4th International Conference on Web Research
April 25-26, 2018; Tehran, IRAN
ICWR2018
کنفرانس بین المللی وب پژوهی
چهارمین
۵ و ۶ اردیبهشت ۱۳۹۷؛ تهران، ایران

Link Prediction on Social Networks Based on Deep Learning

[1] Mohammad Mehdi Keikha, [2]Maseud Rahgozar

[1] PhD Candidate, University of Tehran, Tehran, Iran
Mehdi.keikha@ut.ac.ir

[1] Faculty member, University of Sistan and Baluchestan, Zahedan, Iran

[2] Associate Professor, University of Tehran, Tehran, Iran
rahgozar@ut.ac.ir

Abstract

Link prediction on social networks is one of the issues that has attracted many researchers in recent years. In this problem, missing and future links are predicted by using existing links in the. One of the newest approaches to this problem is the use of deep learning to extract the vector of the features of each node and then find missing and future links. This paper presents a method for learning the vector representation of network nodes based on the information of the nodes adjacent to each node in the social network and the various links present on the network. The results show that the proposed method provides good results for link prediction in comparison with other methods.

Keywords: Link prediction, Deep Learning, feature vector of node, network communities, social networks.

Archive of SID

www.SID.ir

4th International Conference on Web Research
April 25-26, 2018; Tehran, IRAN
ICWR2018
کنفرانس بین المللی وب پژوهی
چهارمین
۵ و ۶ اردیبهشت ۱۳۹۷؛ تهران، ایران

پیش بینی روابط در شبکه های اجتماعی مبتنی بر یادگیری عمیق

محمد مهدی کیخا[۱]، مسعود رهگذر[۲]

[۱] دانشجوی دکترا، دانشکده مهندسی برق و کامپیوتر ، دانشگاه تهران، تهران، ایران
عضو هیئت علمی، دانشگاه سیستان و بلوچستان، زاهدان، ایران
mehdi.keikha@ut.ac.ir

[۲] دانشیار، دانشکده مهندسی برق و کامپیوتر ، دانشگاه تهران، تهران، ایران
rahgozar@ut.ac.ir

Archive of SID

چکیده

پیش بینی ارتباطات در شبکه های اجتماعی، یکی از مسائلی است که در سال‌های اخیر مورد توجه بسیاری از محققین قرار گرفته است. در این مساله سعی بر اینست که به کمک ارتباطات موجود در شبکه، ارتباطات از بین رفته و یا ارتباطات آتی در شبکه شناسایی شوند. یکی از جدیدترین رویکردهای این مسئله، استفاده از یادگیری عمیق جهت استخراج بردار ویژگی های هر گره و یافتن ارتباطات ناشناخته برای آنهاست. در این مقاله روشی برای یادگیری بردار نمایش گره های شبکه بر اساس اطلاعات گره های مجاور هر گره در شبکه اجتماعی و جوامع مختلف موجود در شبکه ارائه شده است. نتایج بدست آمده حاکی از این است که روش ارائه شده در مقایسه با سایر روش ها، نتایج خوبی برای پیش بینی ارتباطات ارائه نموده است.

کلمات کلیدی

پیش بینی ارتباطات، یادگیری عمیق، بردار نمایش گره، جوامع گراف، شبکه اجتماعی.

۱- مقدمه

شبکه های اجتماعی آنلاین مانند فیس بوک، توییتر، تلگرام و غیره خدمات مختلفی را برای تبادل اطلاعات ارائه می دهند. با توجه به ماهیت بسیار پویای شبکه های اجتماعی، تجزیه و تحلیل آنها در سال های اخیر توجه بیشتری را به خود جلب کرده است. در سال های اخیر بررسی های بسیار مختلفی بر روی شبکه های اجتماعی به منظور استخراج دانش ارزشمند موجود در آنها انجام شده است. یکی از مهمترین مسائل در حوزه شبکه اجتماعی پیش بینی ارتباطات است که تلاش می کند ارتباطات پنهانی موجود در شبکه اجتماعی را بر اساس اطلاعات مختلف موجود از جمله ارتباطات و متون منتشر شده توسط هر گره در شبکه اجتماعی پیدا کند [1]. حل مساله پیش بینی ارتباطات کمک شایانی جهت درک نحوه تکامل شبکه و کشف ارتباطات نهفته در شبکه فعلی می نماید. پیش بینی پیوند در برنامه های متنوع مانند شبکه های بیولوژیکی [2]، پیشنهاد دوستی [3, 4]، پیشنهاد دهی کالاها [5, 6]و غیره استفاده شده است.

