第 45 卷 第 3 期 计 算 机 工 程 2019 年 3 月
Vol. 45 No. 3 Computer Engineering March 2019

·人工智能及识别技术· 文章编号：1000-3428(2019)03-0197-05 文献标志码：A 中图分类号：TP391

一种时序有向社会网络中的链路预测算法

杨瑞琪，张月霞

（北京信息科技大学 信息与通信工程学院，北京 100101）

摘 要： 针对现有多数链路预测算法准确度较低的问题，在网络全局与局部特征相结合基础上，提出一种基于归一化共同邻居与局部聚类（NCNLC）相似性指标的链路预测算法。通过分析节点 NCNLC 属性的相似性指标，为节点间的连边分配累积影响因子。仿真结果表明，与 LAS 指标相似性算法相比，该算法具有较高的预测准确度，能够有效地进行时序有向社会网络中的链路预测。

关键词： 时序有向社会网络；链路预测；共同邻居；局部聚类；相似性指标

中文引用格式： 杨瑞琪，张月霞. 一种时序有向社会网络中的链路预测算法[J]. 计算机工程，2019，45(3)：197-201.

英文引用格式： YANG Ruiqi，ZHANG Yuexia. A link prediction algorithm for temporal directed social network[J]. Computer Engineering，2019，45(3)：197-201.

A Link Prediction Algorithm for Temporal Directed Social Network

YANG Ruiqi，ZHANG Yuexia

（School of Information and Communication Engineering，Beijing Information Science and Technology University，Beijing 100101，China）

【Abstract】 Aiming at the problem that the existing majority link prediction algorithm has low accuracy，a link prediction algorithm based on the Normalized Common Neighbor and Local Clustering（NCNLC）similarity index is proposed based on the combination of global and local features. The similarity index of NCNLC attribute of the node is analyzed，and the cumulative influence factor is assigned to the joint between the nodes. Simulation results show that compared with the LAS index similarity algorithm，the proposed algorithm has higher prediction accuracy and can effectively predict the link in the temporal directed social network.

【Key words】 temporal directed social network；link prediction；common neighbor；local clusteing；similarity index

DOI： 10.19678/j.issn.1000-3428.0049608

0 概述

链路预测是复杂网络研究中的重要问题之一。链路预测问题通常被描述为在已知网络拓扑结构与标签等信息的情况下，对 2 个未连边节点间产生连边的可能性进行预测，广泛应用于脑网络[1]、ad-hoc 网络[2]及社会网络[3]等方面。社会网络中的链路预测主要通过挖掘用户之间的各种信息来预测将会产生的关系，因此，研究链路预测有利于进行网络结构演化的预测与各种社会网络分析。

相似性指标是链路预测中的一种常用方法。文献[4]结合局部聚类系数与同配系数来发掘社会网络中节点间更多的信息，相比经典指标提高了预测准确度。文献[5]考虑局部社团特征，提出一种新的相似性指标来提高预测准确度。文献[6]尝试使用相似性指标、图核与话题模型的方法进行异质网络中的连接预测。文献[7]综合考虑节点进化多样性，以节点角度计算综合相似度值。上述算法均考虑了多种相似性属性并具有较好的准确度，但仅适用于静态无向网络，未考虑时间维度与连接方向等属性。

现实中的社会网络通常是处于一种不断变化的状态，网络的拓扑结构会随时间变化，并且网络中人与人的连接一般是有向性的，时序有向网络可以用来合理地描述社会网络。目前，关于时序网络与有向网络的链路预测方法的研究较多，其中，时序网络采用时间衰减因子[8]、时序潜在空间[9]、马尔卡夫概率模型[10]与二元特征提取[11]等方法进行预测。针对有向网络，国内外学者提出了基于节点不同度值并分配权值排序的方法[12]、基于路径采样结合 Katz 指标的方法[13]、基于粗糙集构建的粗糙相似度方法[14]和基于三角模体和权重时间因子的方法[15]。上述链路预测算法分别在时序网络与有向网络上均具有较好的预

