

基于链接重要性和数据场的链接预测算法

陈巧玉, 班志杰*

(内蒙古大学 计算机学院, 呼和浩特 010021)

(* 通信作者电子邮箱 banzhijie@imu.edu.cn)

摘要:针对现有基于节点相似性的链接预测方法忽略了网络拓扑本身链接强度的信息,带权的拓扑路径方法中权重较难确定等缺陷,提出一种基于链接重要性和数据场的链接预测算法。首先,将所有链接边赋予不同的链接权重;其次,考虑潜在链接节点间的相互影响,对部分没有链接的节点进行链接预估计;最后,利用数据场势函数计算两节点间的相似值。在典型的网络数据进行的实验结果表明,所提方法在分类指标和推荐指标中都有很好的表现:以 AUC 为评价指标时,比同复杂度的局部路径(LP)算法提高了 3 到 6 个百分点;以 DCG 为评价指标时比 LP 算法提高了 1.5 到 2.5 个 DCG 值。算法整体上提高了预测准确性,且由于参数确定简单,复杂度又不高,在实际中易于部署。

关键词:链接预测;数据场;链接重要性;节点相似性;复杂网络

中图分类号: TP391.4 **文献标志码:** A

Link prediction algorithm based on link importance and data field

CHEN Qiaoyu, BAN Zhijie*

(School of Computer Science, Inner Mongolia University, Huihot Nei Mongol 010021, China)

Abstract: The existing link prediction methods based on node similarity usually ignore the link strength of network topology and the weight value in the topological path method with weight is difficult to set. To solve these problems, a new prediction algorithm based on link importance and data field was proposed. Firstly, this method assigned different weight for each link according to the topology graph. Secondly, it took into account the interaction between potential link nodes and pre-estimated the link values for the partial nodes without links. Finally, it calculated the similarity between two nodes with data field potential function. The experimental results on some typical data sets of the real-world network show that, the proposed method has good performance with both classification index and recommended index. In comparison to the Local Path (LP) algorithm with the same complexity, the proposed algorithm raises Area Under Curve (AUC) by 3 to 6 percentages, and raises Discounted Cumulative Gain (DCG) by 1.5 to 2.5 points. On the whole, it improves the prediction accuracy. Because of its easy parameter determination and low time complexity, this new approach can be deployed simply.

Key words: link prediction; data field; link importance; node similarity; complex network

0 引言

自 Liben-Nowell 等^[1]首次提出社会网络链接预测的问题以来,链接预测已成为社会网络分析的热点。链接预测是通过已知的网络结构等信息预测和估计目标网络中尚未被观察到的或者在未来会出现的链接,可应用于科学研究、社会安全、行政商业决策、分子生物学的蛋白质关系预测、犯罪网络调查、各种推荐系统等^[2-4]。此外,通过对链接预测的技术和方法进行研究,可以加深对于复杂网络的总体演变规律的理解,进而推动其他复杂网络研究分支的发展。

目前主流的链接预测方法主要分为基于概率模型的算法^[5-9]、基于节点相似性的预测算法^[1,10-19]。利用概率模型进行链接预测的基本思想是建立含有一组可调参数的模型,然后使用一些优化策略寻找最优的参数值,使得所得到的模型能够更好地再现真实网络的结构和关系特征。网络中两个

没有链接的节点对建立链接的概率等于在这组最优参数下,它们之间建立链接的条件概率。概率模型的优势在于较高的预测精确度,同时使用了网络的结构信息和节点的属性信息;但计算的复杂度以及非普适性的参数使其应用范围受到限制。基于节点相似性算法主要基于网络的拓扑信息进行链接预测,该方法对于集聚系数较低的网络预测准确性较低,但计算简单,可扩展性良好,且可被用于实时预测任务中,因此,本文主要研究基于节点相似性的链接预测算法。很多学者对基于节点相似性的方法作了相关研究,如文献[1,10-19],典型的方法有共同邻居(Common Neighbor, CN)算法^[1]、局部路径(Local Path, LP)算法^[10]和 Katz(A new status index named after Katz L)^[13]等。以 CN 为基础的 10 余种算法都只把网络中有直接链接的表示为 1,没有直接链接的表示为 0,没有考虑链接的重要性信息,而现实的网络链接中显然是存在链接强度的。基于路径拓扑的相似性方法,如 LP 和 Katz 算法的