پیش بینی ارتباطات یکی از مسائل بسیار مهم و قدیمی در حوزه شبکه های اجتماعی است که با استخراج یک مدل از بر اساس اطلاعات گوناگون افراد در شبکه اجتماعی، سعی در شناسایی ارتباطات از بین رفته و ارتباطات آتی دارد.

در سال های اخیر، روش های بسیاری برای پیش بینی ارتباطات در یک شبکه اجتماعی پیشنهاد شده اند وانگ و همکاران در [7]، روش های پیش بینی ارتباطات را به ۳ دسته تقسیم نموده‌اند که عبارتند از: روش های مبتنی بر گره های شبکه، روش های مبتنی بر ساختار شبکه و روش های مبتنی بر تئوری های اجتماعی.

روش های مبتنی بر گره فرض می کنند که بین دو گره مشابه، احتمال وجود ارتباط بیشتر از سایر گره ها می‌باشد. این روش ها از اطلاعات مختلفی مانند پروفایل، محتویات تولید شده توسط هر کاربر و غیره استفاده می کنند تا

www.SID.ir

شباهت میان دو گره را محاسبه نمایند [8, 9, 10] . در این روش ها، پس از محاسبه شباهت بین جفت گره های گوناگون در شبکه ، بالاترین جفت های مشابه، به عنوان پیوند های از بین رفته یا آینده شناخته می شوند.

رویکردهای مبتنی بر ساختار از اطلاعات ساختاری شبکه برای اندازه گیری شباهت گره ها استفاده می کنند. برخی از روش های مبتنی بر ساختار، تنها همسایه‌های مشترک دو گره را به عنوان میزان شباهت [1, 11, 12] در نظر می گیرند. روش های مبتنی بر مسیر که تعداد مسیر بین دو گره را شمارش می کنند، از جمله دیگر روش های مبتنی بر ساختار هستند [13]. پیچیدگی زمان روش های مبتنی بر شمارش تعداد مسیرها بیشتر از روش های مبتنی بر شمارش تعداد همسایه‌های مشترک است، زیرا آنها از اطلاعات ساختاری کلی شبکه ورودی برای اندازه گیری شباهت میان گره ها استفاده می کنند. قدم زنی های تصادفی یکی دیگر از روش های مبتنی بر ساختار است که ابتدا احتمال بازدید از همسایه‌های گره فعلی را محاسبه می کند. سپس، از آنها به عنوان میزان احتمال وجود ارتباط بین دو گره استفاده می کند [1]. دسته سوم روش های پیش بینی ارتباطات، روش های مبتنی بر تئوری های اجتماعی است که از همبستگی میان چندین گره، روابط سه گانه یا مثلثی میان سه گره، توازن ساختاری مجموعه ای از گره ها و غیره به عنوان مبنایی برای یافتن شباهت گره ها [14] استفاده می نماید.

آخرین گروه از روش های پیش بینی ارتباطات، روش های مبتنی بر یادگیری ماشین است که بردار ویژگی هر یک از ارتباطات موجود در شبکه را بر اساس صفات مختلف گره ها [15] استخراج می نماید.

گرچه، تحقیقات بسیاری در حوزه پیش بینی ارتباطات انجام شده است، اما تمام روش های پیشنهادی برای یک شبکه خاص مناسب هستند و کارایی مناسبی در تمامی انواع شبکه ها از خود بروز نمی دهند. همچنین استخراج ویژگی های هر گره، فرایند بسیار مهمی ایت که تاثیر مهمی در نتایج ارزیابی ها دارد. در این مقاله، ما یک رویکرد جدید برای حل مشکل پیش بینی ارتباطات با استفاده از تکنیک های یادگیری عمیق پیشنهاد می کنیم. زیرا روش های مبتنی بر یادگیری عمیق قادر به استخراج ویژگی های مناسب می باشند.