基金项目： 国家自然科学基金（51334003，61473039）；北京市属高等学校高层次人才引进与培养计划（CIT&TCD201504058）。

作者简介： 杨瑞琪（1993—），男，硕士研究生，主研方向为复杂网络；张月霞，副教授、博士。

收稿日期： 2017-12-07 **修回日期：** 2018-01-22 **E-mail：** yangruiqi_bistu@126.com

测准确度,但真实的社会网络通常是同时具有多个特征的网络,因此同时考虑时序与有向属性的算法较少,且使用经典相似性算法的准确度较低。

针对上述问题,本文提出一种基于归一化共同邻居与局部聚类(Normalized Common Neighbors and Local Clustering,NCNLC)相似性指标的链路预测算法。该算法结合全局的归一化共同邻居属性与局部聚类属性构建一种新的相似性指标,并引入时间衰减因子,以在时序网络中进行链路预测。

1 网络描述

在一个网络 $G=(V,E)$ 中,V 为节点集合,E 为节点间的边集合。对于带有时间戳的网络数据,假设 T 为网络数据中最大与最小时间戳的差值,将$[0,T]$划分为多个时间间隔 $T_1,T_2,\cdots,T_i$,称为时间步[10],每个时间步长度相等,时间步长度的选择通常根据需要或场景来确定,时间步 T_i 所对应的网络图为 $G_{T_i}=(V_{T_i},E_{T_i})$。

假设在时序网络中,时间步 T_i 所对应的网络会受到前 $i-1$ 个网络的影响,给节点间的连边分配时间衰减因子 $\alpha(\alpha\leqslant 1)$[8]。定义时间步 T_2 所对应的网络中节点 i 和节点 j 的累积影响因子如式(1)所示。

$$I_{ij}^{T_2}=\sum_{t=T_1}^{T_2}E_{ij}(t)\times\alpha^{T_2-t} \tag{1}$$

其中,T_1 为初始时间步,$E_{ij}(t)$为时间步 t 时节点 i 和节点 j 间的连边,若存在连边则为 1,否则为 0。将 2 个节点间的累计影响因子作为节点间的属性,参与相似度值的计算。

2 节点相似性指标

本文通过计算共同邻居与节点度等属性来计算 2 个节点间的相似度,根据相似度值判断是否会产生连边。

2.1 常用相似性指标

在利用网络局部信息的相似性指标中,CN、Jaccard、Slaton 与 Sorenson 指标是根据共同邻居对节点度不同的影响来计算相似度值,AA 指标是根据节点共同邻居度的影响,其思想是度小的共同邻居贡献度大于度大的共同邻居,RA 指标是根据节点间的资源分配量来定义相似度。LAS(Local Affinity Structure)[5]指标是一种综合考虑共同邻居与局部社团相似性的指标,通过计算节点间共同邻居个数与节点度商的总和来描述一组节点对与它们共同邻居间的关联相似性,该指标在常用无向网络数据集上的链路预测性能优于 Salton、Jaccard 与 Sorenson 等相似性指标。本文符号定义如表 1 所示。

表 1 符号定义

符号	描述
$\Gamma(x)$	节点 x 的邻居节点集合
$\Gamma_{\text{in}}(x)$	节点 x 的入度邻居节点集合
$\Gamma_{\text{out}}(x)$	节点 x 的出度邻居节点集合
$k(x)$	节点 x 的度
$k_{\text{in}}(x)$	节点 x 的入度
$k_{\text{out}}(x)$	节点 x 的出度