收稿日期: 2014-04-13; **修回日期:** 2014-05-13。 **基金项目:** 国家自然科学基金资助项目(61063018);内蒙古自治区高等学校科学研究项目(NJZZ12005);内蒙古自治区自然科学基金资助项目(2014MS0603)。

作者简介: 陈巧玉(1986-),女,河北唐山人,硕士研究生,主要研究方向:数据挖掘、复杂网络分析; 班志杰(1976-),女,内蒙古赤峰人,副教授,主要研究方向:数据挖掘、在线网络分析。

权值确定不直观,且 Katz 的权值必须满足一定的条件,计算时间复杂度较高,每次权值的改变都需要重新计算,实际使用效率不高。

针对以上问题,本文提出了基于链接重要性和数据场的链接算法——WCDF(Weighted Complete Data Field)。该算法对现有的链接作加权处理,同时考虑了潜在链接间的影响,对一些没有链接的边进行链接预估计,最后利用数据场势函数计算两节点间的相似值,在整体上提高了预测的准确率。

1 相关工作

目前,基于节点相似性的链接预测算法已有大量研究。两个节点之间相似性越大,则它们之间存在链接的可能性就越大,这是应用节点相似性进行链接预测的重要前提假设。Liben-Nowell 等^[1]提出共同邻居(CN)的方法。该方法认为如果两节点有更多的共同邻居,那么它们更倾向于建立链接,因此考虑了节点的共同邻居这一拓扑属性。Zhou 等^[10]提出一种新的基于节点相似性的资源分配(Resource Allocation, RA)算法,比其他基于节点相似方法^[1,12]有更好的实验结果。该方法从网络资源分配的角度提出新的 RA 指标,对于网络中没有直接相连的两个节点 x 和 y , 它们的共同邻居可以成为媒介,协助建立链接。当网络的平均度较大的时候 RA 效果明显。Lü 等^[12]提出 LP 局部路径指标,该算法是在共同邻居 CN 指标的基础上考虑了三阶邻居的贡献,利用一个参数来控制三阶路径的作用大小。Katz^[13]提出的 Katz 算法考虑所有的路径数,对于短路径赋予较大的权重,对于长路径赋予较小的权重。Katz 算法提出的权重不易确定,同时 Katz 中的权值取值须小于邻接矩阵 A 最大特征值的倒数,这样才可以保证数值的收敛性,且计算复杂度较高。

对于链接强度的研究主要集中在具体网络节点的属性上,很少有文献研究从网络拓扑本身的角度挖掘权重信息。部分学者在链接强度上作了相关的工作。孙浩^[21]提出了基于时间信息的在线社交网络的链接预测框架,在该链接预测框架中,使用核函数来提取网络的时间特征来形成基于时间信息的加权概要图,在时间局部性和时间再现性基础上再使用扩展的关系贝叶斯分类器来进行链接预测。李玉华等^[22]给出一种基于链接重要性的动态链接预测方法,引入链接重要性的度量,对拓扑属性和语义相似度等属性进行修正,考虑动态性以反映时间因素对链接形成的影响。Murata 等^[23]针对对于社交网络中两个节点间的事务次数作为链接的权值,进行链接预测。

综上所述,现在的基于节点相似性的算法,如 CN、RA 等算法没有考虑链接权重信息,部分学者如孙浩^[21]、李玉华等^[22]借助网络的特有属性信息,针对特定网络为链接加权处理,然而网络的私有属性通常不容易获得,并且研究的普适性不强;现有的基于带权路径的典型算法,如 LP、Katz 中的权值确定通常采用简单的试探,很难达到最优解。考虑到基于节点相似的算法有简单,可扩展性强,又同时有一定的提升改进空间,本文主要针对基于节点相似算法的不足加以改进,提升预测的准确率。