روش های مبتنی بر یادگیری عمیق یکی از روش های کاهش ابعاد و استخراج ویژگی ها است که در سال های اخیر در کاربردهای گوناگون بسیار مورد استفاده قرار گرفته است[16, 17]. این روش ها ابتدا گراف ارتباطات گره ها را به عنوان ورودی دریافت می نمایند و سپس با استخراج ویژگی های مختلف توسط یادگیری عمیق، سعی در یادگیری بردار نمایش گره های شبکه می نمایند. هنگامی که هر گره شبکه توسط یک بردار با ابعاد بسیار کمتر نمایش داده شد، از این بردارها جهت حل مساله پیش بینی ارتباطات استفاده می نماییم [18-21] . در این مقاله، با استفاده از یک روش بردار نمایش گره های شبکه، یک روش جدید برای حل مساله پیش بینی ارتباطات ارائه شده است. در روش ارائه شده هم از اطلاعات ساختار محلی یک گره مانند همسایه‌های مستقیم در شبکه و هم از اطلاعات ساختاری کلی شبکه مانند جوامع موجود در شبکه استفاده شده است. ادامه¬ی مقاله به صورت زیر سازماندهی شده است:

در بخش ۲ به مرور کارهای انجام شده در حوزه پیش بینی ارتباطات توسط روش های یادگیری بردار نمایش گره های شبکه خواهیم پرداخت. روش پیشنهاد¬ی برای حل مساله پیش بینی ارتباطات در بخش سوم و نتایج آزمایش‌ها صورت گرفته و تحلیل آن¬ها در بخش چهارم ارائه شده است.. در نهایت در بخش ۵ نیز به نتیجه گیری و ارائه مسیرهای تحقیقاتی مرتبط با این مساله می پردازیم.

۲- تحقیقات پیشین

همانطور که در بخش قبل به آن اشاره شد، دسته ای از روش های پیش بینی ارتباطات ، با استفاده از یادگیری عمیق ابتدا یک بردار نمایش برای هر یک از گره ها ارائه می نمایند و سپس با استفاده از یک کلاس بند ساده، ارتباطات جدید را پیش بینی می نمایند. در ادامه چندین روش مختلف برای یادگیری بردار نمایش گره ها در شبکه را مرور خواهیم نمود که نتایج خوبی در پیش بینی ارتباطات ارائه نموده‌اند.

الگوریتم Deepwalk یکی از اولین تحقیقات صورت گرفته در زمینه نمایش گراف توسط شبکه های عصبی عمیق می‌باشد [18] . در این روش از قدم زنی های تصادفی متوالی به عنوان دانش زمینه ای برای یافتن نمایش یک گره استفاده می گردد. در این روش به دلیل آنکه از قدم زنی های تصادفی در گراف استفاده می گردد و اطلاعات جوامع گراف را پوشش داده نمی شود، نمی تواند ساختار کلی گراف را منعکس نماید و در نتیجه یک نمایش کامل برای تمامی گره های گراف ارائه نخواهد شد. از جمله دیگر معایب این روش می توان به عدم کارایی آن برای گراف های خلوت اشاره نمود. در [19] از همسایه‌های مستقیم و درجه دوم (همسایه همسایه) گره های گراف برای نمایش گره های آنها استفاده می شود. روش ارائه شده در این مقاله اگرچه توانسته است خاصیت غیر خطی بودن و خلوت بودن گراف های واقعی را در نظر بگیرد اما ارتباط هر گره را فقط محدود به همسایه های نزدیک به آن دو نموده است حال آنکه دو گره می توانند عضو یک جامعه باشند اما در گراف فاصله بیشتر از ۲ با یکدیگر داشته باشند. در واقع این مقاله، هر دو گرهی که با یکدیگر ارتباط مستقیم داشته باشند یا چندین همسایه مشترک داشته باشند را به عنوان دو گره مشابه از نظر ساختار در نظر می گیرد و تنها از این اطلاعات جهت ایجاد بردار نمایش هر گره استفاده می نماید حال آنکه ممکن است دو گره در قسمت های مختلف گراف حضور داشته باشند اما از لحاظ نقش ساختاری و یا هموفیلی (جوامع) مشابه یکدیگر باشند. روش Node2vec که به تازگی توسط لسکوک[20] ارائه شده است از قدم زنی های تصادفی استفاده نموده است. این روش نیز اگرچه مقیاس پذیری بالایی دارد و از دو استراتژی جستجوی BFS و DFS برای قدم زنی های تصادفی استفاده می نماید، اما نمی تواند تمامی اطلاعات ساختاری محلی و کلی گراف را حفظ نماید زیرا فقط از روابط رتبه اول, و دوم استفاده می نماید و از اطلاعات جوامع گره ها استفاده نمی کند. به همین دلیل برای نمایش گره¬ها، اطلاعات زیادی در رابطه با سختار کلی گراف را در نظر نمی¬گیرد، همچنین این روش نیز مانند Deepwalk برای گراف های خلوت کارایی بالایی ندارد. اخیرا Wang و همکاران در [21] از روش های فاکتورگیری ماتریسی جهت یادگیری نمایش گره ها استفاده نموده‌اند. آنها دو تابع مختلف جهت استخراج اطلاعات جوامع و همسایه‌های تعریف می نمایند که این دو تابع به طور مستقل از یکدیگر بردار نمایش یک گره را می سازند. از آنجایی که دو تابع به طور مستقل از یکدیگر عمل می کنند و عملیات روی ماتریس ها در فاکتورگیری ماتریسی بسیار هزینه بر است، این روش مقیاس پذیری بالایی نداشته و برای گراف های بزرگ اصلا کارایی ندارد.