相似性指标的计算公式通常用于无向网络,若应用于时序有向网络需要进行修改,3 种网络中的常用相似性指标计算公式如表 2 所示,其中 T 表示时间。

表 2 不同网络中的相似性指标计算公式

相似性指标	无向网络	有向网络	时序有向网络
CN	$S_{xy}=\lvert\Gamma(x)\cap\Gamma(y)\rvert$	$S_{xy}=\lvert\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y)\rvert$	$S_{xy}^T=\sum_{e\in\Gamma_{\text{out}}^T(x)\cap\Gamma_{\text{in}}^T(y)}I_{ex}^T\times I_{ey}^T$
Jaccard	$S_{xy}=\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\lvert\Gamma(x)\cup\Gamma(y)\rvert}$	$S_{xy}=\frac{\lvert\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y)\rvert}{\lvert\Gamma_{\text{out}}(x)\cup\Gamma_{\text{in}}(y)\rvert}$	$S_{xy}^T=\frac{\sum_{e\in\Gamma_{\text{out}}^T(x)\cap\Gamma_{\text{in}}^T(y)}I_{ex}^T\times I_{ey}^T}{\sum_{e\in\Gamma_{\text{out}}^T(x)\cup\Gamma_{\text{in}}^T(y)}I_{ex}^T\times I_{ey}^T}$
Slaton	$S_{xy}=\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\sqrt{k(x)\times k(y)}}$	$S_{xy}=\frac{\lvert\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y)\rvert}{\sqrt{k_{\text{out}}(x)\times k_{\text{in}}(y)}}$	$S_{xy}^T=\frac{\sum_{e\in\Gamma_{\text{out}}^T(x)\cap\Gamma_{\text{in}}^T(y)}I_{ex}^T\times I_{ey}^T}{\sqrt{\sum_{e\in\Gamma_{\text{out}}^T(x)}I_{ex}^T\times\sum_{e\in\Gamma_{\text{in}}^T(y)}I_{ey}^T}}$
Sorensen	$S_{xy}=\frac{2\times\lvert\Gamma(x)\cap\Gamma(y)\rvert}{k(x)+k(y)}$	$S_{xy}=\frac{2\times\lvert\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y)\rvert}{k_{\text{out}}(x)+k_{\text{in}}(y)}$	$S_{xy}^T=\frac{2\times\sum_{e\in\Gamma_{\text{out}}^T(x)\cap\Gamma_{\text{in}}^T(y)}I_{ex}^T\times I_{ey}^T}{\sum_{e\in\Gamma_{\text{out}}^T(x)}I_{ex}^T+\sum_{e\in\Gamma_{\text{in}}^T(y)}I_{ey}^T}$
AA	$S_{xy}=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{\text{lb}\,k(z)}$	$S_{xy}=\sum_{z\in\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y)}\frac{1}{\text{lb}k_{\text{out}}(z)}$	$S_{xy}^T=\sum_{z\in\Gamma_{\text{out}}^T(x)\cap\Gamma_{\text{in}}^T(y)}\frac{1}{\text{lb}\sum_{e\in\Gamma_{\text{out}}^T(z)}I_{ez}^T}$
RA	$S_{xy}=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{k(z)}$	$S_{xy}=\sum_{z\in\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y)}\frac{1}{k_{\text{out}}(z)}$	$S_{xy}^T=\sum_{z\in\Gamma_{\text{out}}^T(x)\cap\Gamma_{\text{in}}^T(y)}\frac{1}{\sum_{e\in\Gamma_{\text{out}}^T(z)}I_{ez}^T}$
LAS	$S_{xy}=\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{k(x)}+\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{k(y)}$	$S_{xy}=\frac{\lvert\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y)\rvert}{k_{\text{out}}(x)}+\frac{\lvert\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y)\rvert}{k_{\text{in}}(y)}$	$S_{xy}^T=\frac{\sum_{e\in\Gamma_{\text{out}}^T(x)\cap\Gamma_{\text{in}}^T(y)}I_{ex}^T\times I_{ey}^T}{\sum_{e\in\Gamma_{\text{out}}^T(x)}I_{ex}^T}+\frac{\sum_{e\in\Gamma_{\text{out}}^T(x)\cap\Gamma_{\text{in}}^T(y)}I_{ex}^T\times I_{ey}^T}{\sum_{e\in\Gamma_{\text{in}}^T(y)}I_{ey}^T}$

