

doi:10.3969/j.issn.1000-2162.2024.02.006

基于离散时间量子漫步的链路预测算法

侍伟敏, 梁佳伟, 周艺华, 杨宇光

(北京工业大学 北京可信计算重点实验室, 北京 100124)

摘要:量子漫步算法能模拟游走粒子在图上的量子相干演化,粒子的运动状态由量子态的相干叠加而成.与经典随机游走算法相比,量子漫步算法具有寻找目标节点时间少和源节点扩散至其他节点时间少的优点.提出一种基于离散时间量子漫步的链路预测(link prediction based on discrete time quantum walk,简称LP-DTQW)算法.研究表明:相对于其他7种算法,LP-DTQW算法有更高的预测精度;LP-DTQW算法的时间复杂度远低于经典RWR(random walk with restart)链路预测算法的时间复杂度.因此,LP-DTQW算法具有更强的预测性能.

关键词:复杂网络;链路预测;离散时间量子漫步;拓扑相似性

中图分类号:TP309

文献标志码:A

文章编号:1000-2162(2024)02-0034-06

Link prediction algorithm based on discrete time quantum walk

SHI Weimin, LIANG Jiawei, ZHOU Yihua, YANG Yuguang

(Beijing Key Laboratory of Trusted Computing, Beijing University of Technology, Beijing 100124, China)

Abstract: The quantum walk algorithm can simulate the quantum coherent evolution of a walking particle on a graph. The motion state of the walking particle is a coherent superposition of quantum states. Compared with classical random walk algorithms, quantum walk algorithms have the advantages of less time for finding the target node and less time for diffusion from the source node to other nodes. A link prediction based on discrete-time quantum walk (LP-DTQW) algorithm was proposed. The experimental results indicated that the LP-DTQW algorithm exhibited higher prediction accuracy compared with the other seven algorithms. Additionally, the time complexity of the LP-DTQW algorithm was significantly lower than that of the classical RWR (random walk with restart) link prediction algorithm. Therefore, the LP-DTQW algorithm exhibited stronger predictive performance.

Keywords: complex network; link prediction; discrete time quantum walk; topology similarity

链路预测是复杂网络研究的重要分支.在众多预测算法中,基于经典随机游走的链路预测算法,因能充分利用网络拓扑结构且有较高的预测精度而受到关注.文献[1]改变了游走者随机游走过程中等概率无偏向采样邻居节点的策略,提出了MHRW(metropolis-hasting random walk)算法,该算法得到的样本能更全面反映原始网络的拓扑结构.文献[2]提出了一种基于OWA(ordered weighted averaging)

收稿日期:2023-03-29

基金项目:国家自然科学基金资助项目(62071015)

作者简介:侍伟敏(1978—),女,北京人,北京工业大学副教授,硕士生导师,博士,E-mail: shiweimin@bjut.edu.cn.

算子的链路预测算法,该算法解决了局部信息算法预测稳定性差的问题.针对目前预测算法要么侧重于节点属性、要么侧重于结构属性的问题,文献[3]提出了一种结合节点属性和结构属性的链路预测算法,该算法与其他算法相比具有更高的预测精度.文献[4]提出了一种基于随机游走和深度信念的机会网络链路预测算法,能提取机会网络动态演化过程中的时域特征,具有较高的准确性和稳定性.文献[5]提出了一种简单、高效、可扩展的基于网络社区的链路预测算法,该算法在不牺牲预测精度的前提下有较高的预测效率.

机器学习最近几年得到迅速发展,越来越多的研究人员将机器学习应用于链路预测算法,并取得了一系列成果.文献[6]为了解决网络结构多样性导致的单一链路预测指标无法应用于各类网络问题,提出了一种适应各类网络结构的链路预测算法,该算法将各种传统的预测指标作为未知链接的多维数据,通过聚类分析将链路预测问题转化为分类问题.文献[7]提出了一种基于节点中心性和机器学习分类器的通用链路预测算法,使用 LightGBM(light gradient boosted machine)进行预测.文献[8]通过基于路径的相似性度量,构建监督机器学习的特征向量,预测网络产生的链接.文献[9]基于网络拓扑特征和链接生成次数,构建动态网络的特征时间序列,结合统计模型和监督学习,提出了一种加权动态网络的链路预测算法,与传统的静态链路预测算法和未加权的动态链路预测算法相比,该算法能一定程度上提高预测精度.文献[10]提出了一种基于对抗训练框架的预测算法,改进了不同领域的高度偏斜和不平衡图的链路预测.文献[11]提出了一种基于时间注意力机制和对抗网络的新型深度学习模型,用于动态加权网络的链路预测,该模型包括判断器和生成器,判断器能判断网络是真实的还是由生成器伪造的,生成器通过图的拓扑结构特征及时间自注意力机制学习网络的演化.