Archive of SID
www.SID.ir

همانطور که گفته شد، تحقیقات زیادی در زمینه پیش بینی ارتباطات وجود دارد اما هیچ معماری وجود ندارد که از تمام اطلاعاتی که در شبکه ارائه شده است استفاده نماید.

۳- روش پیشنهادی

همانگونه که در بخش قبل مشاهده نمودیم، تمامی روش های یادگیری بردار گره های شبکه به کمک یادگیری عمیق، ابتدا تعدادی از گره های مرتبط با هر گره را توسط استراتژی های مختلف استخراج می نمایند و سپس از این گره ها به عنوان دانش زمینه برای یادگیری بردار نمایش هر گره استفاده می کنند. تفاوت روش های مختلف ارائه شده، در نحوه استخراج گره های مرتبط با یک گره خاص می‌باشد. اکثر روش های ارائه شده فقط از همسایه‌های مستقیم و همسایه‌های درجه دوم به عنوان گره های مرتبط استفاده می نمایند. حال آنکه روش پیشنهادی ارائه شده که به اختصار CARE[۱] نامیده می شود، ابتدا برای هر گره تعدادی مسیر از گره های مرتبط با آن ایجاد می نماید. هنگام تشکیل مسیرها، از اطلاعات جوامع موجود در گراف نیز استفاده می گردد و بر اساس یک پارامتر، در مسیر ایجاد شده برای هر گره از برخی گره ها که در همسایگی گره نمی باشند اما با گره مورد نظر در یک جامعه قرار دارند، نیز به عنوان دانش زمینه ای برای یادگیری نمایش برداری آن گره استفاده می شود. با توجه به اینکه اطلاعات جوامع یک گراف، حاوی ساختار کلی گراف است و اکثر روش های پیشین از این اطلاعات هیچگونه استفاده ای ننموده‌اند، استفاده از این اطلاعات در هنگام یادگیری بردار نمایش گره های شبکه می تواند نتایج خوبی داشته باشد.

در ابتدا برای تعیین برچسب گره های شبکه، از الگوریتم لوین برای تشخیص جوامع مختلف شبکه ورودی استفاده شده است[22] .

این روش به کمک بیشینه نمودن خاصیت پیمانه ای در شبکه سعی در تشخیص جوامع آن دارد. خاصیت پیمانه ای یک معیار برای مقایسه چگالی گره هایی است که با گره های داخل یک جامعه نسبت به گره های خارج از آن جامعه ارتباطات بیشتری دارند. لوین برای حل این مساله بهینه سازی ابتدا تمامی گره ها را به کل جوامع شبکه اختصاص می دهد و سپس در تکرارهای مختلف الگوریتم، هر گره را به جامعه جدیدی اختصاص می دهد به نحوی که در آن جامعه، مقدار پیمانه ای بودن بودن برای آن گره حداکثر شود.

