشکده علوم و فنون نوین، دانشگاه تهران، تهران، ایران، (email: h.zare@ut.ac.ir).

Archive of SID

www.SID.ir

گره نیز استفاده نماید. مهم‌ترین و اصلی‌ترین موضوع مطرح در زمینه پیش‌بینی ارتباط به دست آوردن شباهت بین گره‌ها می‌باشد.

از طرف دیگر با توجه به ساختار شبکه و بر اساس ویژگی‌های استخراج‌شده از آن و بررسی شباهت بین افراد می‌توان پیش‌بینی ارتباط را از دیدگاه روش‌های یادگیری مورد بررسی قرار داد. الگوریتم‌های بی‌ناظر جفت گره‌هایی که دارای درجه شباهت بیشتری هستند را به عنوان نامزد مناسب برای ارتباطات آینده در نظر می‌گیرند. در رویکرد باناظر، بر اساس ارتباطات کنونی، می‌توان مسأله پیش‌بینی ارتباط را همانند یک مسأله طبقه‌بندی دوکلاسه دانست که در آن یکی از کلاس‌ها به منزله وجود ارتباط در حال حاضر است و دیگری برای عدم وجود ارتباط بین گره‌ها در نظر گرفته می‌شود [۱۵].

مسأله مهمی که وجود دارد این است، هنگامی که ماهیت شبکه به یک شبکه ناهمگن تغییر پیدا می‌کند، نمی‌توان از معیارهای مطرح در شبکه‌های همگن استفاده نمود. چرا که با در نظر گرفتن ناهمگنی شبکه، اطلاعات ارزشمندی به مسأله افزوده می‌شود که در صورت استفاده از روش‌های پیشین و نادیده‌گرفتن این اطلاعات، پیش‌بینی و مدل‌سازی انجام‌گرفته دارای دقت پایین‌تری خواهد بود [۱۶].

از طرف دیگر بهره‌گیری از اطلاعات ارزشمند موجود در شبکه‌های ناهمگن در پیش‌بینی ارتباط می‌تواند چالش‌های موجود در این زمینه را تا حدی مرتفع کند. به عنوان مثال، بسیاری از شبکه‌های اجتماعی به تنهایی دارای ارتباطات کمی در میان کاربران خود هستند و به اصطلاح دارای داده‌های خلوتی می‌باشند که باعث مشکلات زیادی در روش‌های یادگیری ماشین و کاوش دادگان انبوه شده است [۱۷]. در نظر گرفتن چندین شبکه اجتماعی با یکدیگر می‌تواند موجب افزایش اطلاعات در دسترس و در نتیجه حل مشکل خلوتی داده‌ها شود تا دقت پیش‌بینی را از یک حد نازل به یک حد مطلوب ارتقا بخشد.

در این مقاله یک روش نوین برای پیش‌بینی ارتباط در شبکه‌های ناهمگن بر اساس رویکرد یادگیری باناظر ارائه می‌شود. ابتدا با بهره‌گیری از معیارهای پیشین و در نظر گرفتن ارتباط و همبستگی بین شبکه‌های مختلف معیارهای جدیدی برای محاسبه درجه مشابهت گره‌ها در شبکه‌های ناهمگن تعریف می‌شود و این معیارهای محاسبه‌شده به عنوان بردار ویژگی‌های الگوریتم‌های باناظر برای پیش‌بینی ارتباط در شبکه‌های ناهمگن استفاده می‌شوند. برای ارزیابی روش پیشنهادی از مجموعه داده‌های استاندارد موجود استفاده شده است. نتایج به دست آمده حاکی از بهبود بامعنای نتایج با استفاده از روش پیشنهادی نسبت به روش‌های قبلی است.

ساختار کلی این مقاله به شرح ذیل می‌باشد. در بخش دوم مروری بر ادبیات پیشین انجام گرفته است. در بخش سوم تعاریف اولیه ارائه شده و مسأله پیش‌بینی ارتباط در شبکه‌های ناهمگن به صورت شفاف توضیح داده می‌شود. در بخش چهارم روش پیشنهادی برای پیش‌بینی ارتباط در شبکه‌های ناهمگن معرفی می‌شود. در بخش پنجم مجموعه داده‌های استفاده‌شده تشریح و نتایج به دست آمده از روش‌های پیشین و روش پیشنهادی ارائه می‌شود. نتیجه‌گیری و کارهای آینده در بخش پایانی مورد بحث قرار گرفته است.

۲- مرور ادبیات پیشین

شکل‌گیری رسمی مسأله پیش‌بینی ارتباط در سال ۲۰۰۷ با مقاله مروری کلینبرگ و همکارانش [۱۶] پررنگ‌تر شد به نحوی که بعد از آن مطالعات فراوانی در این زمینه بر اساس دیدگاه‌های مختلف و رویکردهای گوناگون انجام گرفته است. به طور معمول این روش‌ها نیازمند محاسبه شباهت میان افراد هستند تا بتوانند بر اساس آن به پیش‌بینی رفتار آنها در آینده و ارتباطات آنها بپردازند.

برای تعیین درجه شباهت بین گره‌های یک شبکه، روش‌های مختلفی وجود دارد. برخی از این روش‌ها مانند روش همسایگان مشترک[1] [۱۸]، ضریب جاکارد[2] و شاخص سورنسن[3] [۱۹] برای پیش‌بینی تنها از اطلاعات محلی شبکه استفاده می‌کنند، اما برخی دیگر از روش‌ها مانند روش مسیر محلی[4] [۵] و کتز[5] [۲۰] علاوه بر اطلاعات محلی از اطلاعات کل شبکه و اطلاعات مسیرهای موجود در شبکه نیز استفاده می‌کنند. این دسته از الگوریتم‌ها نسبت به دسته قبل از پیچیدگی زمانی بالاتری برخوردار هستند. در طیف دیگری از الگوریتم‌ها از روش‌های احتمالاتی و قدم‌زنی تصادفی برای پیش‌بینی ارتباط استفاده شده است [۲۱] و [۲۲]. یکی از روش‌های متداول برای پیش‌بینی ارتباط در شبکه‌های ناهمگن، همگن فرض کردن ارتباطات و پیش‌بینی ارتباط بر اساس روش‌های متداول می‌باشد [۲۳]. بر اساس این رویکرد در هر شبکه، در هر لایه از شبکه‌های ناهمگن پیش‌بینی ارتباط بر اساس رویکردهای متداول صورت می‌پذیرد و سپس برای پیش‌بینی نهایی ارتباط بین دو گره از وزن‌هایی برای هر شبکه استفاده و یک میانگین وزنی حساب می‌شود. بدیهی است که پیش‌بینی ارتباط بر این اساس رویکرد مناسبی نیست زیرا به طور مثال در شبکه‌های چندلایه- که در هر لایه شبکه‌ای متفاوت از دیگر لایه‌ها وجود دارد- فرض یکسان‌بودن پارامترها برای همه شبکه‌های موجود در لایه‌های مختلف فرض معقولی به نظر نمی‌رسد. ممکن است توزیع درجه، چگالی و ضریب خوشه‌بندی دو شبکه به صورت معناداری با یکدیگر متفاوت باشند به طوری که قابل چشم‌پوشی نباشد. همچنین گاهی اوقات یک شبکه در مقایسه با شبکه دیگری که متراکم است، ماهیت خلوت دارد و در صورت ادغام این دو شبکه با فرض همگنی شبکه، بسیاری از اطلاعات مفید از دست خواهد رفت.