2 基于链接重要性和数据场的链接预测算法

研究发现,在现实的网络链接通常有强弱之分,在网络中距离较近的两个节点可能存在某种相似性。因此,本文从链接的重要性和节点之间的吸引力方面计算节点之间的相似度。图 1 给出 WCDF 算法的流程。

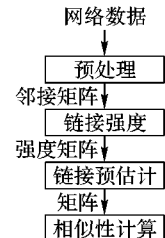


图1 WCDF 算法的流程

WCDF 算法先对网络中的边加权,在此基础上再对未链接的部分进行预估计,最后根据数据场定义的节点相似性计算最终的相似矩阵。以下将对链接强度、链接预估计和基于数据场的相似性 3 个部分进行详细介绍。

2.1 链接强度

Granovetter^[24]讨论了关系强度在信息流动及网络扩散中重要作用,指出一个关系的强度取决于周围网络的结构而不是关系本身的内在特征,并给出任选两个个体 x 和 y 以及与之相关的关系集合 S , x 与 y 之间的关系越强, S 中与 x, y 都相关的比例越大的结论。根据上述观点,将网络表示为带权值的邻接矩阵,无任何关系的两个节点之间用 0 表示,有链接的两个节点 x 和 y 之间的值,即 x 和 y 之间的链接强度计算如下:

$$Q_{xy} = \frac{\left| \frac{\varphi(x) \cap \varphi(y)}{\varphi(x) \cup \varphi(y)} \right| + \alpha}{k \times m + \alpha} \quad (1)$$

其中: $\varphi(x)$ 表示 x 的邻居集合, $\varphi(y)$ 表示 y 的邻居集合, $\varphi(x) \cup \varphi(y)$ 表示 x 和 y 的邻居的并集, $\varphi(x) \cap \varphi(y)$ 表示 x 和 y 的共同邻居, k 和 α 是一个调节参数, m 是整个网络关系强度的均值。很明显,现实网络中的关系通常是有紧密、强弱之分的,如朋友关系的亲密程度、合著网中的合作紧密度等,所以本文中的链接关系不是用简单的 0/1 矩阵来表示,而是参照 Granovetter^[24]对连接强度的定义进行计算而得。

2.2 链接预估计

现实中的网络中很大一部分是随机产生的,即节点之间是不存在共同邻居的,没有特定的规则。本文的出发点是对没有直接连接的两个节点之间相连的可能性进行预估计,考虑潜在链接的影响使其计算出随机链接的部分。先根据现有网络的链接情况,利用网络的拓扑信息对没有链接的两点连接概率进行估计,作为潜在可能的概率,再加上已存在的链接,从而实现对网络进行拓展,以降低网络的稀疏程度,且直接相连的值应大于无链接的值,同时随着路径长度的增加预估计值减小。此处以 RA 算法为基础,对未连接的部分节点 x 和节点 y 之间的链接预估计如下:

$$C_{xy} = \frac{\sum_{m \in \varphi(x) \cap \varphi(y)} \frac{1}{k(m)}}{l \times f\left(\sum_{z \in \varphi(e_1) \cap \varphi(e_2)} \frac{1}{k(z)}\right)} \quad (2)$$

其中: $\varphi(x) \cap \varphi(y)$ 表示 x 和 y 的共同邻居, $k(m)$ 表示节点 m

的度, $f\left(\sum_{z \in \varphi(e_1) \cap \varphi(e_2)} \frac{1}{k(z)}\right)$ 表示对网络中所有可能链接的两个节点的相似性求均值, l 为可调参数。

