

基于交互意见和地位理论的符号网络链接预测模型

王鑫^{1,3,4} 王英^{2,3} 左万利^{2,3}

¹(长春工程学院计算机技术与工程学院 长春 130012)
²(吉林大学计算机科学与技术学院 长春 130012)
³(符号计算与知识工程教育部重点实验室(吉林大学) 长春 130012)
⁴(广西可信软件重点实验室(桂林电子科技大学) 广西桂林 541004)
(xinwangjlu@gmail.com)

Exploring Interactional Opinions and Status Theory for Predicting Links in Signed Network

Wang Xin^{1,3,4}, Wang Ying^{2,3}, and Zuo Wanli^{2,3}

¹(School of Computer Technology and Engineering, Changchun Institute of Technology, Changchun 130012)
²(College of Computer Science and Technology, Jilin University, Changchun 130012)
³(Key Laboratory of Symbolic Computation and Knowledge Engineering (Jilin University), Ministry of Education, Changchun 130012)
⁴(Guangxi Key Laboratory of Trusted Software (Guilin University of Electronic Technology), Guilin, Guangxi 541004)

Abstract With the wide spread and pervasion of social network, it brings more opportunities and novel problems for deep research on signed network, where link prediction is one of key problems in signed network. Interactional opinions and status theory are contributed to explain the construction and sign property of link relations, and provide theoretical principles for improving prediction quality. Therefore, this paper investigates link prediction problem in signed network from the perspective of interactional opinions and status theory, and constructs link prediction model by studying the strong correlation between two inducements and link relationship. Firstly, it explores interactional opinions to enhance the reliability of the decomposed matrix, and makes up for the limitations of status theory. Then, it models interactional opinions as enhanced reliability factor of matrix, and models status theory as the regularization terms. Finally, we construct the model of link prediction in signed network, namely MF-SI. Experimental results demonstrate that the model of MF-SI owns the best prediction quality compared with other baseline methods, which shows that the method of integrating interactional opinions with status theory implements link prediction in signed network.

Key words signed network; link prediction; sparse learning; interactional opinions; status theory

摘要 随着社会网络的普遍和流行,社会网络为符号网络(signed network)的深入研究提出了更多的全新问题,其中链接预测是符号网络研究的核心问题之一.交互意见和地位理论能够较好地解释符号网络中链接关系的构建和链接的符号属性,二者作为链接构建的诱因为提高链接预测的质量提供了理论

收稿日期:2015-12-12;修回日期:2016-02-03
基金项目:国家自然科学基金项目(61300148);吉林省科技计划青年科研基金项目(20130522112JH);广西可信软件重点实验室研究课题资助(kx201533)
This work was supported by the National Natural Science Foundation of China (61300148), the Science and Technology Development Program of Jilin Province of China (20130522112JH), and Guangxi Key Laboratory of Trusted Software (kx201533).
通信作者:王英(wangying2010@jlu.edu.cn)

基础.因此,通过研究交互意见和地位理论与链接关系的强相关性,构建符号网络链接预测模型.首先,利用交互意见增强待分解矩阵的可靠度,弥补由地位理论的对称性所带来的局限性;然后,在稀疏学习矩阵分解模型基础上,将交互意见建模为增强可靠度因子,同时将地位理论建模为稀疏学习模型的正则化项;最后,构建基于交互意见和地位理论的符号网络链接预测模型 MF-SI.实验结果表明相比于其他基本方法,MF-SI 模型获得了较好的预测质量,说明集成交互意见和地位理论能够较好地实现符号网络链接预测问题.

关键词 符号网络;链接预测;稀疏学习;交互意见;地位理论

中图法分类号 TP181

符号网络(signed network)是具有正或负符号属性标记的有向网络,其中正向链接(positive link)表示积极关系,负向链接(negative link)表示消极关系^[1].具体而言,符号网络中的正向链接使用“+”表示,表达了2个节点之间具有信任、朋友、支持和喜欢等积极关系;而负向链接使用“-”表示,表达了2个节点之间具有敌对、不信任、讨厌和反对等消极关系.特别是在社交媒体领域,用户之间借助社交网络平台表达信任与不信任关系、构建朋友与敌对关系、对论坛内容发表支持与反对观点等.符号网络中链接预测问题是利用网络拓扑结构数据和用户行为数据预测尚未建立链接的2个节点之间是否可能建立链接关系以及标记链接关系的符号属性.