پس از اینکه جوامع مختلف گراف ورودی شناسایی شد، اقدام به تولید مسیرهای ویژه ای می شود که علاوه بر اطلاعات همسایه‌های یک گره، اطلاعات گره هایی که با گره مبدا مسیر در یک جامعه قرار دارند را نیز شامل شود. برای تولید مسیرها، برای هر گره مشاهده شده در طی روند ایجاد مسیر، با احتمال α یکی از همسایه‌های مستقیم گره مشاهده شده را انتخاب و با احتمال α -1 نیز یکی از گره هایی برگزیده می شود که با گره مشاهده شده در یک جامعه قرار گرفته باشد. به این ترتیب CARE قادر خواهد بود هم ساختار محلی یک گره و هم ساختار کلی گراف را هنگام استخراج ویژگی های یک گره مد نظر قرار دهد. برای یادگیری بردار نمایش هر گره با استفاده از مسیرهای ویژه ایجاد شده از الگوریتم Skip-Gram استفاده می شود که پیش از این در پردازش زبان طبیعی از آن استفاده شده است [23]. الگوریتم ۱ مراحل انجام این روش را تشریح می نماید:

Algorithm 1: CARE (G, w, d, μ, l)
Input:
Graph G (V, E)
Window length w
Representation size d
Number of random walks per node μ
Random walk max length l
Output:
Matrix of node representations $f \in \mathbb{R}^{|V| \times d}$
1: Com = ModularityMaximization(G)
2: sample f from $\mathbb{R}^{|V| \times d}$
3: while (i < μ)
4: s = shuffle (V)
5: for each vi $\in$ s do
6: $\mathcal{W}_{v_i}$ = CommunityawareRW (G, vi , Com, l)
7: SkipGram (f, $\mathcal{W}_{v_i}$, w)
8: end for
9: end while

همانگونه که در الگوریتم CARE مشاهده می شود، در گام ۱ ابتدا جوامع مختلف گراف ورودی تشخیص داده می شوند. سپس در گام ۲ بردار نمایش هر گره مقدار دهی اولیه می شود. سپس برای هر گره در شبکه به تعداد μ مسیر ویژه ایجاد می گردد تا اطمینان حاصل شود که تمامی اطلاعات ساختار محلی و ساختار کلی گراف برای هر گره حفظ خواهد شد. برای جلوگیری از وابستگی روش به ترتیب مشاهده گره های شبکه، در گام ۴ ترتیب آنها بصورت تصادفی تغییر داده می شود. در گام ۶ به ازای هر گره یک مسیر تصادفی ایجاد می شود که این مسیر حاوی اطلاعات همسایه‌های و اطلاعات گره هایی است که با گره های مشاهده شده در یک جامعه قرار دارند. سپس مسیر ایجاد شده برای یک گره به الگوریتم Skip-Gram تحویل داده می شود تا به کمک این روش، بردار نمایش هر گره یادگرفته شود.

پس از آنکه ویژگی های هر گره استخراج شد، الگوریتم CARE بردار ویژگی های دو گره را توسط عملگر هادامارد در هم ضرب نموده تا بردار ویژگی های هر یال از گراف مطابق رابطه (۱) بدست آید.

$$Hadamard(u.v): f(u)_i * f(v)_i \qquad (۱)$$

در رابطه (۱) $f(u)_i$. $f(v)_i$ ویژگی i ام گره های u v, است.
البته می توانیم به جای عملگر هادامارد از سایر عملگرها مانند جمع یا میانگین نیز برای استخراج بردار ویژگی یالها بر اساس بردار ویژگی گره ها استفاده نمود، اما طبق آزمایش‌ها و ارزیابی‌های انجام شده در [20] عملگر هادامارد بهترین نتایج را در مساله پیش بینی ارتباطات ارائه نموده است.

٤- نتایج عملی و ارزیابی

جهت ارزیابی روش پیشنهادی برای مساله ارتباطات، از چندین الگوریتم استخراج بردار نمایش گره های شبکه استفاده شده است که این الگوریتم ها عبارتند از:

الگوریتم همسایه‌های مشترک (CN) [11] : این روش با شمارش تعداد همسایه های مشترک میان دو گره، احتمال ایجاد ارتباط میان دو گره را محاسبه می نماید. این روش یکی از بهترین روش های پیش بینی ارتباطات است که مبتنی بر ساختار شبکه می‌باشد.