از دیگر روش‌های مطرح برای پیش‌بینی ارتباط درشبکه‌های ناهمگن می‌توان به تحلیل ساختار هر شبکه به صورت مجزا و بدون در نظر گرفتن ارتباط بین شبکه‌های مختلف در [۲۴] اشاره کرد. عیب این روش از بین رفتن بخشی از اطلاعات می‌باشد زیرا ممکن است ارتباطات موجود در شبکه‌های مختلف روی هم تأثیرگذار باشند و هنگامی که به تنهایی مورد تجزیه و تحلیل قرار می‌گیرند، این تأثیرات در نظر گرفته نشوند.

از دیگر فعالیت‌های انجام‌شده در زمینه پیش‌بینی ارتباط در شبکه‌های ناهمگن می‌توان به [۲۵] اشاره کرد که در آن یک روش ترکیبی بر اساس نزدیکی مثلثی و ضریب جاکارد پیشنهاد شده است. از دیگر کارهای انجام‌شده در این زمینه، [۲۶] است. در این مقاله علاقه‌مندی‌های هر کاربر در هر شبکه به صورت جداگانه مورد بررسی قرار می‌گیرد، سپس ارتباط بین علاقه‌مندی کاربر در شبکه مورد مطالعه و سایر شبکه‌ها به دست آمده و در نهایت با توجه به این علاقه‌مندی‌ها و ارتباطات آنها به پیش‌بینی ارتباط پرداخته می‌شود.

تانگ و همکارانش در [۲۷] با استفاده از مفاهیم نظریه اجتماعی مانند حفره‌های ساختاری[6] به بررسی مسأله پیش‌بینی ارتباط در شبکه‌های

1. Common Neighbor
2. Jaccard Coefficient
3. Sorenson Index
4. Local Paths
5. Katz
6. Structural Holes

Archive of SID

www.SID.ir

جدول ۱: مقایسه روش‌های پیش‌بینی ارتباط در شبکه‌های همگن.

نام الگوریتم	پیچیدگی زمانی	نرمال‌شده	خصوصیات
CN	$O(n^2)$	خیر	روش بسیار ساده و قابل درک. هرچه تعداد همسایگان مشترک دو گره بیشتر باشد احتمال برقراری ارتباط بین آنها بیشتر است [۱۷].
JC	$O(2n^2)$	بله	نسبت همسایه‌های مشترک به تعداد کل همسایه‌های دو گره [۱۹]
SI	$O(n^2)$	بله	احتمال برقراری ارتباط بین گره‌ها با درجه پایین‌تر، بیشتر است [۱۶].
SC	$O(n^2)$	بله	معیار کسینوس [۹]
AA	$O(2n^2)$	خیر	همسایگان با درجه کمتر، در تعیین مشابهت دو گره مؤثرتر می‌باشند [۱۸].
PA	$O(2n)$	خیر	روشی آسان و قابل درک. هرچه درجه دو گره بیشتر باشد احتمال برقراری ارتباط بین آنها بیشتر است [۱۷].
RA	$O(2n^2)$	خیر	مانند آدامیک و آدار عمل می‌کند اما دارای تفاوت‌های کوچکی می‌باشند [۱۷].

جدول ۲: مقایسه روش‌های پیش‌بینی ارتباط در شبکه‌های ناهمگن.

سال	شبکه	نوع شبکه	در نظر گرفتن تأثیر روابط
۲۰۱۱ [۲۳]	یوتیوب	ناهمگن از نظر ارتباطات	بله
	ژن- بیماری	ناهمگن از نظر ارتباطات و گره‌ها	
۲۰۱۱ [۱۴]	آب و هوایی	ناهمگن از نظر ارتباطات	خیر
	نویسندگان همکار		
	هنرپیشگان همکار		
۲۰۱۲ [۱۳]	نویسندگان همکار	ناهمگن از نظر ارتباطات و گره‌ها	بله
	ژن- بیماری	ناهمگن از نظر ارتباطات و گره‌ها	
۲۰۱۲ [۲۷]	همکاری نویسندگان، اسلش- دات	ناهمگن از نظر ارتباطات	بله
۲۰۱۴ [۲۴]	توییتر، فوراسکور	ناهمگن از نظر ارتباطات و گره‌ها	بله
۲۰۱۴ [۴۴]	فیس‌بوک، توییتر، گیت‌هاب	ناهمگن از نظر ارتباطات	بله
۲۰۱۵ [۳۰]	دی‌بی‌ال‌بی	ناهمگن از نظر ارتباطات	بله
۲۰۱۵ [۳۱]	اسکولار مت	شبکه دوبخشی	خیر
۲۰۱۷ [۱۲]	توییتر، فوراسکور	شبکه ناهمگن از نظر ارتباطات	بلی
۲۰۱۷ [۴۵]	ارکایو فیزیک، فیس‌بوک	ناهمگن از نظر ارتباطات	بلی

ناهمگن پرداخته‌اند.

در [۲۳] دیویس و همکارانش با استفاده از اطلاعات سه‌تایی‌های موجود در شبکه روش آدامیک و آدار[1] را توسعه داده و با استفاده از این معیار به پیش‌بینی ارتباط پرداخته‌اند. همچنین پژوهش دیگری که در زمینه پیش‌بینی ارتباط انجام گرفته شده است [۲۸] می‌باشد. در این مقاله با استفاده از الگوریتم پیمایش اول سطح و الگوی انتشار در شبکه‌ها به پیش‌بینی ارتباط پرداخته شده است. همچنین پژوهش‌هایی مانند [۲۴] و [۲۹] با استفاده از فرامسیرهای موجود به پیش‌بینی ارتباط پرداخته‌اند.

پژوهش دیگری که اخیراً در زمینه پیش‌بینی ارتباط در شبکه‌های ناهمگن انجام گرفته است [۳۰] می‌باشد. این پژوهش با توجه به لایه‌های موجود در شبکه به شناسایی ویژگی‌های جدید در شبکه پرداخته است. اساس کار به این صورت است که در هر لایه به صورت جداگانه به شناسایی معیارهای شباهت بین دو گره پرداخته می‌شود. سپس با توجه به بردار شباهت به دست آمده از لایه‌های مختلف به شناسایی معیار شباهت‌های ناهمگن پرداخته می‌شود.