2.2　NCNLC 相似性指标

常用相似性指标一般是利用网络的局部或全局拓扑信息来计算节点相似性,本文算法综合利用网络的全局与局部拓扑信息,提出一种结合节点全局归一化共同邻居属性与局部聚类系数的 NCNLC 相似性指标。

时间步 T 对应的网络中节点 x 和 y 归一化共同邻居系数 NCN_{xy}^{T} 定义如下:

$$NCN_{xy}^{T}=\frac{|CNI_{xy}^{T}|}{\max\limits_{i\in V_T,j\in V_T,i\neq j}|CNI_{ij}^{T}|} \tag{2}$$

其中,CNI_{xy}^{T} 为时间步 T 的网络中节点 x 和节点 y 之间累积影响因子非零的共同邻居节点,$|\cdot|$ 表示集合中的元素个数,显然 NCN_{xy}^{T} 的取值范围为[0,1]。

时间步 T 对应的网络中节点 x 和 y 的局部聚类系数 LCC_{xy}^{T} 定义如下:

$$LCC_{xy}^{T}=\frac{|E_{ij}|}{m_{ij}},i,j\in\Gamma_{\text{out}}(x)\cap\Gamma_{\text{in}}(y),i\neq j \tag{3}$$

其中,E_{ij} 表示节点 i 和节点 j 之间累积影响因子非零的连接,m_{ij} 表示节点 i 和节点 j 之间的最大有向边数。

结合式(2)与式(3)的定义构建一种共同邻居特征与局部聚类特征的 NCNLC 相似性指标,计算公式为:

$$S_{xy}^{T}=\theta\times NCN_{xy}^{T}+(1-\theta)\times LCC_{xy}^{T} \tag{4}$$

其中,θ 为分配参数,取值范围为[0,1],用来调整相似性指标中 NCN 与 LCC 所占的权重比例。该指标综合了节点的归一化共同邻居系数和节点邻居间的连接数量,不仅考虑共同邻居对节点的影响,而且考虑周围邻居间连接强度[16]的影响。该指标与 LAS 指标相比,LAS 仅利用节点的度与共同邻居属性,没有考虑共同邻居之间的联系,而 NCNLC 中不仅包含共同邻居属性,还考虑共同邻居之间的聚类属性。

3　仿真结果与分析

3.1　数据集

为了验证算法的有效性,本文在 email-eu-core 数据集[17]上进行仿真验证。该数据集是一个大型欧洲研究机构中成员间收发邮件的匿名数据集,数据表示机构成员之间的交流情况,包含发送用户 ID、接收用户 ID 与时间戳信息,数据集的网络统计属性如表 3 所示。

表 3　email-eu-core 数据集网络统计属性

数据集名称	节点数	时序边数	静态网络边数	时长/d
email-eu	986	332 334	24 929	803
email-eu-d1	309	61 046	3 031	803
email-eu-d2	162	46 772	1 772	803
email-eu-d3	89	12 216	1 506	802
email-eu-d4	142	481 441	1 375	803

在实验中,考虑到邮件交流的周期较长,将 email-eu-core 数据集的时间步 T 取值为 30 d,提取数据集的拓扑结构信息与时间特征信息来构建时序有向网络,将选定时间步之前的网络邻接矩阵作为训练集,将当前时间步的网络邻接矩阵作为测试集。