量子漫步已有多种应用,如空间搜索^[12]、图同构^[13]、量子路由^[14]等.相对于经典随机游走,量子漫步具有更快的多项式分解及粒子扩散^[15-16]优点.鉴于此,该文拟提出一种基于离散时间量子漫步的链路预测(link prediction based on discrete time quantum walk,简称 LP-DTQW)算法.

1 基于离散时间量子漫步的链路预测算法

1.1 离散时间量子漫步模型

网络 $G(V, E)$ 的节点集合 $V = \{v_i, i \in \{1, 2, \dots, n\}\}$ 构成整个离散时间量子漫步的位置状态空间.构造基于离散时间量子漫步模型需引入游走粒子和硬币,游走粒子的希尔伯特空间 H_p 由粒子位置态 $\{|i\rangle, i \in \{1, 2, \dots, n\}\}$ 张起,且这些位置态相互正交.硬币的希尔伯特空间 H_c 由硬币态 $\{|k\rangle, k \in \{1, 2, \dots, n\}\}$ 张起.整个量子系统的希尔伯特空间定义为

$$H = H_p \otimes H_c. \quad (1)$$

量子漫步的时间演化算符定义为

$$U = SC, \quad (2)$$

其中: S 为条件移位算符, C 为硬币算符.条件移位算符根据游走粒子的方向将粒子迁移至下一节点,硬币算符将游走粒子的方向变为各方向的相干叠加.

条件移位算符 S 有 2 种表现形式,一种为转移算符,它使游走粒子跳跃且始终指向同一方向,另一种为触发器移位算符,它使游走粒子跳跃和翻转,且有 $S(|j\rangle \otimes |k\rangle) = |k\rangle \otimes |j\rangle$.该文使用触发器移位算符作为条件移位算符.

图 1 为触发器移位算符作用示意图.由图 1 可知,触发器移位算符使游走粒子改变了游走方向.触发器移位算符的定义式为

$$S = \sum_{j,k=1}^n |j, k\rangle \langle k, j|. \quad (3)$$

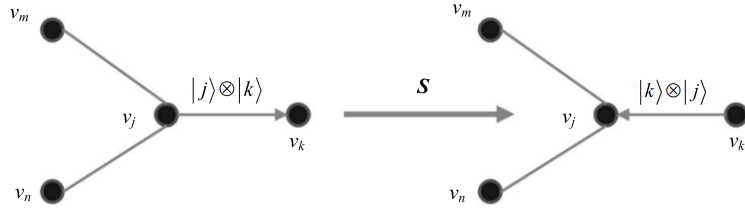


图 1 触发器移位算符作用示意图

图 2 为硬币算符作用示意图. 由图 2 可知, 硬币算符使游走粒子游走时有更多选择. 硬币算符定义式为

$$C = \sum_{j=1}^n |j\rangle\langle j| \otimes C_j, \quad (4)$$

其中: $C_j = 2|s_j\rangle\langle s_j| - I_{\deg(v_j) \times \deg(v_j)}$, $\deg(v_j)$ 表示节点 v_j 的度, $I_{\deg(v_j) \times \deg(v_j)}$ 为 $\deg(v_j)$ 维 \times $\deg(v_j)$ 维的单位矩阵, $|s_j\rangle = \frac{1}{\sqrt{\deg(v_j)}} \sum_{k=1}^n |k\rangle$.