به صورت اجمالی روش‌های مهم پیش‌بینی ارتباط در شبکه‌های همگن در جدول ۱ به همراه ویژگی‌های آنها آورده شده است. همچنین مهم‌ترین رویکردهای موجود برای پیش‌بینی ارتباط در شبکه‌های ناهمگن بر اساس ویژگی‌های آنها در جدول ۲ تبیین شده است.

بیشتر روش‌های پیشنهادی در ادبیات موضوع از وزن‌هایی برای لایه‌های مختلف شبکه ناهمگن بهره گرفته و از رویکردهای متداول برای محاسبه شباهت میان گره‌ها استفاده می‌کنند [۳۰]. در حالی که در این مقاله یک معیار شباهت نوین بر اساس شهود احتمالاتی در شبکه‌های ناهمگن معرفی می‌شود. همچنین این رویکرد از بار محاسباتی کمتری نسبت به روش‌های قبلی معرفی‌شده مانند [۸] و [۲۳] برخوردار می‌باشد و نتایج به دست آمده نشان از برتری آن نسبت به روش‌های پیشین بر اساس معیارهای استاندارد دقت و صحت دارد. علاوه بر آن یک تفسیر و شهود احتمالاتی برای روش پیشنهادی وجود دارد که آن را از کارهای مشابه متمایز می‌کند.

۳- تعاریف اولیه

۳-۱ تعریف مسأله

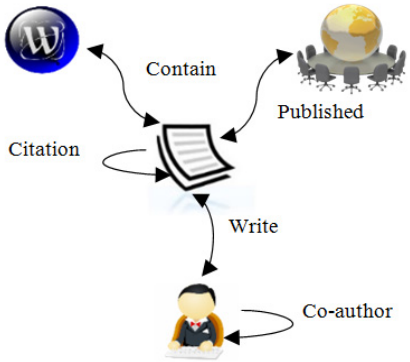
با افزایش استفاده از شبکه‌های اجتماعی، چگونگی تحلیل و استخراج الگوها و مدل‌سازی آنها به منظور کنترل، پیش‌بینی و مطالعه رفتار کاربران از اهمیت فراوانی برخوردار شده است. یکی از حوزه‌های مهم در علوم شبکه که توجه پژوهشگران زیادی را به خود جلب کرده است، پیش‌بینی ارتباط می‌باشد [۱۶].

پیش‌بینی ارتباط به شناسایی ارتباطات نهفته در شبکه یا پیش‌بینی ارتباطاتی که در حال حاضر در شبکه وجود ندارند، اما ممکن است در آینده به وجود آیند اطلاق می‌شود. فعالیت افراد در شبکه‌های مختلف تنها به یک شبکه محدود نمی‌شود، به بیان دیگر می‌توان گفت افراد مختلف تنها در یک زمینه با هم به تعامل نمی‌پردازند و می‌توانند در زمینه‌های

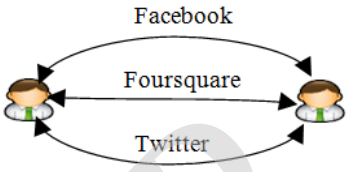
1. Adamic and Adar

Archive of SID

www.SID.ir



شکل ۱: شبکه ناهمگن همکاری نویسندگان (ناهمگن از نظر گره و ارتباط) [۳۱].



شکل ۲: شبکه ناهمگن دوستی افراد (ناهمگن از نظر ارتباط).

مختلف با یکدیگر ارتباط برقرار کنند [۱۸]. برای مثال می‌توان فعالیت کاربران در شبکه‌های اجتماعی مختلف مانند فیس‌بوک، توییتر، فوراسکور را در نظر گرفت. از دیگر مثال‌ها در این زمینه می‌توان به شبکه همکاران دانشگاهی در زمینه‌های مختلف مثل علوم کامپیوتر، مهندسی برق و فناوری اطلاعات اشاره کرد.

تعامل افراد در بستر شبکه‌های مختلف منجر به ظهور شبکه‌های ناهمگن اجتماعی شده است. در حقیقت، شبکه‌های ناهمگن شبکه‌هایی هستند که در آنها نوع گره‌ها یا ارتباطات تنها به یک نوع، محدود نمی‌شود. نمونه بارزی از شبکه‌های ناهمگن در شکل ۱ مشاهده می‌شود. در شبکه ناهمگن نمایش داده شده در شکل ۱ گره‌ها شامل نویسنده، مقاله، کنفرانس و کلمه کلیدی می‌باشد. ارتباطات ناهمگن موجود در این شبکه شامل ارتباط همکاری بین دو نویسنده، نوشتن مقاله توسط نویسنده، چاپ مقاله در کنفرانس، ارجاع‌دهی مقالات به هم و مقاله حاوی کلمه کلیدی‌بودن می‌باشد [۳۱].

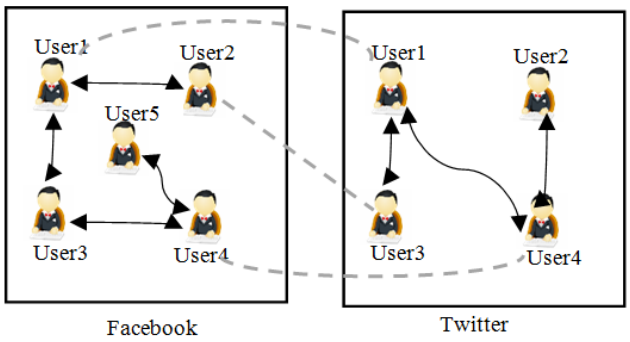
نوع متداول از شبکه‌های ناهمگن گونه‌ای از شبکه‌ها است که تنها از نظر ارتباطات ناهمگن باشند. در حقیقت در این شبکه‌ها گره‌ها یکسان است و ارتباطات می‌تواند در شبکه‌های مختلف متفاوت باشد. از جمله این شبکه‌ها می‌توان به همکاری نویسندگان در زمینه‌های مختلف اشاره نمود.

شکل ۲ بیانگر شبکه ناهمگن از نظر ارتباط است. این شکل نشان‌دهنده دوستی افراد در شبکه‌های اجتماعی مختلف می‌باشد.

همان طور که مشاهده می‌شود تعریف شبکه ناهمگن به سطح تجزیه و تحلیل پژوهشگر بستگی دارد.

۲-۳ تعاریف اولیه

شبکه ناهمگن به صورت رسمی به شبکه‌ای اطلاق می‌شود که در آن انواع مختلفی از یال‌ها و گره‌ها وجود داشته باشد. تعریف عمومی این شبکه‌ها به صورت $G=(V,E)$ می‌باشد که در آن V نشانگر انواع گره‌ها است که به صورت $V=\cup_{i=1}^{n}V_i$ تعریف می‌شود. در این رابطه V_i بیانگر نوع خاصی از گره و n نشان‌دهنده تعداد انواع گره‌ها است. همچنین E ارتباطات موجود در شبکه را در قالب $E=\cup_{i=1}^{m}E_i$ نمایش می‌دهد که E_i و m به ترتیب بیانگر نوع خاصی از ارتباطات و تعداد انواع ارتباطات می‌باشد.