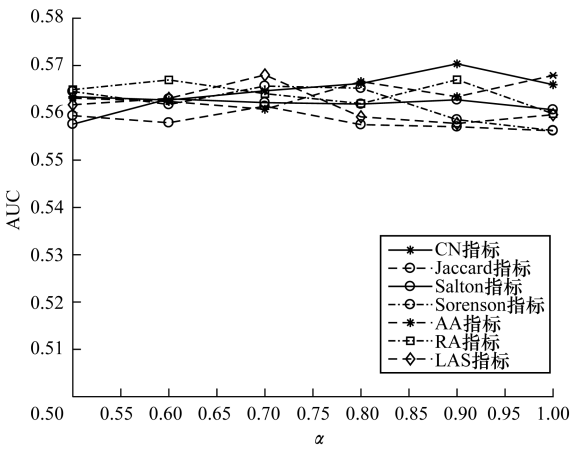
本文采用 AUC(Area Under the receiver operating characteristic Curve)指标从整体上衡量链路预测算法的准确度,其可以描述为:在测试集中,随机抽取的已存在连边的相似度值比随机抽取的不存在连边的相似度值高的概率,定义为:

$$AUC=\frac{n'+0.5n''}{n} \tag{5}$$

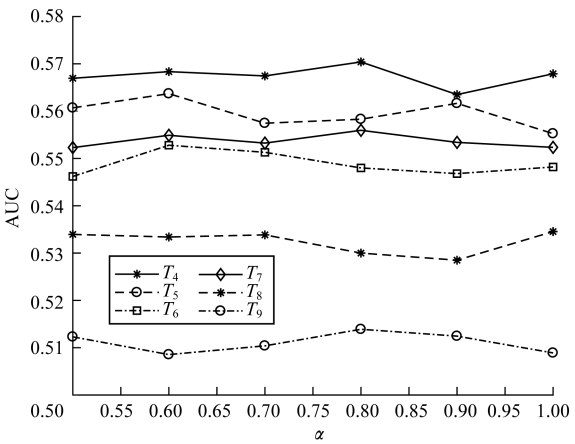
其中,n 为独立重复抽取的次数,n' 为测试集中已存在连边的相似度值大于不存在连边的相似度值的次数,n'' 为测试集中已存在连边的相似度值等于不存在连边的相似度值的次数。

3.2　时间衰减因子 α 分析

为了分析时间衰减因子 α 对预测结果的影响,选取不同的 α 值进行仿真。图 1 所示为不同 α 取值的仿真分析。图 1(a)为常用相似性指标在不同 α 取值的时间步 T_4 网络上的对比结果,图 1(b)为 NCNLC 指标在不同 α 取值的时间步 T_4 ~ T_9 网络上的对比结果。可以看出,各时间步在不同 α 取值网络上的 AUC 值均较为稳定,说明 α 的取值对时序网络中预测结果的影响较小。



(a)各指标的AUC值对比结果



(b)NCNLC指标在不同α取值的时间网络上的对比结果

图 1　时间衰减因子 α 与 AUC 的关系

3.3 分配参数 θ 分析

为了分析分配参数 θ 对 NCNLC 指标的影响，选取不同的 θ 值进行仿真。图 2 所示是使用不同分配参数 θ 的 NCNLC 指标在不同时间步的 email-eu 网络上进行的对比仿真。

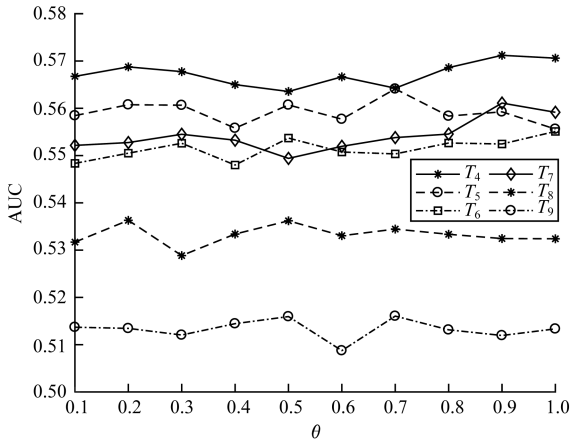


图 2 AUC 与分配参数 θ 的关系

从图 2 可以看出，在时间步 T_4、T_5、T_8 和 T_9 网络上 AUC 的变化不大，说明在这些网络上 NCNLC 指标中的NCN和LCC的权重相当，而在时间步 T_6 和 T_7 网络上 AUC 有略微递增的趋势，说明 NCN 对结果的影响略大于 LCC。NCNLC 指标中 NCN 和 LCC 的权重相当，对算法的预测准确度影响较小。