符号网络在机器学习和数据挖掘领域研究中具有重要的研究意义和应用价值,并引起越来越多研究人员的关注.因此,提出了高效的链接预测模型用于实现正向链接和负向链接的预测,并在 Epinions 数据集上构建多组实验,实验结果证明所提出的模型能够准确地预测符号网络中的链接关系.主要贡献如下:

1) 提出有效的量化策略分别量化交互意见和地位理论;

2) 将交互意见模拟为增强可靠度因子,模拟地位理论为稀疏学习模型的正则化项,利用矩阵分解学习模型融合交互意见和地位理论构建链接预测模型 MF-SI 预测链接关系,进而构建符号网络;

3) 在 Epinions 数据集上评估所提出的模型 MF-SI,构建多组实验验证模型的有效性,并详细地阐述不同链接构建诱因以及不同量化策略对链接正负预测质量的影响.

1 相关性研究

最早的符号网络研究起源于20世纪40年代的

社会心理学领域,Heider^[2]提出不同类型的社会关系对个体心理平衡产生不同影响,从心理学角度研究人在认知过程中所产生的正向关系和负向关系之间的相互作用. Cartwright 和 Harary^[3]在 Heider 的基础上系统地将线性的数学理论应用于 Heider 平衡理论中,对其形式化描述为正、负属性的符号网络.现有符号网络的链接预测问题采用的方法包含2个方面:监督学习和非监督学习方法.监督学习将链接预测问题视为在已知链接标识的基础上利用监督学习算法,如决策树^[4]、迁移学习^[5]、组合分类器^[6]等实现分类问题;非监督学习利用网络结构的拓扑属性和交互行为进行分析预测,比如节点相似性^[7-8]、链接存在的最大似然^[9]和矩阵处理^[10-11]等.

监督学习方法首先收集带有符号属性的链接标记数据,然后为用户对构建各种类型的特征.针对网络拓扑结构特征而言,正负链接中节点的出入度特征^[12]和基于平衡理论的三角形特征^[13]较为常见. Leskovec 等人^[12]重点考虑节点出入度等局部测度特征和待预测边所处的结构平衡理论三角形关系模式特征,并利用逻辑回归(logistic regression)方法训练模型对边进行分类并获得了较高预测准确度.实验证明符合结构平衡理论和地位理论的三元组对于待预测边的符号属性预测起到了重要作用. Chiang 等人^[14]认为平衡理论三角形关系特征并不具备鲁棒性,因为有的符号社会网络较为稀疏并且大多数用户并没有基于平衡理论三角形关系特征,进而提出引入长度为 k 的环特征量化不平衡指标,再利用分类算法预测边的符号属性.该方法从全局的角度出发利用网络的不平衡程度进行预测,从而实现对 Leskovec 等人局部度量方法的扩展.除了网络拓扑结构数据之外,用户的属性信息^[15]和情境信息^[6]也可作为参考特征,比如兴趣、职业、性别、家乡等. Patidar 等人^[15]首先通过训练基于用户属性信息的分类器预测未知链接,然后根据平衡理论判断是否

保持或扩大平衡指数来预测未知链接. Zolfaghar 等人^[6]利用符号网络拓扑结构和情境信息构建熟识度、声誉度、相似度和个性化 4 个因素的特征集合,然后利用组合分类器算法预测网络中链接的符号属性.

非监督学习方法通常是根据符号网络中拓扑结构数据以实现链接关系的预测. 一般来说,非监督学习方法依据不同策略可将链接预测分成 2 个方面:

1) 基于相似度的方法. 首先定义相似度量化策略用于衡量节点的相似度,然后根据相似度推断链接的符号属性. 2) 基于矩阵处理的方法. 该方法主要包括传播模型、矩阵分解^[16]或矩阵填充方法预测链接的符号属性. 在基于相似度的方法中, Javari 等人^[7]提出一种基于聚类 and 协同过滤的方法用于预测符号网络中正向和负向链接,首先在符号网络中引入簇之间的条件相似度概念,并将网络聚类成一定数量的簇,使其每个簇都满足平衡性,然后应用基于用户的协同过滤算法实现链接的预测,有效地解决了符号网络的稀疏性问题. Symeonidis 等人^[8]提出一种基于朋友推荐的算法,首先实现拉普拉斯矩阵谱聚类算法,然后定义同一簇的节点相似度和不同簇的节点相似度的计算方法,最后借助朋友推荐算法实现符号网络中的链接预测. 在矩阵处理方法中,最早的信任传播模型研究是由 Ramanathan 等人^[17]将负关系融入信任传播模型,利用 4 种基本的信任传播模式,实现在拥有较少的信任或不信任关系基础上以较高的准确度预测网络中 2 个用户之间的信任或不信任关系. 矩阵分解是将矩阵分解成多个矩阵的乘积,利用低秩矩阵不断地逼近原矩阵进而补齐原矩阵中的缺失值,达到预测符号网络链接正负属性目的. 基于矩阵分解或矩阵填充方法, Hsieh 等人^[10]验证在符号网络中弱结构平衡往往会产生全局的低秩模型,然后将链接预测问题转化为低秩矩阵填充问题.