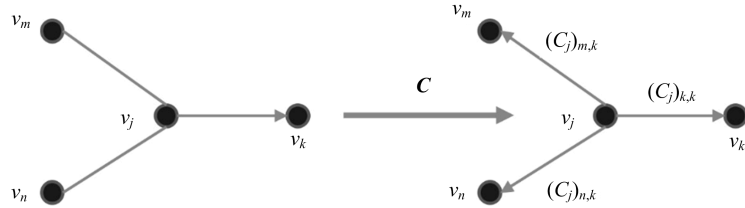


图 2 硬币算符作用示意图

为了实现基于离散时间量子漫步的链路预测算法, 该文引入 Szegedy 量子漫步^[17]重新定义上述公式中的 $|s_j\rangle$, 该量子漫步是一种在二部图上由反射算符驱动的离散时间量子漫步, 二部图是通过复制一个与离散马尔科夫链相关的简单图而得到的. 设 x 和 y 分别是二部图节点集合 X 和 Y 中的节点, 则演化算符 A, B 分别为

$$A = \sum_{x \in X} |\alpha_x\rangle\langle x|, \quad (5)$$

$$B = \sum_{y \in Y} |\beta_y\rangle\langle y|, \quad (6)$$

其中: $|\alpha_x\rangle = |x\rangle \otimes \left(\sum_{y \in Y} \sqrt{P_{xy}} |y\rangle \right)$, $|\beta_y\rangle = \left(\sum_{x \in X} \sqrt{P_{yx}} |x\rangle \right) \otimes |y\rangle$.

反射算符 R_A, R_B 分别为

$$R_A = 2AA^T - I_{n \times n}, \quad (7)$$

$$R_B = 2BB^T - I_{n \times n}, \quad (8)$$

其中: $I_{n \times n}$ 为 n 维 \times n 维的单位矩阵.

演化算符 W_p 为

$$W_p = R_B R_A. \quad (9)$$

硬币算符中的 $|s_j\rangle$ 为

$$|s_j\rangle = \sum_{e_{jk} \in E} \sqrt{P_{jk}} |k\rangle, \quad (10)$$

其中: $e_{jk} \in E$ 表示边 e_{jk} 属于边集 E , P_{jk} 表示游走粒子从节点 v_k 转移到 v_j 的概率.

设 $|\varphi(0)\rangle$ 为初始态, 经过时间 t 的量子漫步后量子态演化为

$$|\varphi(t)\rangle = U_t |\varphi(0)\rangle. \quad (11)$$

经过时间 t 后, 游走粒子到达节点 v_i 的瞬时概率为

$$p(i, t) = \sum_{j \in \{1, 2, \dots, n\}} |\langle i, j | U_t | \varphi(0) \rangle|^2 = \langle \varphi(0) | U_t^\dagger (|i\rangle \otimes I_{n \times n}) (\langle i | \otimes I_{n \times n}) U_t | \varphi(0) \rangle = \langle \varphi(0) | U_t^\dagger (|i\rangle\langle i| \otimes I_{n \times n}) U_t | \varphi(0) \rangle, \quad (12)$$

其中: $|i\rangle$ 为 n 维希尔伯特空间的基态, $|i\rangle \otimes \mathbf{I}_{n \times n}$ 表示 $|i\rangle$ 和 $\mathbf{I}_{n \times n}$ 的张量积.

为了实现一个稳态概率,该文定义了游走粒子到达节点 v_i 的平均概率,其表达式为

$$\langle p(i) \rangle = \frac{1}{T} \sum_{t=0}^{T-1} p(i, t). \quad (13)$$

当 T 足够大时,到达各节点的平均概率可收敛至稳态概率.

1.2 基于离散时间量子漫步的链路预测算法

该文提出 LP-DTQW 算法. 根据游走者从当前节点出发到达各节点的平均概率,定义两节点的相似值,计算 AUC(area under the curve). 采用 AUC 平均值度量算法的预测精度.