Archive of SID
www.SID.ir

الگوریتم Spectral clustering [24] : این الگوریتم با استفاده از برش های[۲] مختلف گراف، سعی در شناسایی برچسب های مختلف یک گره می کند. برای یافتن برش های مختلف گراف، ابتدا ماتریس لاپلاسین نرمال شده گراف را محاسبه می نماید و سپس d کوچکترین مقدار بردار ویژه[۳] را به عنوان بردار نمایش هر گره انتخاب می نماید.

الگوریتم DeepWalk [18]: این روش نخستین الگوریتمی است که با استفاده از بردار نمایش گره ها که توسط یادگیری عمیق بدست می آورد برای حل مساله دسته بندی چند برچسبی استفاده می نماید. این روش در ابتدا به ازای هر گره، تعدادی قدم زنی تصادفی ایجاد می کند و از این مسیرهای تصادفی به عنوان اطلاعات زمینه برای یادگیری بردار نمایش گره ها در فضای d بعدی استفاده می کند. در نهایت با استفاده از یک دسته بند ساده اقدام به شناسایی برچسب های یک گره می کند.

الگوریتم LIINE [19]: این روش دو تابع جداگانه برای جمع آوری اطلاعات همسایه‌های مستقیم و همسایه‌های درجه دوم گره استفاده می کند و از اطلاعات جمع آوری شده توسط این دو تابع جهت یادگیری بردار نمایش گرهه ها استفاده می کند.

الگوریتم Node2vec [20]: این روش نیز مانند DeepWalk از قدم زنی های تصادفی برای جمع آوری اطلاعات گره ها استفاده می نماید. تفاوت اصلی آن با DeepWalk در این است که این روش فقط از همسایه‌های مستقیم و همسایه‌های درجه دوم گره برا ایجاد مسیرهای تصادفی استفاده می کند.

برای ارزیابی کامل CARE از دو مجموعه داده استاندارد استفاده شده است که اطلاعات آنها در جدول زیر ارائه شده است:

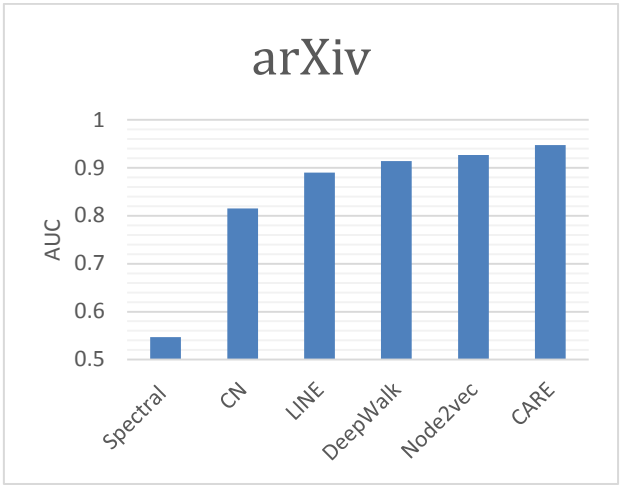
جدول ۱: مشخصات مجموعه داده های مورد استفاده در آزمایش‌ها

نام مجموعه داده	تعداد گره ها	تعداد یالها
PPI [25]	19706	390633
arXiv ASTRO-PH [26]	18772	198110