2.3 基于数据场的相似性

受重力场的启发,本文拟采用场来描述节点间非接触的相互作用。空间任意一点的势值与某个代表场源强度的参量成正比,与该点到场源间的距离成递减关系。在距离场源很远的地方仍然存在着场力的作用,代表长程场;而势值随着距离的增长急剧下降,相应力场很快衰减为0,代表短程场。根据场论知识,数据场的势函数形态需满足:定义在空间 Ω 上的连续、光滑、有限函数; $\varphi(x)$ 具有各向同性; $\varphi(x)$ 是对象 O 到场点 x 的距离的单值递减函数。距离为0时, $\varphi(x)$ 达到最大值;距离趋于无穷大时, $\varphi(x) \rightarrow 0$ 。考虑到现实网络的小世界效应,本文拟采用具有良好数学性质的高斯函数来定义数据场的标量势。

根据六度分割和三度分割的原理,短程场作用应更接近于现实的网络。给定网络 $G = \langle V, E \rangle$, 计算某节点对网络其他点的吸引力,吸引力越大,则将来链接的可能性越强。节点 v_i 与节点 v_j 间的相似性定义为:

$$g_{ij} = m_i \times m_j \times \sum_{d=2}^{ld} S(d) \times e^{-(d/\sigma)^2} \quad (3)$$

其中: d 为 v_i 与 v_j 间的路径长度; ld 用来表示需要计算的路径长度上线,公式中表示要计算 v_i 与 v_j 间长度为2到 ld 的所有路径; m_i 与 m_j 为节点的活跃度; $S(d)$ 为 v_i 与 v_j 间距离为 d 的路径条数; σ 为可调参数,决定节点的影响范围。根据高斯函数的数学性质,对于给定的 σ 值,每个节点的影响范围近似为 $3\sigma/\sqrt{2}$ 跳的局域区域,当距离大于 $3\sigma/\sqrt{2}$ 跳时,单位势函数快衰减为0。实际应用中根据网络的网络稀疏度和网络半径来确定 σ 值。根据普尔和科亨三度分割原理^[25] 此处 $ld = 3$ 。

2.4 WCDF 算法流程及复杂度分析

2.4.1 算法主要步骤

主要步骤如下:

- 1) 将网络表示为一个无向不加权图的邻接矩阵 A ;
- 2) 根据式(1)得到带权重的网络邻接矩阵 W ;
- 3) 根据式(2)对矩阵 W 进行补全,得到矩阵 WC ;
- 4) 根据矩阵 WC 计算两点间的路径长度为2,3的路径条数 $WC1$ 、 $WC2$;

5) 分别计算式(3)得到路径长度取值为1,2的场的势值记为 x 、 y ,代入 $WCDF = x * WC1 + y * WC2$,计算出最终的相似值。

2.4.2 算法复杂度分析

设 e 为平均每个节点的链接边数, n 为网络中总的节点数,由相关文献[10,12-13],CN、RA算法的复杂度为 $O(ne^2)$,LP算法的复杂度为 $O(ne^3)$,Katz算法的复杂度为 $O(n^3)$ 。WCDF算法的时间复杂度包含3个方面:1) 在处理链接强度中,因为是基于共同邻居计算 W 矩阵,所以计算复杂度等同于CN的计算复杂度,为 $O(ne^2)$;2) 在处理链接预估中,是对RA算法的改进,复杂度为 $O(ne^2)$;3) 在计算相似性时,主要是计算二、三阶路径时,算法的复杂度等同于LP算法,复杂度为 $O(n(e_1 + e_2)^3)$, e_1 、 e_2 分别为平均每个节点的直接链接边数和间接链接边数。因此,WCDF算法的复杂度为