شکل ۳: شبکه ناهمگن جزئی هم‌پوشان.

در این مقاله تمرکز ما روی شبکه‌های ناهمگن از نوع ارتباط می‌باشد. در این شبکه‌ها $n=1$ است اما مقدار m هر عددی است که بزرگ‌تر از یک باشد.

گره‌ها در شبکه‌های ناهمگن می‌توانند کاملاً هم‌پوشانی داشته باشند، به این معنی که تعداد گره‌ها در تمام شبکه‌ها یکسان است و تناظر یک به یکی بین گره‌های شبکه‌ها وجود دارد.

کشف شبکه‌هایی که گره‌های آنها کاملاً همپوشانی دارند کار آسانی نخواهد بود، به این دلیل که تقریباً شبکه‌های موجود شبکه‌هایی هستند که به صورت جزئی هم‌پوشان هستند. منظور از هم‌پوشانی جزئی این است که شبکه‌ها دارای گره‌هایی از یک نوع هستند اما تعداد گره‌های دو شبکه الزاماً برابر نیست و هر شبکه می‌تواند شامل برخی گره‌های منحصر به فرد نیز باشد [۳۲]. شکل ۳ نشان‌دهنده یک شبکه ناهمگن متشکل از دو شبکه اجتماعی جزئی هم‌پوشان است که در آن کاربرانی که با خطوط خط‌چین به هم وصل هستند کاربران مشترک دو شبکه را نشان می‌دهد.

۳-۳ پیش‌بینی ارتباط در شبکه‌های ناهمگن

پیش‌بینی ارتباط در شبکه‌های ناهمگن به این صورت اتفاق می‌افتد که در یک لایه خاص، وجود ارتباط بین دو گرهی که در حال حاضر بین آنها ارتباطی وجود ندارد، $e(u,v)\notin E_i$، در زمان آینده پیش‌بینی می‌شود. برای انجام یک پیش‌بینی دقیق باید از اطلاعات موجود در شبکه ناهمگن به درستی استفاده نمود. به منظور تعریف یک معیار شباهت دقیق باید ادغام صحیحی از اطلاعات تمام لایه‌های موجود در شبکه با اطلاعات لایه‌ای که قصد پیش‌بینی ارتباط در آن داریم، داشته باشیم.

در قسمت بعد یک روش جدید برای محاسبه معیار شباهت بین گره‌ها ارائه می‌شود.

۴- روش پیشنهادی

هدف این بخش ارائه روشی برای محاسبه مشابهت بین دو گره در یک لایه خاص از شبکه ناهمگن است به طوری که بتوان با استفاده از این معیار پیش‌بینی خوبی انجام داد. قبل از ارائه الگوریتم پیشنهادی به توضیح برخی از روش‌های پایه می‌پردازیم.

۴-۱ روش‌های پایه در زمینه پیش‌بینی ارتباط

۴-۱-۱ همسایگان مشترک

ایده اصلی در این روش این است که هرچه تعداد همسایگان دو گره بیشتر باشند درجه شباهت بین آنها نیز افزایش یافته و با احتمال بیشتری در آینده بین آنها ارتباطی به وجود می‌آید [۱۶]

$$CN(u,v)=\left|\Gamma(u)\cap\Gamma(v)\right| \qquad (۱)$$

Archive of SID

www.SID.ir

در (۱)، $\Gamma(u)$ نشان‌دهنده مجموعه همسایگان گره u و $|\Gamma(u)|$ نشانگر تعداد همسایگان آن گره می‌باشد.

٤-۱-۲ ضریب آدامیک و آدار

این روش، تعمیم‌یافته همسایگان مشترک است. ایده مورد استفاده در این روش به این صورت است که همسایگان مشترکی که دارای درجه پایین‌تری هستند دارای ارزش بیشتری می‌باشند [۳۲]

$$AA(u,v)=\sum_{z\in\Gamma(u)\cap\Gamma(v)}\frac{1}{\log|\Gamma(z)|} \qquad (۲)$$

برای مطالعه سایر روش‌های پیش‌بینی ارتباط می‌توان به مقالات مروری نوشته‌شده در این زمینه مراجعه کرد [۹] و [۱۸].

٤-۲ روش پیشنهادی

برای راحتی کار، در شبکه‌های ناهمگن به جای هر کدام از شبکه‌ها از اصطلاح لایه استفاده می‌کنیم.

پیش‌بینی ارتباط در شبکه‌های ناهمگن به این صورت اتفاق می‌افتد که قصد داریم بین دو گره غیر متصل در لایه‌ای خاص وجود ارتباط پنهانی یا ارتباط در زمان آینده را پیش‌بینی کنیم. برای این منظور باید میزان شباهت دو گره را در لایه مورد مطالعه محاسبه نماییم. به منظور تعیین یک معیار شباهت بین گره‌ها ابتدا معیارهای لازم شرح داده می‌شود و در نهایت معیار شباهت تعیین‌شده ارائه می‌گردد.

تعریف میزان تأثیرپذیری لایه‌ها: برای محاسبه میزان تأثیرپذیری لایه‌های مختلف بر روی یکدیگر از نسبت ارتباط‌های مشترک دو لایه به روی سهم ارتباط‌های موجود در لایه مورد مطالعه به صورت رابطه زیر بهره گرفته می‌شود

$$P(Correctness(L_1|L_2))=\frac{P(Edges\ of\ L_1|Edges\ of\ L_2)}{P(Edges\ of\ L_2)} \qquad (۳)$$

جدول ۳ میزان تأثیرپذیری دو شبکه موجود در شکل ۴ را نشان می‌دهد.

به عنوان نمونه می‌توان گفت توییتر شامل ۴ ارتباط است که تعداد ۳ ارتباط از ارتباطات این شبکه با فیس‌بوک مشترک است و بنابراین تأثیرپذیری این شبکه از فیس‌بوک برابر ۰٫۶ می‌باشد. این احتمال به این معناست که اگر ارتباطی در شبکه فیس‌بوک ایجاد شود به احتمال ۰٫۶ درصد در توییتر نیز به وجود خواهد آمد.

در ادامه مقاله نماد $\Gamma(u,1)$ بیانگر همسایگان گره u در لایه شماره ۱ می‌باشد. همچنین بدون از دست رفتن کلیات مسأله، به منظور شهود بیشتر فرض می‌کنیم شبکه مورد نظر دارای دو لایه است و قصد پیش‌بینی ارتباط در لایه اول را داریم. این ایده قابل تعمیم به تعداد دلخواهی از لایه‌ها می‌باشد.