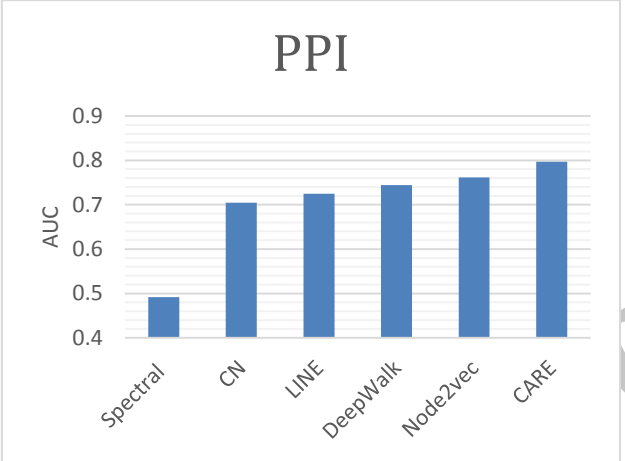
جهت ارزیابی روش پیشنهادی از معیار AUC استفاده می نماییم که پیش از این نیز در [20] جهت ارزیابی پیش بینی ارتباطات مورد استفاده قرار گرفته است. معیار AUC بیانگر میزان دقتی است که احتمال وجود ارتباط بین دو گره توسط روش پیشنهادی را با احتمال پیش بینی ارتباط به صورت تصادفی مقایسه می نماید و در واقع کیفیت روش پیشنهادی را در مقایسه با روش تصادفی اررزیابی می‌نماید.

در ادامه مقدار AUC روش های ذکر شده در مجموعه داده های مختلف در شکل های ۱ و ۲ نمایش داده شده است.

همانگونه که در شکل های ۱ و ۲ مشاهده می گردد، روش CARE در هر دو مجموعه داده در مقایسه با سایر روش های پیش بینی ارتباطات عملکرد بهتری را نشان می دهد. روش همسایه مشترک نسبت به روش Spectral clustering نتایج بهتری داشته است که دلیل آن تضمین در نظرگرفتن اطلاعات ساختاری محلی می‌باشد. همانگونه که اشاره شد روش Spectral clustering با استفاده از بردار ویژه ماتریس مجاورت اقدام به



شکل ۱: مقایسه روش های مختلف پیش بینی ارتباطات توسط arXiv



شکل ۲: مقایسه روش های مختلف پیش بینی ارتباطات توسط PPI

استخراج ویژگی های گره های شبکه می نماید و هیچ تضمینی مبنی بر حفظ ویژگی های ساختاری محلی و کلی شبکه وجود ندارد.

روش های DeepWalk و Node2vec برای استخراج بردار ویژگی هر گره، مسیرهایی به طور تصادفی تولید می کنند و هیچ اطلاعاتی در مورد جوامع گره ها در آنها وجود ندارد. اگر چه، Node2vec تلاش می کند تا با استفاده از دو پارامتر کنترل شده، اطلاعات جوامع را مورد بررسی قرار دهد، اما این پارامترها نمی توانند اطمینان حاصل کنند که اطلاعات مربوط به گره ها را در یک مسیر تصادفی گنجانده اند. در مقابل Node2vec، CARE این اطلاعات را به یک پیاده روی تصادفی سفارشی جاسازی می کند در حالیکه با احتمال α به گره هایی که در همان جامعه با آخرین گره در طول مسیر هستند پرش می کنیم.

۵- نتیجه‌گیری

در این مقاله روشی برای حل مساله پیش بینی ارتباطات مبتنی بر یادگیری بردار نمایش گره های شبکه ارائه شده است. روش پیشنهادی هنگام یادگیری بردار نمایش گره های شبکه هم ساختار محلی شامل همسایه‌های و هم ساختار کلی گراف شامل جوامع موجود در گراف را در نظر می گیرد. نتایج ارزیابی‌های انجام شده بر روی دو مجموعه داده استاندارد نشان از عملکرد

Archive of SID
www.SID.ir

مطلوب روش پیشنهادی در مقایسه با سایر روش های نمایش بـرداری بـرای حل مساله پیش بینی ارتباطات است.

مراجع

[1] Liben-Nowell, D., and Kleinberg, J. M., "*The link-prediction problem for social networks*", Journal of the American Society for Information Science and Technology, Vol. 58, pp. 1019-1031, 2007.

[2] Almansoori, W., Gao, S., Jarada, T. N., and et al. *"Link prediction and classification in social networks and its application in healthcare and systems biology"*. Network Modeling Analysis Health Informatics Bioinformatics, Vol. 1, pp. 27-36, 2012.

[3] Aiello, L. M., Barrat A., Schifanella, R., and et al. "*Friendship prediction and homophily in social media*", ACM Transactions on the Web, Vol. 6, No. 9, 2012.

[4] Mori, J., Kajikawa, Y., Kashima, H., and et al. "*Machine learning approach for finding business partners and building reciprocal relationships"*, Expert Systems with Applications, Vol. 39, pp. 10402-10407, 2012.