综上所述,虽然对符号网络链接预测模型的研究已有很多,大部分工作都是在 2 个节点之间已经存在链接关系的假设前提下预测链接的符号属性^[18]. 但是,同时完成符号网络中任意 2 个节点是否建立链接关系以及预测链接关系的符号属性,在预测难度和准确率方面遇到了极大困难和挑战,这正是本文的工作重点和贡献. 社会学理论有助于解释社会现象,从而获得对社会现象发生、发展的规

律性认识,而带有符号属性的社会网络是由多个节点构成的一种社会结构,能够反映现实世界中用户之间的联系. 因此,将社会学理论引入符号网络链接预测的研究,将会在很大程度上提高链接预测的质量^[19-20].

2 问题形式化描述

用 $D=[Z, E, R, G]$ 表示 Epinions 数据集中所有数据的集合. 其中, $Z=\{u_1, u_2, \dots, u_n\}$ 表示用户集合, n 为用户数量; $E \in \mathbb{R}^{n \times m}$ 表示用户-评论关联矩阵, m 为评论数量, $E_{ij}=1$ 表示用户 u_i 发表评论 r_j , 否则 $E_{ij}=0$; $R \in \mathbb{R}^{n \times m}$ 表示用户对评论的帮助评级矩阵,如果用户 u_i 评级了评论 r_j 的帮助程度,那么用 R_{ij} 表示帮助评级的分值; $G \in \mathbb{R}^{n \times n}$ 表示用户之间符号网络矩阵,如果 $G_{ij}=1$ 表示用户 u_i 与用户 u_j 之间存在正向链接关系,用矩阵 $P \in \mathbb{R}^{n \times n}$ 表示,而 $G_{ij}=-1$ 表示用户 u_i 与用户 u_j 之间存在负向链接关系,用矩阵 $N \in \mathbb{R}^{n \times n}$ 表示.

针对以上描述,符号网络链接预测问题转变为设计预测框架 f ,利用网络拓扑结构数据和用户的评级行为数据预测未知的正向和负向链接关系,即给定已有的符号网络关系矩阵 G 、用户-评论矩阵 E 、用户对评论的帮助评级矩阵 R ,目的是学习预测框架 f 预测未知的正负向链接关系矩阵 P' 和 N' ,即:

$$f = \{G, E, R\} \rightarrow \{P', N'\}. \tag{1}$$

3 数据分析以及链接构建诱因量化

3.1 数据分析

Epinions.com 是著名的商品评价的社交网站,是提供消费者对所购买的商品进行评论的平台,公正的评论值得用户信任,帮助用户确定购买决策. Epinions 数据集^①的处理过程如下: 1) 获取由用户创建的信任和不信任链接关系,以及由用户行为数据构成的用户评价信息. 对于链接关系,主要包括施信者(trustor)与受信者(trustee)之间建立的正向链接和负向链接,以及链接关系构建时间;对于用户行为数据,主要包括用户对商品的评论以及用户对评论的帮助评级. 2) 将用户的所有评级按照时间偏序排列,过滤掉构建链接关系时间之后的所有评级信息. 3) 由于不活跃或不重要的用户在一定程度上会

① http://www.trustlet.org/wiki/Downloaded_Epinions_dataset

影响评估结果,将少于 2 个施信者、2 个评论及 2 个评级信息的用户过滤掉.表 1 为处理后的 Epinions 数据集统计结果.

Table 1 Statistics of Epinions Dataset

表 1 Epinions 数据集统计结果

Item	Value
Number of Users	12 135
Number of Positive Links	251 465
Number of Negative Links	46 817
Number of Reviews	575 156
Number of Ratings	6 443 575