برای پیش‌بینی ارتباط در یک لایه از شبکه ناهمگن می‌توان همسایگان مشترک دو گره را در همان لایه به دست آورد

$$CN1(u,v)=\Gamma(u,1)\cap\Gamma(v,1) \qquad (۴)$$

رابطه (۴) نشان‌دهنده تعداد همسایگان مشترک دو گره در لایه اول است. مرحله بعدی این است که به طریقی تأثیر لایه دوم نیز در پیش‌بینی ارتباط در نظر گرفته شود. برای این منظور ابتدا باید همسایگان مشترک دو گره مورد نظر را در لایه دوم نیز به دست آورد. این مجموعه را با نماد CN۲ نمایش می‌دهیم

$$CN2(u,v)=\Gamma(u,2)\cap\Gamma(v,2) \qquad (۵)$$

کمیت جدیدی که باید تعریف گردد تعداد عناصر یکسان در دو مجموعه حاصل از (۴) و (۵) است. این کمیت را با نماد CN_CN نمایش می‌دهیم

$$CN_CN=|CN1\cap CN2| \qquad (۶)$$

اگر مقدار CN_CN مقداری مخالف صفر باشد، یعنی دو گره دارای همسایگان یکسانی در دو لایه مورد نظر هستند، پس میزان تأثیرپذیری از لایه دوم عددی بزرگ‌تر و برابر (۷) خواهد بود

$$Layer2_effect=\frac{CN_CN}{|CN2|} \qquad (۷)$$

اما اگر مقدار CN_CN برابر صفر باشد، یعنی دو گره در لایه‌های مختلف با افراد مختلف تعامل داشته‌اند و میزان تأثیرپذیری از لایه دوم عددی کوچک‌تر خواهد بود که تنها از چیدمان شبکه تأثیر می‌پذیرد. احتمال این تأثیرپذیری برابر (۸) است

$$Layer2_effect=|CN2|\times correctness(1|2) \qquad (۸)$$

در شبکه‌های اجتماعی واقعی، در صورتی که عدد حاصل از (۷) مقداری مخالف صفر باشد، این عدد در مقایسه با عدد حاصل از (۸) عدد بزرگ‌تری خواهد بود. دلیل این امر این است که تعداد ارتباط‌های موجود در لایه نسبت به همسایگان مشترک گره‌ها در دو لایه عددی بسیار بزرگ است.

پس از این که مقدار تأثیرپذیری از لایه دوم با توجه به مقدار CN_CN مشخص گردید می‌توان رابطه نهایی برای همسایگان مشترک دو گره را به صورت (۹) تعریف نمود

$$CN_Expanded=|CN1|+Layer2_effect \qquad (۹)$$

رابطه (۹) توسعه‌یافته همسایگان مشترک در شبکه‌های ناهمگن است که می‌تواند به عنوان معیار شباهت جدید مورد استفاده قرار گیرد. شبه‌کد روش پیشنهادی در الگوریتم زیر ارائه شده است:

```
Input: G = [G_1, G_2, …, G_k], G_In = Layer of interest
#G is set of available layers for a given Multilayer graph
#G_In is Layer of interest for prediction of links
Output: L: Predicted Links matrix among nodes
Counter = 1;
Calculate interrelation between layer1 and other layers based on (3)
For each(u,v) ∉ E1 do
   Calculate CN1 based on (4) #common neighbors of node u and v
   in G_In
   For each G/G_In #all layers axcept layer1
      Compute CN2 based on (5) #common neighbor of node u and v
      Compute CN_CN based on (6) #common elements of CN1 and
      CN2
      If (CN_CN != 0)
         Effect_of_this_layer = |CN_CN|/|CN2|
      Else
         Effect_of_this_layer = CN2 * Correctness (G_In|this layer)
      End
      Layers_effect+= Effect_of_this_layer
   End
   CN_Expended(u,v) = CN1 + Layers_effect
   S(Counter,1) = u
   S(Counter,2) = v
   S(counter,3) = CN_Expended(u,v)
   Counter++
End
Compute L = f(S) where f is a well-known binary classifier
```

٤-۳ پیاده‌سازی

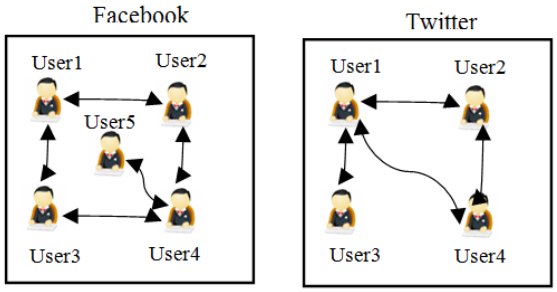
ما در این مقاله ۵ معیار شباهت تعریف‌شده برای شبکه‌های همگن را برای استفاده در شبکه‌های ناهمگن توسعه داده‌ایم که ایده بسط این

Archive of SID

www.SID.ir

جدول ۳: میزان تأثیرپذیری شبکه‌های موجود در شکل ۴.

ارتباطات مشترک	ارتباطات توییتر	ارتباطات فیس‌بوک	تأثیرپذیری شبکه‌ها از هم
۳	۴	۵	$P(Correctness(T\|F))=\frac{۳}{۵}=۰/۶$ $P(Correctness(F\|T))=\frac{۳}{۴}=۰/۷۵$



شکل ۴: مثالی برای نمایش تأثیر روابط بر هم.

جدول ۴: نتایج حاصل از اجرای الگوریتم رگرسیون لجستیک.

مجموعه داده	روش ارزیابی	Direct	Multiplex	Direct_Multiplex	EC-MLP
CKM-Physicians	Precision	۰/۵۲۳۹	۰/۷۶۳۸	۰/۷۵۹۹	۰/۷۶۶۷
	Accuracy	۰/۹۵۹۵	۰/۸۹۵۹	۰/۹۰۰۹	۰/۹۸۵۳
ArXiv	Precision	۰/۶۵۸۲	۰/۷۲۹۰	۰/۷۲۹۹	۰/۷۳۹۳
	Accuracy	۰/۹۷۶۱	۰/۹۹۵۵	۰/۹۵۴۰	۰/۹۹۵۵
شبکه دولایه‌ای	Precision	۰/۴۶۹۸	۰/۵۹۶۵	۰/۵۷۴۲	۰/۵۱۸۳
	Accuracy	۰/۷۹۵۸	۰/۶۱۳۱	۰/۶۳۶۶	۰/۸۲۸۴
شبکه چهارلایه‌ای	Precision	۰/۴۹۸۳	۰/۶۵۸۹	۰/۵۹۴۵	۰/۶۷۴۳
	Accuracy	۰/۷۸۱۳	۰/۷۷۶۲	۰/۵۸۴۴	۰/۸۶۰۳

معیارها مانند ایده شرح داده شده در قسمت قبل می‌باشد. این معیارهای شباهت عبارتند از همسایگان مشترک، ضریب آدامیک و آدار، ضریب جاکارد، پیوست تدریجی و تخصیص منابع.