3.4 性能分析

在 email-eu-core 数据集上进行的本文算法与常用相似性算法的性能对比结果如图 3 所示。从图 3(a)可以看出，在最大的 email-eu 网络数据集中，从预测精度上来看，NCNLC 指标的预测效果明显优于 LAS 指标与 CN、Jaccard 等常用指标。从图 3(b)～图 3(e)可以看出，在大部分中小型网络数据集中，NCNLC 指标也具有明显的优势：在 email-eu-d1 网络数据集中 NCNLC 指标效果略优于其他算法；在 email-eu-d2 数据集网络中算法的整体效果不稳定，没有在全局时间步上效果均较好的算法，这是由于 email-eu-d2 训练集中的网络经过不同时间步网络的累积影响因子叠加后网络的平均聚类系数较大，较难区分出性能最优的算法；在 email-eu-d3 网络数据集中 NCNLC 指标的准确度明显优于其他指标，其次是 LAS 指标；在 email-eu-d4 网络数据集中在多数时间步上 NCNLC 指标的效果更优。仿真结果表明基于 NCNLC 指标的算法在时序有向网络的链路预测中具有一定的可行性和有效性。

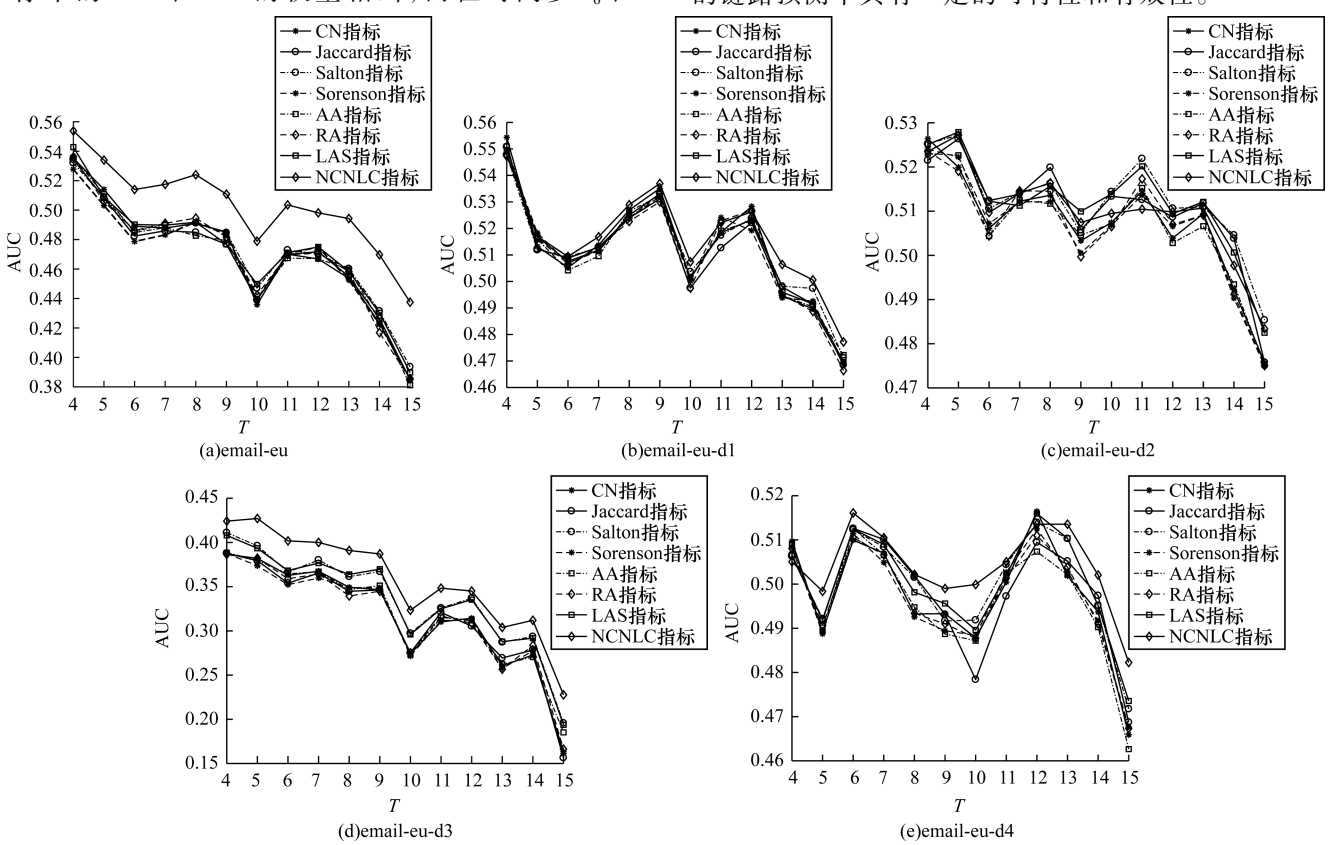


图 3 各指标在不同数据集上的 AUC 对比结果

4 结束语

本文提出一种基于 NCNLC 相似性指标的链路预测算法。该算法综合考虑了共同邻居与局部聚类特征，在真实社会网络数据集上进行仿真对比，结果表明，该算法的预测效果优于常用相似性算法，并且使用本文算法可以有效地提高时序有向社会网络上的链路预测准确度。

参考文献

[1] YAN H, YANG F, YU Y. Link prediction in brain networks with epilepsy disease [C]//Proceedings of International IEEE/EMBS Conference on Neural Engineering. Washington D. C., USA: IEEE Press, 2017:138-141.

[2] GITE P. Link stability prediction for mobile Ad-hoc network route stability[C]//Proceedings of International Conference on Inventive Systems and Control. Washington D. C., USA: IEEE Press, 2017.