3.2 链接构建诱因量化

3.2.1 交互意见量化

在 Epinions 中,交互意见倾向通过用户对其他用户评论的帮助评级体现,这种意见倾向在一定程度上体现了链接关系方向属性和符号属性.直观理解,如果用户 u_i 对用户 u_j 评级越高,表明认可程度越大,那么更易产生从用户 u_i 到用户 u_j 的正向链接关系,相反评级越低越可能产生负向链接关系.用 $H_{ij} = R_{i,*} \cdot E_{j,*}^T$ 表示用户 u_i 对用户 u_j 的所有评论帮助评级集合, $H_{ij} = \{r_{ij1}, r_{ij2}, \dots, r_{ijm}\}$, 其中 r_{ijk} 表示用户 u_i 对用户 u_j 的第 k 个评论帮助评级分值.首先,定义施信者评级分数的平均值 $\bar{r}_{tr}$ 为

$$\bar{r}_{tr} = \frac{\sum_k \sum_m r_{ikm}}{|R_{i,*}|}, \quad (2)$$

然后,针对施信者评级分数的平均值 $\bar{r}_{tr}$,通过交互数量量化用户 u_i 对用户 u_j 的正向意见倾向 $ET^+(u_i, u_j)$:

$$ET^+(u_i, u_j) = |(H_{ij} - \bar{r}_{tr}) > 0|, \quad (3)$$

相反,负向意见倾向 $ET^-(u_i, u_j)$ 的量化公式如下:

$$ET^-(u_i, u_j) = |(H_{ij} - \bar{r}_{tr}) < 0|, \quad (4)$$

其中, $|\cdot|$ 表示集合的大小.

3.2.2 地位理论量化

在 Epinions 中,由于缺乏权威的第三方认定机构对用户社会地位的评估,只能借助网络拓扑结构量化策略衡量用户在社会网络中的地位.在符号网络中,正向链接入度有助于提升节点的地位,而负向链接入度降低节点的地位.由于经典 PageRank 算法未考虑入度链接的符号属性对节点地位的影响,因此提出引入正负向链接的地位理论的量化策略,用于计算用户的社会地位,即 *PolarityRank* 和 *PolarityIndegree*.

1) *PolarityRank*. 借鉴 PageRank 算法的思想,从全局角度衡量节点在符号网络中的重要地位. *PolarityRank* 被定义为

$$PolarityRank(u_i) = \frac{PR^+(u_i) - PR^-(u_i)}{PR^+(u_i) + PR^-(u_i)}, \quad (5)$$

其中, $PR^+(u_i)$ 和 $PR^-(u_i)$ 分别表示正向链接和负向链接对用户 u_i 社会地位变化的贡献度,其公式为

$$PR^+(u_i) = \frac{1-\alpha}{n} + \alpha \left(\sum_{j \in In^+(u_i)} \frac{PR^+(u_j)}{|Out(u_j)|} - \sum_{j \in In^-(u_i)} \frac{PR^-(u_j)}{|Out(u_j)|} \right), \quad (6)$$

$$PR^-(u_i) = \frac{1-\alpha}{n} + \alpha \left(\sum_{j \in In^-(u_i)} \frac{PR^-(u_j)}{|Out(u_j)|} - \sum_{j \in In^+(u_i)} \frac{PR^+(u_j)}{|Out(u_j)|} \right),$$

其中,参数 α 表示阻尼系数,通常将其设置为 0.85; $Out(u_i)$ 表示节点 u_i 的出度数; $In^+(u_i)$ 和 $In^-(u_i)$ 分别表示节点 u_i 正向入度数和负向入度数.

2) *PolarityIndegree*. 借鉴投票机制,利用正、负向的入度链接数量,从局部角度衡量节点在符号网络中的重要地位. *PolarityIndegree* 被定义如下:

$$PolarityIndegree = \frac{PI^+(u_i) - PI^-(u_i)}{PI^+(u_i) + PI^-(u_i)}, \quad (7)$$

其中, $PI^+(u_i)$ 和 $PI^-(u_i)$ 分别采用 sigmoid 函数形式衡量正、负向链接入度的贡献程度,定义如下:

$$PI^+(u_i) = \frac{1}{1 + e^{-\alpha \times In^+(u_i)}}, \quad (8)$$

$$PI^-(u_i) = \frac{1}{1 + e^{-\alpha \times In^-(u_i)}},$$

其中,参数 α 表示控制入度链接的权重系数.

4 链接构建诱因的动机观察

利用 Epinions 数据集,以交互意见和地位理论为理论依据分析用户之间建立链接的诱因.由于很难准确识别出建立诱因,仅仅通过最简单和直接的方式,即观测社会网络中真实的数据,针对不同诱因,利用网络拓扑数据和用户行为数据展开深入分析,进而产生设计符号网络链接预测框架的动机.

4.1 交互意见

为了验证交互意见与关系符号属性之间的关联关系,利用正向和负向交互意见量化策略尝试回答 2 个问题:

1) 施信者的正向意见倾向大于负向意见倾向是否更可能与受信者建立正向链接关系?

2) 施信者的负向意见倾向大于正向意见倾向是否更可能与受信者建立负向链接关系?