برای برنامه نویسی از محیط Visual Studio و زبان برنامه‌نویسی C# و محیط برنامه‌نویسی MATLAB استفاده شده است. برای هر جفت گره ۵ معیار شباهت نام برده شده محاسبه گردید و سپس این معیارها به عنوان بردار ویژگی‌ها برای الگوریتم‌های باناظر رگرسیون لجستیک و شبکه عصبی و ماشین بردار پشتیبان در نظر گرفته شده‌اند و مسأله طبقه‌بندی حل شده است. دلیل بهره‌گیری از این جداسازها این است که در مسایلی با دو کلاس، این الگوریتم‌ها از بهترین عملکرد برخوردارند. روش‌های رگرسیون لجستیک و ماشین بردارهای پشتیبان بر اساس یک رویکرد آماری و روش شبکه عصبی بر اساس یک رویکرد هوش محاسباتی بنا نهاده شده‌اند [۳۳]. برای رگرسیون لجستیک از روش گرادیان نزولی برخط برای آموزش پارامترهای مدل بهره گرفته شده است. همچنین در صورت غیر خطی بودن کلاس‌ها از روش کرنل و تابع کرنل پایه‌ای شعاعی بهره گرفته شده است (خواننده می‌تواند برای جزئیات بیشتر به فصل ۷ کتاب [۳۴] رجوع کند). برای شبکه عصبی مورد استفاده در این پژوهش تعداد لایه‌های پنهان مورد استفاده ۷ لایه در نظر گرفته شده و دلیل این امر این است که تعداد ویژگی‌های مورد استفاده کم می‌باشد. برای تعداد ویژگی‌های کم معمولاً تعداد لایه‌های مخفی، تعداد ویژگی‌ها به اضافه ۲ در نظر گرفته می‌شود. هنگامی که این تعداد افزایش یابد، تعداد لایه‌های مخفی مورد استفاده برابر نصف تعداد ویژگی‌های مورد استفاده می‌باشد [۳۵]. برای جداسازی با استفاده از شبکه عصبی پرسپترون پس‌رونده چندلایه و با شرط توقف بر روی میزان خطا و از تابع استاندارد پترن‌نت Matlab استفاده شده است. به منظور جداسازی کلاس‌ها بر اساس الگوریتم ماشین‌های بردارهای پشتیبان از کتابخانه استاندارد Libsvm و بر اساس رویکرد غیر خطی مبتنی بر توابع کرنل گاوسی بهره گرفته شده [۳۶] و برای جزئیات تئوری بیشتر در مورد ماشین‌های بردار پشتیبان خواننده علاقمند می‌تواند به فصل ۵ کتاب [۳۷] مراجعه کند.

۵- آزمایش‌ها و نتایج

۵-۱ مجموعه داده

برای ارزیابی روش ارائه‌شده از دو مجموعه داده استاندارد استفاده شده است. مجموعه داده‌های مورد استفاده در این مقاله مجموعه داده‌های همکاری بین نویسندگان مختلف در حوزه‌های مختلف است [۳۸] و [۳۹]. اولین مجموعه داده مورد استفاده، مجموعه داده CKM-Physicians برای همکاری بین پژوهشگران است که شامل ۲۳۱ نویسنده متمایز است که بین این نویسندگان سه نوع ارتباط دوستی، رایزنی و مباحثه وجود دارد. دومین مجموعه داده استخراج‌شده از پایگاه مقالات علمی ArXive[1] مشتمل بر ۲۹۵۶ پژوهشگر است که در دو شاخه تحلیل داده و بیوفیزیک به همکاری با هم پرداخته‌اند. همچنین یک شبکه ۵۰۰ گرهی شامل دو لایه و شبکه‌ای دیگر دارای ۲۵۰ گره و ۴ لایه با خصوصیت شبکه‌های دنیای واقعی و با توزیع درجه قانون توانی شبیه‌سازی شده است.

۵-۲ نتایج

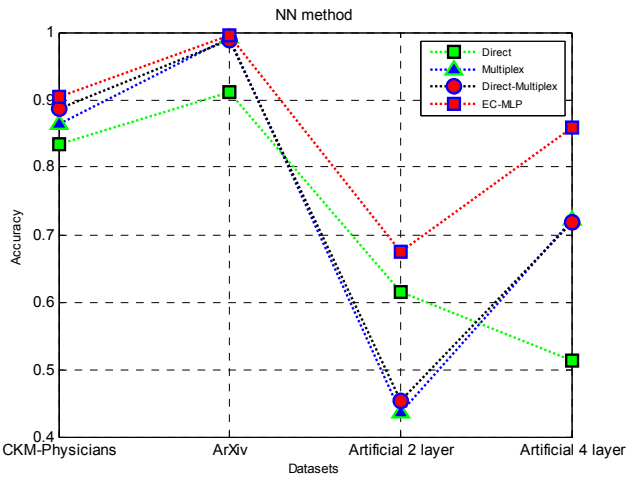
برای مقایسه با کارهای پیشین از روش Direct [۲۳] و روش‌های اخیر Direct_Multiplex و Multiplex ارائه‌شده در [۳۰] استفاده شده است. روش پیشنهادی به صورت EC-MLP در نتایج نمایش داده شده و در هر قسمت بهترین نتایج با رنگ قرمز آمده است.

برای معیارهای ارزیابی، از معیارهای دقت و صحت که معیارهای استاندارد ارزیابی در روش‌های یادگیری باناظر هستند بهره گرفته شده است [۴۰]. بر اساس این معیارها و جداساز رگرسیون لجستیک روش پیشنهادی از دقت بالاتری نسبت به روش‌های دیگر برخوردار است که جدول ۴ این نتایج را نشان می‌دهد. نکته مهم، برتری روش پیشنهادی نسبت به روش Direct است که در آن شبکه همگن فرض شده و

1. https://arxiv.org/

Archive of SID

www.SID.ir



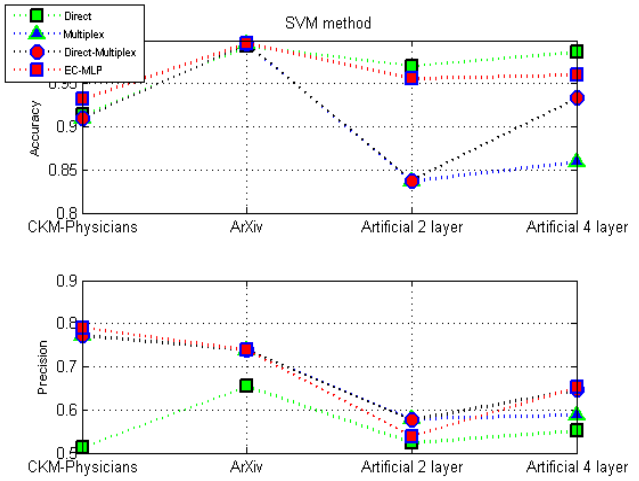
شکل ۵: نتایج حاصل از اجرای روش یادگیری شبکه عصبی با روش‌های مختلف، معیار صحت.

همچنین نسبت به روش‌های اخیر مبتنی بر شبکه‌های ناهمگن، از برتری محسوسی برخوردار می‌باشد.