$O(ne^2) + O(n(e_1 + e_2)^3)$,与LP算法的时间复杂度同级,远小于Katz算法的时间复杂度。

3 实验与分析

3.1 评价指标

链接预测主要应用于推荐系统和分类系统,评价指标如下。

1) 推荐系统常用的度量方法是 DCG (Discounted Cumulative Gain),其定义如下:

$$DCG(x) = \sum_{i=1}^{|x|} \frac{2^{R(x_i)} - 1}{\ln(1 + i)} \quad (4)$$

其中: x_i 代表 x 中排在第 i 位的文档, $R(x_i)$ 表示 x_i 相关度。在实验中,存在链接的两个节点 $R(x_i)$ 取值为1;否则为0。根据DCG的定义,排位相同时,文档的相关度越高,其对DCG的贡献就越高,而相关度相同时,文档的排位越前,其对DCG的贡献就越高。

2) 分类系统常用度量方法是 AUC (Area Under Curve),其定义如下:

$$AUC = \frac{n' + 0.5 * n''}{n} \quad (5)$$

其中: n 为随机抽取一个正反例的实验次数, n' 代表 n 次实验中正例的相似值大于反例的次数, n'' 表示 n 次实验中正例的相似值等于反例的次数。

3.2 实验数据

实验采用文献[26]提供的两个具有代表性的社会网络数据 Flickr 和 Arnet。Flickr 是一个关于图片分享的社交网络数据集,来源于 <http://www.flickr.com>,数据包含从2006-11-01—2007-03-20的网络数据;数据以10 d为间隔,按时间顺序共分成14个数据集,分别记为 d1, d2, ..., d14。Arnet 是一个关于学术研究网络的数据集,来源于 <http://www.arnetminer.org>;该数据以年为间隔,共分10个数据集,分别记为 ar1, ar2, ..., ar10。实验中,本文仅取 Flickr 的部分数据,分别为 Flickr-1 和 Flickr-2。数据的详细信息如表1所示。

表1 数据信息

数据集	训练节点数	训练边数	测试节点数	测试边数	实际预测的边数	网络密度
Flickr-1	2152	11 649	2152	13 859	1836	0.005 0
Flickr-2	2152	11 093	2152	11 649	474	0.004 8
Arnet	2062	6 590	2062	8 056	102	0.003 1

表1中,训练节点数、训练边数、测试节点数、网络密度等是通过实验统计得到的。

3.3 实验内容

实验分成两组:一组实验是对 WCDF 算法中的参数 σ 进行实验,在 Flickr、Arnet 数据集中分别取不同的 σ 值进行对比,分析不同取值的影响;另一组实验是对 WCDF 的性能与其他算法进行了比较测试。

3.3.1 WCDF 参数确定

式(3)中, σ 因子决定了不同路径长度所起的作用。本文采用简单试探方法求不同的 σ 对预测准确度的影响。根据高斯函数的数学性质和式(3),为了直观表示,设 $\sigma = \sqrt{2}k/3$,可以得出 σ 和 k 是线性关系,从而用 k 的变化对预测的影响来反