[5] Akcora, C. G., Carminati, B., Ferrari, E. "Network and profile based measures for user similarities on social networks", In Proc. of the 12th IEEE International Conference on Information Reuse and Integration, Las Vegas, USA, 2011.

[6] Tang, J., Wu, S., Sun, J. M., and et al. "Cross-domain collaboration recommendation", In Proc. of the 18th ACM SIGKDD International Conference on Discovery and Data Mining (KDD'12), Beijing, China, pp. 1285-1293, 2012.

[7] Wang, P., Xu, B., Wu, Y., and Zhou, X. "*Link prediction in social networks: The state-of-the-art*". Science China Information Sciences, Vol. 58, No. 1, pp. 1–38, 2015.
Anderson, A., Huttenlocher, D., Kleinberg, J., and et al. "Effects of user similarity in social media", In Proc. of the 5th ACM International Conference on Web Search and Data Mining (WSDM'12), Seattle, USA, pp. 703-712, 2012.

[8] Bhattacharyya, P., Garg, A., and Wu, F. S. "*Analysis of user keyword similarity in online social networks*", Social Network Analysis and Mining, Vol. 1, pp. 143-158, 2011.

[9] Akcora, C. G, Carminati, B., and Ferrari, E. "*User similarities on social networks*", Social Network Analysis and Mining, Vol. 3, pp. 475-495, 2013.

[10] Adamic, L. A., and Adar, E. "*Friend and neighbors on the web*", Social Networks, Vol. 25, pp. 211-230, 2003.

[11] Newman, M. E. J, "*Clustering and preferential attachment in growing networks*", Physical Review Letters E, Vol. 64, 2001.

[12] Lichtenwalter, R. N., and Chawla, N. V. "Vertex collocation profiles: subgraph counting for link analysis and prediction", In Proc. of the 21st World Wide Web Conference(WWW'12), Lyon, France, pp. 1019-1028, 2012.

[13] Liu, H., Hu, Z., Haddadi, H., and et al. "*Hidden link prediction based on node centrality and weak ties*". EPL (Europhysics Letters), Vol. 101, 2013.

[14] Wohlfarth, T., and Ichise, R., "Semantic and event-based approach for link prediction", In Proc. of the 7th International Conference on Practical Aspects of Knowledge Management (PAKM'08), Yokohama, Japan, pp. 50-61, 2008.

[15] Bengio, Y., Courville, A., and Vincent, P., "*Representation learning: A review and new perspectives"* IEEE TPAMI, Vol. 35, No. 8, pp.1798–1828, 2013.

[16] Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., Norick, B., and Han, J., "Personalized entity recommendation: A heterogeneous information network approach," In WSDM, pp. 283-292, 2014.

[17] Perozzi, B., Al-Rfou, R., and Skiena, S,. "DeepWalk: Online learning of social representations," In KDD, 2014.

[18] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q., "LINE: Large-scale Information Network Embedding," In WWW, 2015.

[19] Grover, A., and Leskovec, J., "Node2vec: Scalable feature learning for networks," In Proc. of the 22nd ACM SIGKDD, pp. 1225–1234, 2016.

[20] Wang, X., Cui, P., Wang, J., Pei, J., Zhu, W., and Yang, S., "Community preserving network embedding," in AAAI, pp. 203–209, 2017.

[21] Blondel, V. D., Guillaume, J., Lambiotte, R., Lefebvre, E., "Fast unfolding of communities in large networks" Journal of Statistical Mechanics: Theory and Experiment, Vol. 10, 2008.

[22] Mikolov, T., Chen, K., Corrado, G., and Dean, J., "Efficient estimation of word representations in vector space," In ICLR, 2013.

[23] Breitkreutz, B. J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D. H., Bähler, J., and Wood, V., "*The BioGRID interaction database*" Nucleic acids research, Vol. 36, pp. 637–640, 2008.

[24] Mahoney, M., Large text compression benchmark. www.mattmahoney.net/dc/textdata, 2011.

[25] Leskovec, J., and Krevl, A., SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.

زیرنویس

[1] Community aware random walk for network embedding

[2] Graph cuts

[3] Eigen vectors

Archive of SID
www.SID.ir