همچنین در شکل ۵ نتایج روش پیشنهادی در مقایسه با دیگر روش‌ها بر اساس جداساز شبکه عصبی و معیار صحت قرار داده شده است. بر اساس این نتایج، روش پیشنهادی از دقت بالاتری برخوردار است که نشانگر برتری این روش نسبت به روش‌های قبلی است. از طرف دیگر نتایج به دست آمده برای رگرسیون لجستیک و معیار دقت به دلیل تشابه زیاد آورده نشده‌اند. شکل ۶ نشان‌دهنده نتایج حاصل از بهره‌گیری جداساز ماشین‌های بردار پشتیبان بر روی مجموعه داده‌های مورد مطالعه و روش‌های مورد مقایسه است. این نتایج نشانگر برتری روش پیشنهادی یا نتایجی نزدیک به بهترین نتایج در اکثر موارد است. بنابراین با استفاده از نتایج به دست آمده بر روی مجموعه دادگان مختلف می‌توان نتیجه گرفت که روش پیشنهادی عملکرد بهتری نسبت به روش‌های مشابه پیشین دارد.

٦- نتیجه‌گیری

در این مقاله، ابتدا اهمیت موضوع پیش‌بینی ارتباط و کاربردهای مختلف آن بیان شد و رویکردهای مختلف برای حل آن مورد بررسی قرار گرفت. شبکه‌های اجتماعی ناهمگن، چرایی و مزایای آنها مطرح گردید و سپس به صورت رسمی این شبکه‌ها که تعمیمی بر شبکه‌های اجتماعی همگن هستند، معرفی شدند. در ادامه، ابتدا یک معیار شباهت نوین برای شبکه‌های ناهمگن ارائه کردیم. بر اساس این معیار نوین، الگوریتم جدیدی برای پیش‌بینی ارتباط در شبکه‌های ناهمگن بر اساس رویکردهای احتمالاتی ارائه دادیم. این روش قابلیت اجرا روی هر شبکه ناهمگن در هر حوزه‌ای و با هر تعداد لایه دلخواه با انواع مختلفی از گره‌ها و ارتباطات را دارد. بر اساس رویکرد حل مسأله از طریق یادگیری باناظر، از بهترین جداسازهای مسایل دوکلاسه برای آزمون روش پیشنهادی شامل رگرسیون لجستیک، شبکه‌های عصبی پرسپترون چندلایه و بردارهای پشتیبان تصمیم استفاده کردیم. با استفاده از معیارهای استاندارد ارزیابی و بر اساس مجموعه داده‌های مصنوعی و واقعی به ارزیابی روش پیشنهادی با روش‌های مطرح پیشین پرداخته شد که نتایج حاصل نشان از بهبود معنادار پیش‌بینی ارتباط با استفاده از رویکرد پیشنهادی ما داشته است. روش پیشنهادی می‌تواند به عنوان یک روش معتبر برای پیش‌بینی ارتباط در شبکه‌های اجتماعی ناهمگن مورد استفاده قرار گیرد.



شکل ۶: نتایج حاصل از اجرای روش ماشین بردار پشتیبان با روش‌های مختلف.

به عنوان پیشنهاد کارهای آتی می‌توان از پیش‌بینی ارتباط و روش پیشنهادی در مسایلی همچون سامانه‌های توصیه‌گر و حل مسأله پراکندگی ماتریس داده‌ها [۴۱]، توسعه روش پیشنهادی برای شبکه‌های پویا [۴۲] و چگونگی ارائه رویکردی بدون ناظر برای پیش‌بینی ارتباط در شبکه‌های چندلایه [۴۳] می‌تواند زمینه‌هایی باشد که در آینده مورد توجه قرار گیرد.

مراجع

[1] M. Newman, *Networks: An Introduction*, Oxford, New York: Oxford University Press, 2010.

[2] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *Proc. 4th Workshop on Link Analysis, Counter-Terrorism and Security, SDM'06*, 10 pp., 2006.

[3] A. Anil, *et al.*, "Link prediction using social network analysis over heterogeneous terrorist network," in *Proc. IEEE Int. Conf. on Smart City/SocialCom/SustainCom (SmartCity)*, Chengdu, China, 19-21 Dec. 2015.

[4] L. M. Aiello, *et al.*, "Friendship prediction and homophily in social media," *ACM Trans. Web*, vol. 6, no. 2, Article No. 9, Jun. 2012.

[5] L. Lu, C. H. Jin, and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," *Phys. Rev. E*, vol. 80, no. 4, pp. 046122, 26 Oct. 2009.

[6] F. Folino and C. Pizzuti, "Link prediction approaches for disease networks," in *Proc. Int. Conf. on Information Technology in Bio- and Medical Informatics, ITBAM '11*, pp. 99-108, Toulouse, France, 31 Aug.- 1 Sept., 2011.

[7] N. Rummele, R. Ichise, and H. Werthner, "Exploring supervised methods for temporal link prediction in heterogeneous social networks," in *Proc. of the 24th Int. Conf. on World Wide Web*, pp. 1363-1368, Florence, Italy, 18-22 May 2015.

[8] C. A. Bliss, M. R. Frank, C. M. Danforth, and P. S. Dodds, "An evolutionary algorithm approach to link prediction in dynamic social networks," *J. Comput. Sci.*, vol. 5, no. 5, pp. 750-764, Sept. 2014.

[9] P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: the state-of-the-art," *Sci. China Inf. Sci.*, vol. 58, no. 1, pp. 1-38, Jan. 2015.

[10] F. Liu and S. T. Xia, "Link prediction in aligned heterogeneous networks," in T. Cao, *et al.* (eds) *Advances in Knowledge Discovery and Data Mining*, LNCS, vol 9077. pp. 33-44, Springer, Cham, 2015.

[11] G. D. Lyu, C. J. Fan, L. F. Yu, B. X. Xiu, and W. M. Zhang, "Predicting missing links via structural similarity," *Int. J. Mod. Phys. B*, vol. 29, no. 15, Article No. 1550095, Apr. 2015.

[12] M. Jalili, Y. Orouskhani, M. Asgari, N. Alipourfard, and M. Perc, "Link prediction in multiplex online social networks," *Open Sci.*, vol. 4, no. 2, Article No. 160863, Feb. 2017.

[13] Y. Yang, N. Chawla, Y. Sun, and J. Hani, "Predicting links in multi-relational and heterogeneous networks," in *Proc. of the IEEE 12th Int. Conf. on Data Mining*, pp. 755-764, Washington, DC, USA, 10-13 Dec. 2012.