映 σ 的影响。图 2 ~ 5 给出了 k 值对评价指标的影响曲线。

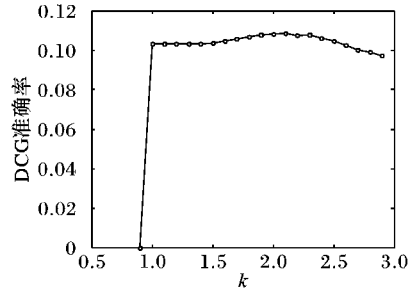


图 2 Flickr-1 中不同 k 值对 DCG 准确率的影响

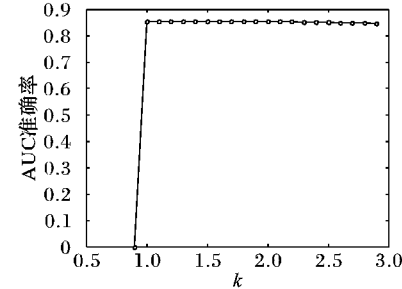


图 3 Flickr-1 中不同 k 值对 AUC 准确率的影响

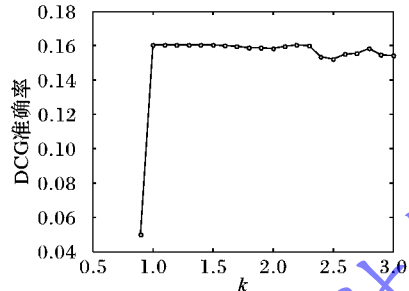


图 4 Arnet 中不同 k 值对 DCG 准确率的影响

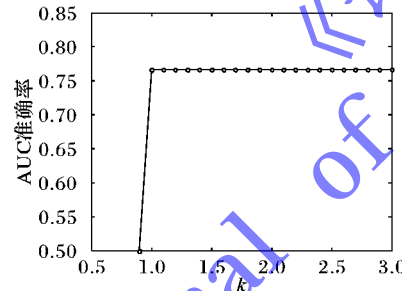


图 5 Arnet 中不同 k 值对 AUC 准确率的影响

从图 2 ~ 5 可以看出, σ 的取值达到一定的范围后, 在很长区间内对 AUC 和 DCG 的影响基本趋于稳定状态。考虑到计算代价, 对于 WCDF 算法, 使用时可以在固定区间 $[1, 3]$ 内使用几次简单试探法寻求 σ 的次优解。

3.3.2 性能测试

实验中, 用 Flickr-1 的 d1 到 d9 的数据作为训练集, d14 数据为测试集; 用 Flickr-2 的 d1 数据作为训练集, d10 数据作为测试集。对于 Arnet 数据集, 用其中的 ar1 数据集作为训练集, ar10 数据集作为测试集。对带权路径算法 LP、Katz 采用简单试探法在 $[0, 1]$ 区间找出最优的权值, 此处的间隔取 0.005。数据属于极度不平衡的数据, 正反例比为 1:1000, 因此实验中为了避免数据的不平衡性对结果的影响, 随机选出等数量的反例进行, 并且多次进行, 每个链接进行 10000 次的随机实验。实验中证明进行 10000 次实验结果趋于稳定, w 为最优的权值。表 2 给出了 AUC 的实验对比结果。

表 2 几种算法的 AUC 的实验结果对比

算法	AUC		
	Flickr-1	Flickr-2	Arnet
CN 算法	68.45	69.53	65.45
RA 算法	68.54	69.48	65.51
LP 算法	79.28	81.79	71.88
Katz 算法	83.57	83.74	81.93
WCDF 算法	85.40	84.85	76.62

从表 2 中可以看出, 以 AUC 为精确度指标时, 对于 Flickr-1 和 Flickr-2 数据, WCDF 略好于 Katz, 提高约 1 到 2 个百分点, 优于 LP 算法 3 到 6 个百分点; 对于 Arnet 数据 Katz 优于 WCDF 约 5 个百分点, WCDF 优于 LP 约 5 个百分点。WCDF 和 Katz 的优势比较明显。CN 和 RA 算法明显落后。WCDF 算法比其他算法整体上具有优势。

本文同时也进行了 DCG 的性能对比实验。由于在推荐系统中排名太靠后的结果没有太大意义, 所以只取前 50 个推荐结果的 DCG 值, 并将最后结果归一化为预测点为 100 的 DCG 值, 且 w 为实验得出的最优的权值。表 3 给出了 DCG 的实验结果。

表 3 几种算法的 DCG 的实验结果对比

算法	DCG		
	Flickr-1	Flickr-2	Arnet
CN 算法	7.87	7.27	11.42
RA 算法	9.15	8.82	14.13
LP 算法	9.32	9.44	13.45
Katz 算法	9.22	8.93	13.68
WCDF 算法	10.71	10.85	16.04

从表 3 可知, 以 DCG 为精确度指标时, 在 Flickr-1、Flickr-2 和 Arnet 数据中 WCDF 好于 LP、Katz; 提高了约 1.5 到 2.5 个 DCG 分值。在 Arnet 中 RA 方法比 LP 方法有优势, Katz 方法不占优势, WCDF 方法较有优势, LP 算法和 RA 算法次之。