定义转移概率矩阵 \mathbf{P} 的矩阵元 P_{ij} 为

$$P_{ij} = \frac{\text{CN}(v_i, v_j)}{\min\{\deg(v_i) - 1, \deg(v_j) - 1\}}, \quad (14)$$

其中: P_{ij} 为从节点 v_i 到 v_j 的转移概率, $\text{CN}(v_i, v_j)$ 为节点 v_i 和 v_j 的共同邻居数. 对上述矩阵进行归一化处理,使矩阵的每行元素之和为 1,得到不等概率转移矩阵. 归一化后的转移概率为

$$p_{ij} = \frac{P_{ij}}{\sum_{j=1}^n P_{ij}}. \quad (15)$$

在转移概率矩阵 \mathbf{P} 的基础上,定义触发器移位算符 \mathbf{S} , 设置足够大的游走步数 T , 通过硬币算符 \mathbf{C} 和触发器移位算符 \mathbf{S} 控制游走粒子进行离散时间量子漫步.

定义节点 v_i 和 v_j 的相似值,其表达式为

$$s_{ij} = \frac{k_i}{M} p_{ij} + \frac{k_j}{M} p_{ji}, \quad (16)$$

其中: M 为整个网络的边数. 显然, $s_{ij} = s_{ji}$.

LP-DTQW 算法的流程如下.

- (1) 初始化转移概率矩阵和相似值矩阵;
- (2) 根据式(2)~(4)分别定义演化算符、交换算符、硬币算符;
- (3) 通过式(14)~(15)计算转移概率矩阵,设置游走步数;
- (4) 在图上进行离散时间量子漫步;
- (5) 通过式(12)~(13)计算两节点间的平均概率;
- (6) 通过式(16)计算两节点间的相似值;
- (7) 返回相似值矩阵.

2 仿真分析

在 4 个网络上进行实验,每个网络上分别独立进行 60 次实验. 实验在具有 intel i5-12490F CPU 3.00 GHz, 64 GB RAM 和 64 位操作系统的电脑上进行,使用 MatlabR2021b 进行仿真分析. 所用网络情况介绍如下.

(1) Metabolic 网络. 该网络为线虫的新陈代谢网络,网络节点表示生物基底,连边表示生物基底之间的生化反应. 网络有 453 个节点和 2 025 条连边.

(2) Polbooks 网络. 该网络为 2004 年美国亚马逊在线售卖的有关政治书的关系网络,网络节点表示政治书,连边表示同一客户同时购买两个不同节点对应的两本书. 网络有 105 个节点和 441 条连边.

(3) Dolphins 网络. 该网络为新西兰的海豚关系网络,网络节点表示海豚,若两只海豚经常一起频繁活动,则相应两节点间就存在一条边. 网络有 62 个节点和 159 条连边.

(4) USAir 网络. 该网络为美国航空网络,网络节点对应机场,连边表示航线. 网络有 332 个节点和 2 126 条连边.

表 1 为 4 个网络的拓扑结构特征.

表 1 4 个网络的拓扑结构特征

网络	节点数	连边数	平均度	平均聚类系数	平均最短路径长度	度异质性
Metabolic	453	2 025	8.94	0.647	2.66	4.49
Polbooks	105	441	8.40	0.488	3.08	1.42
Dolphins	62	159	5.13	0.303	3.36	1.66
USAir	332	2 126	12.81	0.749	2.74	3.46

为验证该文提出的 LP-DTQW 算法的优越性,将其与 AA(adamic-adar),RA(resource allocation),CN(common neighbors),ACT(average commute time),Cos+(cosine based on L^+),LRW(local random walk),SRW(superposed random walk)算法进行对比.表 2 为 8 种算法在 4 个网络上 60 次实验得到的预测精度.由表 2 可知,LP-DTQW 算法在 4 个网络上的预测精度均达到 0.830 1 以上,均高于其他 7 种算法,表明 LP-DTQW 算法有更高的预测精度.

表 2 8 种算法在 4 个网络上 60 次实验得到的预测精度

网络	AA	RA	CN	ACT	Cos+	LRW	SRW	LP-DTQW
Metabolic	0.952 5	0.957 2	0.919 3	0.763 7	0.898 0	0.953 7	0.968 1	0.969 7
Polbooks	0.895 0	0.896 6	0.886 1	0.761 5	0.895 9	0.925 0	0.920 0	0.927 2
Dolphins	0.790 9	0.789 3	0.787 8	0.770 4	0.820 6	0.808 2	0.828 0	0.830 2
USAir	0.966 7	0.972 5	0.954 8	0.903 4	0.957 6	0.962 3	0.973 3	0.977 3

3 时间复杂度分析

LP-DTQW 算法的时间复杂度是 $O(n^2)$,其中 n 表示网络节点数.经典 RWR 链路预测算法的时间复杂度为 $O(n^3)$.可见,该文提出的 LP-DTQW 算法的时间复杂度远远低于经典 RWR 链路预测算法的时间复杂度.