[14] G. Rossetti, M. Berlingerio, and F. Giannotti, "Scalable link prediction on multidimensional networks," in *Proc. IEEE 11th Int.*

Archive of SID

www.SID.ir

Conf. on Data Mining Workshops, ICDMW'11, pp. 979-986, Vancouver, Canada, 11-11 Dec. 2011.
[15] J. Leskovec and J. D. Ullman, *Mining of Massive Datasets*, 2nd Ed., Cambridge: Cambridge University Press, 2014.
[16] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Am Soc Inf Sci Technol*, vol. 58, no. 7, pp. 1019-1031, May 2007.
[17] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Phys. Rev. E*, vol. 64, no. 2, 4 pp., Jul. 2001.
[18] V. Martinez, F. Berzal, and J. C. Cubero, "A survey of link prediction in complex networks," *ACM Comput Surv.*, vol. 49, Article No. 69, Dec. 2016.
[19] T. Sorensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons," *Biol Skr*, vol. 5, pp. 1-34, 1948.
[20] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39-43, Mar. 1953.
[21] F. Fouss, A. Pirotte, J. M. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 3, pp. 355-369, Mar. 2007.
[22] G. Jeh and J. Widom, "SimRank: a measure of structural-context similarity," in *Proc. of the 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining,* pp. 538-543, Edmonton, Alberta, Canada, 23-26 Jul. 2002.
[23] D. Davis, R. Lichtenwalter, and N. V. Chawla, "Multi-relational link prediction in heterogeneous information networks," in *Proc. Int. Conf. on Advances in Social Networks Analysis and Mining*, pp. 281-288, Kaohsiung, Taiwan, 25-27 Jul. 2011.
[24] J. Zhang, P. S. Yu, and Z. H. Zhou, "Meta-path based multi-network collective link prediction," in *Proc. of the 20th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 1286-1295, New York, NY, USA, 24-27 Aug. 2014.
[25] B. R. Memon and U. K. Wiil, "Predicting links in multi-relational networks," in *Proc. IEEE Joint Intelligence and Security Informatics Conf., JISIC'14*, pp. 107-114, Hague, The Netherlands, 24-26 Sept. 2014.
[26] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: meta path-based top-k similarity search in heterogeneous information networks," in *Proc. VLDB Endow.*, vol. 4, no 11, pp. 992-1003, 2011.
[27] J. Tang, T. Lou, and J. Kleinberg, "Inferring social ties across heterogenous networks," in *Proc. of the 5th ACM Int. Conf. on Web Search and Data Mining,* pp. 743-752, Seattle, Washington, USA, 8-12 Feb. 2012.
[28] D. Li, Y. Zhang, Z. Xu, D. Chu, and S. Li, "Exploiting information diffusion feature for link prediction in Sina Weibo," *Sci. Rep.*, vol. 6, Article No. 20058, Jan. 2016.
[29] J. Li, B. Ge, K. Yang, Y. Chen, and Y. Tan, "Meta-path based heterogeneous combat network link prediction," *Phys. Stat. Mech. Its Appl.*, vol. 482, no. Supplement C, pp. 507-523, 15 Sep. 2017.
[30] M. Pujari and R. Kanawati, "Link prediction in multiplex networks," *NHM*, vol. 10, no. 1, pp. 17-35, 2015.
[31] C. Yang, J. Sun, J. Ma, S. Zhang, G. Wang, and Z. Hua, "Scientific collaborator recommendation in heterogeneous bibliographic networks," in *Proc. 48th Hawaii Int. Conf. on System Sciences, HICSS'15,* pp. 552-561, Kauai, HI, USA, 5-8 Jan. 2015.
[32] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Soc. Netw.*, vol. 25, no. 3, pp. 211-230, Jul. 2003.
[33] The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition, New York, NY: Springer, 2016.
[34] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.
[35] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 1st Ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1994.
[36] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol. TIST*, vol. 2, no. 3, Article No. 27, Apr. 2011.
[37] C. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2007.
[38] J. Coleman, E. Katz, and H. Menzel, "The diffusion of an innovation among physicians," *Sociometry*, vol. 20, no. 4, pp. 253-270, Dec. 1957.
[39] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, "Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems," *Phys. Rev. X 5*, arXiv:1408.2925 2015.
[40] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, 2nd Ed., Berlin: Springer Series in Statistics, 2011.
[41] M. Ranjbar, P. Moradi, M. Azami, and M. Jalili, "An imputation-based matrix factorization method for improving accuracy of collaborative filtering systems," *Eng. Appl. Artif. Intell.*, vol. 46, Issue A, pp. 58-66, Nov. 2015.
[42] L. Yao, L. Wang, L. Pan, and K. Yao, "Link prediction based on common-neighbors for dynamic social network," in Proc. 7th Int. Conf. Ambient Syst. Netw. Technol. ANT 2016 6th Int. Conf. Sustain. Energy Inf. Technol. SEIT-2016 Affil. Workshop, vol. 83, Supplement C, pp. 82-89, Jan. 2016.
[43] Y. Dong, N. V. Chawla, and A. Swami, "Metapath2Vec: scalable representation learning for heterogeneous networks," in *Proc. of the 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining,* pp. 135-144, Halifax, NS, Canada, 13-17 Aug. 2017.
[44] E. Zhong, E. W. Xiang, W. Fan, N. N. Liu, and Q. Yang, *Friendship Prediction in Composite Social Networks*, ArXiv Prepr. ArXiv: 14024033, 2014.
[45] B. Pachev and B. Webb, Fast Link Prediction for Large Networks Using Spectral Embedding, ArXiv Prepr. ArXiv: 1703.09693, 2017.

سعیده رضاوندی شعاعی در سال ۱۳۹۰ مدرک کارشناسی مهندسی نرم‌افزار خود را از دانشگاه علمی و کاربردی تراکتورسازی تبریز و در سال ۱۳۹۴ مدرک کارشناسی ارشد مهندسی فناوری اطلاعات خود را از دانشگاه تهران دریافت نمود. از سال ۱۳۸۴ تا سال ۱۳۹۵ به عنوان برنامه‌نویس به صورت فری‌لنسر مشغول به فعالیت بود، وی از سال ۱۳۹۵ تا کنون به عنوان برنامه‌نویس در یک شرکت خصوصی مشغول به کار است. همچنین از نیمسال اول سال ۱۳۹۷ به عنوان مربی در دانشگاه غیرانتفاعی ارشد دماوند مشغول به تدریس است. زمینه‌های تحقیقاتی ایشان عبارتند از: پیش‌بینی ارتباط در انواع شبکه‌ها، تحلیل شبکه‌های اجتماعی، تشخیص انجمن در شبکه‌های اجتماعی و تشخیص سرقت علمی در مقالات می‌باشد.

هادی زارع تحصیلات خود را در مقطع کارشناسی آمار ریاضی در سال ۱۳۸۴ از دانشگاه شیراز و کارشناسی ارشد در سال ۱۳۸۶ در رشته آمار و دکتری خود را در ریاضیات کاربردی از دانشگاه صنعتی امیرکبیر در سال ۱۳۹۰ دریافت کرده است. از سال ۱۳۹۱ تاکنون در دانشکده علوم و فنون نوین دانشگاه تهران به عنوان استادیار مشغول به کار می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: تحلیل شبکه‌های اجتماعی، یادگیری ماشین آماری، مدلهای گرافی احتمالاتی و روش‌های خلاصه‌سازی دادگان انبوه می‌باشد.

Archive of SID
www.SID.ir