[3] WANG T, LIAO G. A review of link prediction in social networks[C]//Proceedings of International Conference on Management of e-Commerce and e-Government. Washington D. C., USA: IEEE Computer Society, 2014: 147-150.

[4] 杨旭华,凌非. 一种基于局部社团和全局信息的链路预测算法[J]. 浙江工业大学学报,2017,45(1):10-13,63.

[5] SUN Q, HU R, YANG Z, et al. An improved link prediction algorithm based on degrees and similarities of nodes [C]//Proceedings of the 16th International Conference on Computer and Information Science. Washington D. C., USA: IEEE Press, 2017:13-18.

[6] ANIL A, KUMAR D, SHARMA S, et al. Link prediction using social network analysis over heterogeneous terrorist network [C]//Proceedings of IEEE International Conference on Smart City/Socialcom/Sustaincom. Washington D. C., USA: IEEE Press, 2015:267-272.

[7] WANG H, HU W, QIU Z, et al. Nodes' evolution diversity and link prediction in social networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(10):2263-2274.

[8] 陈莎,朱福喜,阳小兰,等. 一种基于混合相似性指标的网络动态链路预测方法[J]. 小型微型计算机系统,2016,37(8):1798-1801.

[9] ZHU L, GUO D, YIN J, et al. Scalable temporal latent space inference for link prediction in dynamic social networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(10):2765-2777.