为了回答以上 2 个问题,首先针对每对用户关系量化施信者对受信者的正向意见倾向 ET^+ 与负向意见倾向 ET^- 的差值;然后统计正向关系和负向关系情感差异的密度分布.从图 1 可以发现,大多数信任关系(几乎达到 92%)分布在意见差异值 $[0, 1]$ 的区间内;相反,不信任关系的分布情况稍微复杂,不信任关系分布最多的 2 个位置分别在 -1 和 $+1$,但是,仍然可从分布的结果中发现将近 69%的不信任关系分布在意见差异值的 $[-1, 0]$ 区间内.不信任关系分布复杂的原因是:在 Epinions 中,用户的评级数据较为倾斜(skewed),即大多数评级都分布在分值 4~5 之间.这说明绝大部分用户更加愿意对评论做出较为积极正向的评级,即使建立的关系是不信任关系.

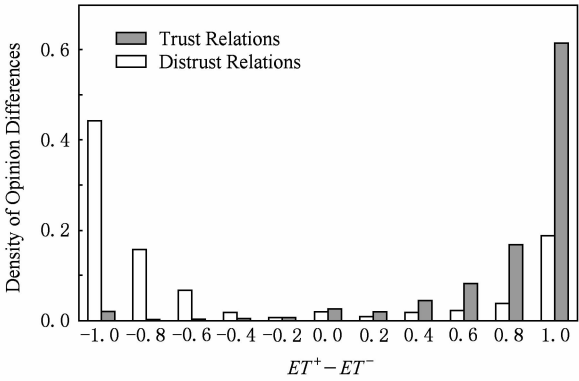


Fig. 1 Density distribution of opinion differences.
图 1 意见差异密度分布情况

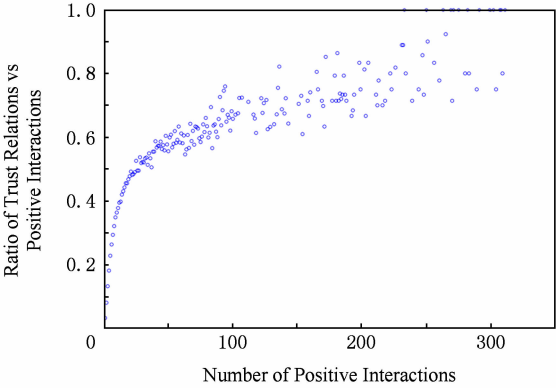
另外,为了进一步验证交互意见的数量与建立链接关系的正负属性之间存在的关联关系,尝试回答 2 个问题:

- 1) 2 个用户之间正向交互意见倾向数量越多,这 2 个用户之间更可能构建正向链接?
- 2) 2 个用户之间负向交互意见倾向数量越多,这 2 个用户之间更可能构建负向链接?

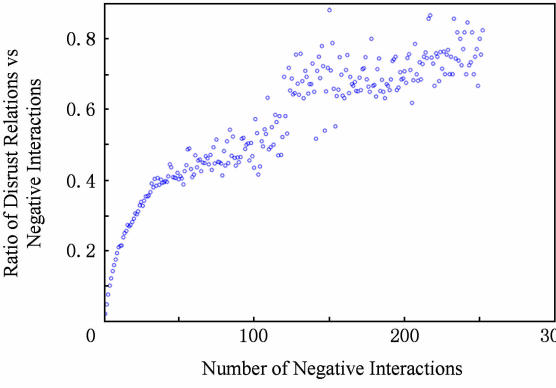
为了回答第 1 个问题,首先形式化定义一些符号,如 PI_{ij} 表示用户 u_i 与用户 u_j 的正向交互数量; U_k 表示用户 u_i 与用户 u_j 正向交互数量不小于 K 的用户对 $\langle u_i, u_j \rangle$ 集合, $U_k = \{ \langle u_i, u_j \rangle \mid PI_{ij} \geq K \}$,其中 K 表示正向交互数量; PR_k 表示在 U_k 中正向关系对的集合, $PR_k = \{ \langle u_i, u_j \rangle \mid (PI_{ij} \geq K) \wedge T_{ij} = 1 \}$.然后,计算正向链接关系数量同正向交互数量比率 P_K :

$$P_K = \frac{|PR_K|}{|U_K|}. \tag{9}$$

正向链接关系数量同正向交互数量比率 P_K 随着 K 变化的比率分布如图 2(a) 所示.随着正向交互数量 K 的增加,比率 P_K 往往也随着增加,表明正向交互数量与正向链接关系存在较大的相关性,大部分正向链接关系是基于交互行为的意见倾向而构建,正向交互越多的用户之间更可能构建正向链接关系.这样的分布结果也正好回答了第 1 个问题.