4 结束语

该文提出了 LP-DTQW.研究表明:相对于其他 7 种算法,LP-DTQW 算法有更高的预测精度;LP-DTQW 算法的时间复杂度远低于经典 RWR 链路预测算法的时间复杂度.因此,LP-DTQW 算法具有优越性.

参考文献:

[1] 吕亮,何敏,易灿.一种基于 MH 改进的重启随机游走链路预测算法[J].云南大学学报(自然科学版),2021,43(2):245-253.

[2] LI T T, ZENG C Y, FENG Y, et al. Research of local similarity index based on OWA integration operator in terrorist network link prediction method[C]//2021 33rd Chinese Control and Decision Conference (CCDC), 2021: 6420-6424.

[3] LI Z C, JI L X, LI J S, et al. A new link prediction in directed networks based on attributes fusion[C]//2020 IEEE International Conference on Smart Internet of Things, 2020: 161-167.

[4] LIAO Z L, LIU L L, CHEN Y B. A novel link prediction method for opportunistic networks based on random walk and a deep belief network[J]. IEEE Access, 2020, 8(2): 16236-16247.

[5] MOKHTARI S, SHAKIBIAN H. An efficient link prediction method using community structures[C]//2021 12th International Conference on Information and Knowledge Technology, 2021: 174-177.

[6] SHAO H, WANG L W, DENG J. A link prediction algorithm by unsupervised machine learning[C]//International Conference on Communications, Information System and Computer Engineering, 2019: 622-625.

- [7] KUMAR S, MALLIK A, PANDA B S. Link prediction in complex networks using node centrality and light gradient boosting machine[J]. World Wide Web, 2022, 25 (6): 2487-2513.
- [8] KUMARI A, SAHOO S P, BEHERA R K, et al. Supervised machine learning for link prediction using path-based similarity features[EB/OL]. [2022-01-23]. <https://ieeexplore.ieee.org/document/9342531/>.
- [9] LIU J C, JIANG Y N, WANG Y S, et al. Link prediction in dynamic networks based on machine learning [C]//2020 3rd International Conference on Unmanned Systems, 2020: 836-841.
- [10] ZHU H, BENDRE M, HASSOUN S, et al. Few-shot link prediction with domain-agnostic graph embedding [C]//2022 21st IEEE International Conference on Machine Learning and Applications, 2022: 659-664.
- [11] TANG S, XIAO X Q. Att-GAN: a deep learning model for dynamic network weighted link prediction[C]//2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer, 2021: 15-20.
- [12] MAGNIEZF, SANTHA M, SZEGEDY M. Quantum algorithms for the triangle problem[J]. SIAM Journal on Computing, 2008, 37 (2): 413-424.
- [13] CHEN Y B, WANG X L, LIU Z, et al. A new meta-baseline for few-shot learning[EB/OL]. [2022-01-26]. <https://arxiv.org/pdf/2003.04390v3.pdf>.
- [14] CHAKRABORTY S, NOVO L, DI GIORGIO S, et al. Optimal quantum spatial search on random temporal networks[J]. Physical Review Letters, 2017, 119 (22): 220503.
- [15] TRAVAGLIONE C, MILBURN G J. Implementing the quantum random walk[J]. Physical Review A, 2002, 65 (3): 032310.
- [16] MOLFETTA G, ARRIGHI P. A quantum walk with both a continuous-time limit and a continuous-spacetime limit[J]. Quantum Information Processing, 2020, 19 (2): 1-16.
- [17] SANTOS R A M. Szegedy's quantum walk with queries[J]. Quantum Information Processing, 2016, 15 (11): 4461-4475.

(责任编辑 郑小虎)