[10] DAS S, DAS S K. A probabilistic link prediction model in time-varying social networks [C]//Proceedings of IEEE International Conference on Communications. Washington D. C., USA: IEEE Press, 2017.

[11] SETT N, BASU S, NANDI S, et al. Temporal link prediction in multi-relational network [J]. World Wide Web Internet and Web Information Systems, 2018, 21(2):395-419.

[12] YU Y, WANG X. Link prediction in directed network and its application in microblog [EB/OL]. [2017-11-20]. https://www.hindawi.com/journals/mpe/2014/509282/.

[13] LIU Q H, CHEN L. Sampling based one-source link prediction in directed network [C]//Proceedings of International Conference on Robotics and Automation Sciences. Washington D. C., USA: IEEE Press, 2017: 139-143.

[14] 张军,夏昊翔. 基于粗糙相似度的有向网络链路预测[J]. 系统工程,2015,33(11):146-152.

[15] BÜTÜN E, KAYA M, ALHAJJ R. A new topological metric for link prediction in directed, weighted and temporal networks [C]//Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Washington D. C., USA: IEEE Press, 2016:954-959.

[16] ZHANG X, WANG C, SU Y, et al. A fast overlapping community detection algorithm based on weak cliques for large-scale networks [J]. IEEE Transactions on Computational Social Systems, 2017, 4(4):218-230.

[17] PARANJAPE A, BENSON A R, LESKOVEC J. Motifs in temporal networks [C]//Proceedings of the 10th ACM International Conference on Web Search and Data Mining. New York, USA: ACM Press, 2017:601-610.

编辑　赵　辉

(上接第 196 页)

[4] BHUNIA A K, BHOWMICK A, BHUNIA A K, et al. Handwriting trajectory recovery using end-to-end deep encoder-decoder network [EB/OL]. [2017-12-05]. https://arxiv.org/ftp/arxiv/papers/1801/1801.07211.pdf.

[5] NOUBIGH Z, KHERALLAH M. A survey on handwriting recognition based on the trajectory recovery technique[C]//Proceedings of International Workshop on Arabic Script Analysis and Recognition. Washington D. C., USA: IEEE Press, 2017:69-73.

[6] LI X, SHU W, LI M, et al. Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring [J]. IEEE Transactions on Vehicular Technology, 2009, 58(4):1647-1653.

[7] TONG C, CHEN H, QI X, et al. A framework for bus trajectory extraction and missing data recovery for data sampled from the internet[J]. Sensors, 2017, 17(2):342.

[8] WU H, XUE M, CAO J, et al. Fuzzy trajectory linking[C]//Proceedings of the 32nd International Conference on Data Engineering. Washington D. C., USA: IEEE Press, 2016:859-870.

[9] 梁天新,杨小平,王良,等. 记忆神经网络的研究与发展[J]. 软件学报,2017,28(11):2905-2924.

[10] 韩冬煦,常宝宝. 中文分词模型的领域适应性方法[J]. 计算机学报,2015,38(2):272-281.

[11] KACHERGIS G, YU C, SHIFFRIN R M. A bootstrapping model of frequency and context effects in word learning [J]. Cognitive Science, 2016, 41(3): 590-622.

[12] ZHAO Y, YANG R, CHEVALIER G, et al. Applying deep bidirectional LSTM and mixture density network for basketball trajectory prediction[J]. Optik, 2018, 158: 266-272.

[13] ZHOU Z H. Ensemble methods: foundations and algorithms[M]. [S.l.]: CRC Press, 2012:68-70.

[14] 付忠良. 通用集成学习算法的构造[J]. 计算机研究与发展,2013,50(4):861-872.

[15] 李凯,崔丽娟. 集成学习算法的差异性及性能比较[J]. 计算机工程,2008,34(6):35-37.

[16] T-Drive trajectory data sample [EB/OL]. [2017-12-05]. http://research.microsoft.com/apps/pubs/? id=152883.

编辑　赵　辉