(a) Positive Interaction



(b) Negative Interaction

Fig. 2 Ratios of the number of links with respect to the number of interactions.

图 2 链接数量同交互数量比率分布情况

为了回答第 2 个问题,采取相反的方式定义负向链接关系数量同负向交互数量的比率 N_K ,其计算公式为

$$N_K = \frac{|NR_K|}{|U_K|}. \tag{10}$$

负向链接关系数量同负向交互数量比率 N_K 随着 K 变化的分布情况如图 2(b) 所示.随着负向交互数量 K 的增加,比率 N_K 往往也随之增加,其结果表明负向交互数量与负向链接关系存在较大的相关性,相似的分布结果正好回答了第 2 个问题.

4.2 社会地位理论

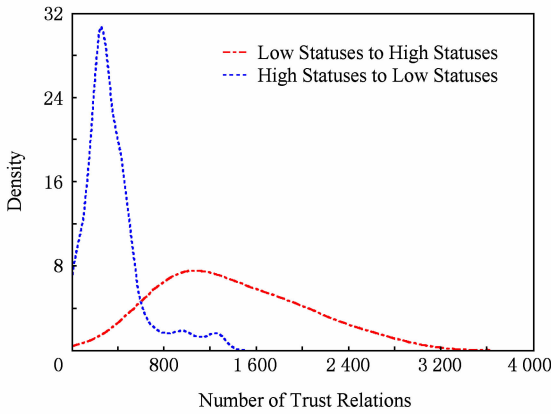
为了验证地位理论与关系符号属性之间的关联

关系,利用 *PolarityRank* 策略量化符号网络节点的地位等级信息以尝试回答 2 个问题:

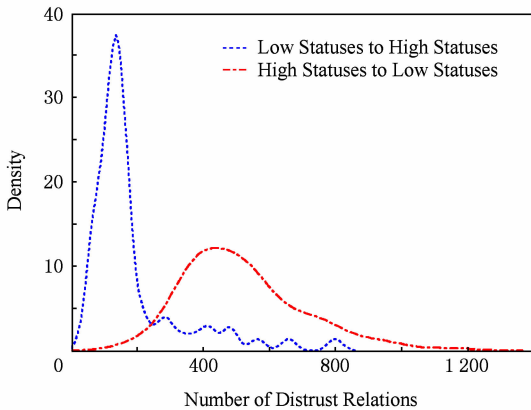
1) 在符号网络中,低地位等级用户是否更可能与高地位等级用户建立正向链接关系?

2) 在符号网络中,高地位等级用户是否更可能与低地位等级用户建立负向链接关系?

为了回答以上 2 个问题,首先将所有的用户依据 *PolarityRank* 值降序排列,并将用户的地位等级划分为 K 个等级,在验证中将设置用户等级的数量 $K=10$. 然后,针对信任关系分布统计构建 2 个样本,如果 $1 \leq i < j \leq 10$,那么 $[Level_i, Level_j]$ 表示高地位等级用户向低地位等级用户建立正向链接;如果 $1 \leq j < i \leq 10$,那么 $[Level_i, Level_j]$ 表示低地位等级用户向高地位等级用户建立正向链接. 对每个 $[Level_i, Level_j]$,统计 $Level_i$ 等级中用户向 $Level_j$ 等级中用户建立正向链接关系的数量. 最后,利用统计学中的核心平滑密度(kernel smoothing density)方法评估信任关系的分布统计,其结果分布图如图 3



(a) Trust relations



(b) Distrust relations

Fig. 3 Density estimate of numbers of trust and distrust relations.

图 3 信任和不信任关系数量的密度评估

(a)所示. 同理,针对不信任关系分布统计,按照相反的方式构建 2 个样本,其结果分布图如图 3(b)所示.

从图 3(a)可以看出,低地位等级信任高地位等级数量同高地位等级信任低地位等级数量相比拥有较大的密度中心,这正好回答了第 1 个问题:低地位等级用户更可能与高地位等级用户建立正向链接关系. 同样,在图 3(b)中,对于不信任关系数量,高地位等级不信任低地位等级的数量的密度中心大于低地位等级不信任高地位等级用户,这也正好回答了第 2 个问题:高地位等级用户更可能与低地位等级用户建立负向链接关系. 这 2 个问题正好提供了关于社会地位理论和链接关系正负属性之间存在一定关联的证据.

5 链接预测模型及优化算法

提出基于低秩矩阵分解方法^[21],以交互意见和地位理论作为模型构建的动机,分别构建相应诱因的链接预测模型实现符号网络中链接关系的预测.