表 4 几种算法的权值 w 的最优值对比

算法	w		
	Flickr-1	Flickr-2	Arnet
LP 算法	0.150	0.405	0.250
Katz 算法	0.015	0.015	0.070
WCDF 算法	2.000	2.000	1.000

总之, 在两个评价指标的实验中, WCDF 算法与其他 4 种方法算法相比整体上有优势。与已有算法相比, 本文算法还有如下优点:

1) 与具有相同时间复杂度的 LP 算法相比, WCDF 整体提高了准确率, 比较高时间复杂度的 Katz 方法略有优势, 并且 WCDF 参数的确定简单, 容易求出次优解。而 Katz 的参数不仅有限制, 而且每次参数的变化 Katz 都要重新计算, 调整参数时间代价较高, 而 WCDF 改变权值时和 LP 一样, 不需要重新计算路径数信息。

2) WCDF 算法在以 AUC 和 DCG 两种衡量指标中都有很好的整体效果, 而 Katz 在以 AUC 为衡量指标时占优势, 以 DCG 为衡量指标时不占优势, LP 和 RA 在以 AUC 为衡量指标时不占优势, 在以 DCG 为衡量指标时较占优势。

3) 原有的 LP、Katz 算法中的 β 值没有固定的区间, 不稳

定,很难确定出优解,而且在 Katz 算法中 β 的取值须小于邻接矩阵 A 最大特征值的倒数,这样才可以保证数列的收敛性,每次参数的改变都需要重新计算,调整参数时间代价很大。WCDF 算法参数的确定简单、直观:当参数 σ 达到一定的值后在很长区间内趋于稳定,采用几次简单的试探便能够在一定区间内找到很好的次优解,参数确定更直观,且参数的改变不需要重新计算路径数,调整参数时代价小。

4 结语

针对已有基于节点相似性的链接预测方法的缺陷,提出一种基于链接重要性和数据场的链接预测方法。该方法从邻居节点的角度考虑了链接重要性的影响,同时在已有算法的基础上对没有直接相连的节点之间进行链接可能性进行预估计,且提出利用数据之间的吸引力来计算不同节点之间的相似性。本文方法在不增加时间复杂度的情况下,与其他算法相比,整体上提高了准确率,并且参数调节简单、直观,代价小,有一定的实用价值。但新方法中暂时只考虑了两节点间的拓扑信息,而节点的活跃性信息还没有计算在内,下一步工作将加入节点的活跃信息以及其他网络属性特征来计算节点相似度。

参考文献:

- [1] LIBEN-NOWELL D, KLEINBERG J. The link prediction problem for social networks [J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(7): 1019–1031.
- [2] HUANG Z, LI X, CHEN H. Link prediction approach to collaborative filtering [C]// JCDL'05: Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries. Piscataway: IEEE Press, 2005: 141–142.
- [3] BISCHOFF K, FIRAN C, GEORGESCU M. Social knowledge-driven music hit prediction [C]// ADMA 2009: Proceedings of the 5th International Conference on Advanced Data Mining and Applications. Berlin: Springer, 2009: 43–54.
- [4] PAVLOV M, ICHISE R. Finding experts by link prediction in co-authorship networks [EB/OL]. [2014-01-15]. <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-290/paper04.pdf>.
- [5] JEH G, WIDOM J, SIMRAN K. A measure of structural context similarity [C]// Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, ACM Press, 2002: 538–543.
- [6] BLONDEL V D, GAJARDO A, HEYMANS M, et al. A measure of similarity between graph vertices: applications to synonym extraction and Web searching [J]. *SIAM Review*, 2004, 46(4): 647–666.
- [7] SARUKKAI R R. Link prediction and path analysis using Markov chains [J]. *Computer Networks*, 2003, 33(6): 377–386.
- [8] FRIEDMAN N, GETOOR L, KOLLER D, et al. Learning probabilistic relational models [C]// IJCAI'99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers, 1999: 1300–1309.
- [9] HECKERMAN D, MEEK C, KOLLER D. Probabilistic entity-relationship models, PRMs, and plate models [C]// Proceedings of the ICML-2004 Workshop on Statistical Relational Learning and its Connections to Other Fields. Banff: [s. n.], 2004: 55–60.
- [10] ZHOU T, LÜ L, ZHANG Y C. Predicting missing links via local information [J]. *The European Physical Journal B*, 2009, 71(4): 623–630.
- [11] DONG Y, KE Q, WU B. Link prediction based on node similarity [J]. *Computer Science*, 2011, 38(7): 162–199. (东昱晓, 柯庆, 吴斌. 基于节点相似性的链接预测[J]. 计算机科学, 2011, 38(7): 162–199.)
- [12] LÜ L, JIN C H, ZHOU T. Similarity index based on local paths for link prediction of complex networks [J]. *Physical Review E*, 2009, 80(4): 046122.
- [13] KATZ L. A new status index derived from sociometric analysis [J]. *Psychometrika*, 1953, 18(1): 39–43.
- [14] SHEN Y. Link prediction in complex networks [D]. Guangzhou: South China University of Technology, 2011. (沈勇明. 复杂网络链接预测[D]. 广州: 华南理工大学, 2011.)
- [15] TENG Z. Research on link prediction for dynamic networks [D]. Jinan: Shandong University, 2012. (滕兆明. 动态多维社会网络链接预测研究[D]. 济南: 山东大学, 2012.)
- [16] YIN H. Research on link prediction for social networks [D]. Changchun: Jilin University, 2012. (殷涵. 社会网络链接预测研究[D]. 长春: 吉林大学, 2012.)
- [17] ZHAO C. Link prediction for social networks [D]. Harbin: Harbin Institute of Technology, 2012. (赵婵媛. 社会网络链接算法研究[D]. 哈尔滨: 哈尔滨工业大学, 2012.)
- [18] GAO S, DENOYER L, GALLINARI P. Temporal link prediction by integrating content and structure information [C]// Proceedings of the 20th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2011: 1169–1174.
- [19] CUI A, FU Y, SHANG M, et al. The local structure of complex networks sprung up: common neighbor drive network evolution [J]. *Acta Physica Sinica*, 2011, 60(3): 803–808. (崔爱香, 傅彦, 尚明生, 等. 复杂网络局部结构的涌现: 共同邻居驱动网络演化[J]. 物理学报, 2011, 60(3): 803–808.)
- [20] ADAMIC L A, ADAR E. Friends and neighbors on the Web [J]. *Social Networks*, 2003, 25(3): 211–230.
- [21] SUN H. Research on link prediction based on transactional information [D]. Qinhuangdao: Yanshan University, 2010. (孙浩. 基于事物信息的链接预测算法研究[D]. 秦皇岛: 燕山大学, 2010.)
- [22] LI Y, XIAO H, LI D, et al. Research of dynamic link prediction method based on link importance [J]. *Journal of Computer Research and Development*, 2011, 48(z2): 40–46. (李玉华, 肖海岭, 李栋才, 等. 基于链接重要性的动态链接预测方法研究[J]. 计算机研究与发展, 2011, 48(z2): 40–46.)
- [23] MURATA T, SAKIKO M. Link prediction of social networks based on weighted proximity measures [C]// Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence. Washington, DC: IEEE Computer Society, 2007: 86–88.
- [24] GRANOVETTER M S. The strength of weak ties [J]. *Journal of American Sociology*, 1973, 78(6): 1368–1178.
- [25] WATTS D. Network dynamics between order and disorder [M]. CHEN Y, translate. Beijing: China Renmin University Press, 2006: 14–17. (WATTS D. 有序与无序之间的网络动力学[M]. 陈禹, 译. 北京: 中国人民大学出版社, 2006: 14–17.)
- [26] TAN C, TANG J, SUN J, et al. Social action tracking via noise tolerant time-varying factor graphs [C]// KDD'10: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2010: 1049–1058.