5.1 低秩矩阵链接预测 MF

符号网络 G 以邻接矩阵形式表示, $G_{ij} = 1$ 表示用户 u_i 与用户 u_j 之间建立正向链接关系,而 $G_{ij} = -1$ 表示用户 u_i 与用户 u_j 之间建立负向链接关系. 由于符号网络矩阵 G 具有低秩结构特征^[10]并且较为稀疏,因此利用矩阵分解方法寻求低秩逼近的表示方式. 将符号网络矩阵 G 分解为低秩矩阵,并通过低秩矩阵的相乘重构符号网络矩阵 G ,补齐原矩阵中的缺失值,如式(11)所示:

$$G \approx UHV^T, \quad (11)$$

其中, $G \in \mathbb{R}^{n \times n}$, $U \in \mathbb{R}^{n \times d}$, $H \in \mathbb{R}^{d \times d}$, $V \in \mathbb{R}^{n \times d}$, 且 $d \ll n$. 矩阵 U 和矩阵 V 表示用户在同一领域的偏好矩阵,每一行表示用户,每一列表示用户在该领域内的偏好方向,每一个元素表示用户在某个方向上的偏好权重;矩阵 H 表示用户偏好方向之间的联系程度,每一行表示用户偏好的方向,每一列也表示用户偏好的方向,每一个元素表示偏好方向之间的相关权重. 由于符号网络矩阵 G 中行和列的特征相同,且行和列均表示用户,导致 $V=U$,即:

$$G \approx UHU^T. \quad (12)$$

由于符号网络矩阵 G 较为稀疏,通过添加合理的正则化项方式来避免过度拟合(over-fitting). 因此, MF 模型定义如下:

$$\min_{U, H} \|G - UHU^T\|_F^2 + \alpha(\|U\|_F^2) + \beta(\|H\|_F^2). \quad (13)$$

5.2 基于交互意见的链接预测模型 MF-I

为了模拟交互意见用于链接预测,尝试利用交互意见倾向与链接关系符号属性之间存在的强相关性构建基于交互意见的符号网络链接预测模型 MF-I. 为了强调交互意见在符号网络链接预测过程中的重要性,采取不依赖于已知的符号网络拓扑结构信息,即仅利用带有意见倾向的评级行为数据构建伪符号网络矩阵 $\mathbf{X}$. 因此,本模型假设已知的符号网络并不存在,利用交互数据构建伪符号网络矩阵 $\mathbf{X}$ 替换真实符号网络矩阵 $\mathbf{G}$. 伪符号网络矩阵 $\mathbf{X}$ 中的元素 x_{ij} 定义如下:

$$x_{ij} = \begin{cases} 1, & PI_{ij} - NI_{ij} > 0, \\ -1, & PI_{ij} - NI_{ij} < 0, \\ 0, & \text{没有任何交互}, \end{cases} \quad (14)$$

其中, PI_{ij} 表示用户 u_i 到用户 u_j 的正向交互数量, NI_{ij} 表示用户 u_i 到用户 u_j 的负向交互数量.

伪符号网络矩阵 $\mathbf{X}$ 依据正负向的交互数量而构建,其元素值并不是真实的符号网络数据,因此该矩阵并不可靠. 根据交互意见与链接关系符号属性之间的强相关性,引入权重矩阵 $\mathbf{W}$ 增强伪符号网络矩阵 $\mathbf{X}$ 的可靠度,当矩阵中 W_{ij} 权重越大,说明伪符号网络矩阵 $\mathbf{X}$ 中用户 u_i 与用户 u_j 构建链接的可能性越高. 在矩阵分解过程中,通过矩阵元素 W_{ij} 控制对符号网络矩阵元素 x_{ij} 学习过程的贡献. 权重矩阵 $\mathbf{W}$ 的元素定义如下:

$$W_{ij} = \begin{cases} g(I_{ij}), & u_i \rightarrow u_j \text{ 存在交互数据}, \\ c, & \text{没有任何交互}, \end{cases} \quad (15)$$

其中, I_{ij} 表示用户 u_i 与用户 u_j 产生的正向(或负向)交互数量,定义如下:

$$I_{ij} = \begin{cases} PI_{ij}, & PI_{ij} - NI_{ij} > 0, \\ NI_{ij}, & PI_{ij} - NI_{ij} < 0, \\ 0, & \text{otherwise}. \end{cases} \quad (16)$$

另外,为保证 $W_{ij} \in [0, 1]$, 设置常量 $c = 0.01$, 而函数 $g(x)$ 被经验地定义为 $g(x) = 1 - \frac{1}{1 + \ln(x+1)}$.

在构建伪符号网络矩阵 $\mathbf{X}$ 的基础上,通过引入权重矩阵 $\mathbf{W}$ 用于增强待分解矩阵 $\mathbf{X}$ 的可靠度,进而扩展基于低秩矩阵分解的链接预测模型,因此, MF-I 模型定义如下:

$$\min_{\mathbf{U}, \mathbf{H}} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{U}\mathbf{H}\mathbf{U}^T)\|_F^2 + \alpha(\|\mathbf{U}\|_F^2 + \beta(\|\mathbf{H}\|_F^2)), \quad (17)$$

其中, $\odot$ 表示 Hadamard 乘积, 针对维度相同的任意 2 个矩阵 $\mathbf{X}$ 和 $\mathbf{Y}$, $(\mathbf{X} \odot \mathbf{Y})_{ij} = X_{ij} \times Y_{ij}$.

5.3 基于交互意见和地位理论的链接预测模型 MF-SI

当 2 个用户之间存在等级差异时,依据地位理论, 2 个用户之间应该同时建立正、负链接关系, 即两者之间具有对称关系属性. 但是, 通过观察 Epinions 数据集发现, 仅仅大约 1.7% 的关系之间符合对称关联属性. 因此, 当构建基于地位理论的符号网络预测模型时, 不仅仅考虑 2 个用户之间的等级差异, 还需利用交互意见的显著方向性特征克服地位理论对称性的局限, 进而构建基于交互意见和地位理论的链接预测模型 MF-SI. 针对 Leskovec 等人^[12,22]所提出的基于三元组的地位理论 8 种关系组合, 用户 u_i 和用户 u_j 之间链接关系的符号属性及链接关系权重 $s(i, j)$ 可被形式化建模如下:

$$s(i, j) = \text{sign}(r_i - r_j) \times g(r_i, r_j), \quad (18)$$

其中, $\text{sign}(\cdot)$ 表示符号函数, $g(r_i, r_j)$ 表示 2 个用户之间等级差异权重. 根据经验设置 2 个用户之间的等级差异权重 $g(r_i, r_j)$ 为

$$g(r_i, r_j) = \sqrt{\left| \frac{1}{1 + \ln(r_j + 1)} - \frac{1}{1 + \ln(r_i + 1)} \right|}. \quad (19)$$

地位理论的正则化项^[23]可被定义如下:

$$\min \sum_{i=1}^n \sum_{j=1}^n \text{sign}(r_i - r_j) \times g(r_i, r_j) \|\mathbf{U}(i, :) - \mathbf{U}(j, :)\|_2^2. \quad (20)$$

在基于低秩矩阵分解的链接预测基础上, 融合交互意见和地位理论构建符号网络链接预测模型, 将交互意见建模为矩阵分解方法的增强可靠度因子, 而地位理论建模为矩阵分解方法的正则化项, MF-SI 模型定义如下:

$$\min_{\mathbf{U}, \mathbf{H}} \|\mathbf{W} \odot (\mathbf{G} - \mathbf{U}\mathbf{H}\mathbf{U}^T)\|_F^2 + \alpha(\|\mathbf{U}\|_F^2 + \|\mathbf{H}\|_F^2) + \lambda \min \sum_{i=1}^n \sum_{j=1}^n \text{sign}(r_i - r_j) g(r_i, r_j) \times \|\mathbf{U}(i, :) - \mathbf{U}(j, :)\|_2^2, \quad (21)$$

其中, λ 表示用于调谐地位理论贡献程度系数.

5.4 优化算法

针对交互意见和地位理论分别构建基于交互意见的链接预测模型(MF-I)以及融合交互意见和地位理论的链接预测模型(MF-SI), 其中交互意见是以增强可靠度因子形式来构建模型, 而地位理论是以稀疏学习正则化项形式来构建模型. 针对式(13)的低秩矩阵分解方法建模符号网络链接预测问题, 采用 Hsieh 等人^[10]所提出的方法, 通过 sigmoid 变换函数惩罚分解时的符号不一致问题, 以有效地解决算法优化问题并提高预测质量.

首先将式(21)的正则化项转化为矩阵形式,具体过程如下:

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^n \text{sign}(r_i - r_j) \times g(r_i, r_j) \|U(i, :) - \\ & U(j, :)\|_2^2 = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^d \text{sign}(r_i - r_j) \times \\ & g(r_i, r_j) (U(i, k) - U(j, k))^2 = \\ & 2 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^d \text{sign}(r_i - r_j) \times g(r_i, r_j) U^2(i, k) - \right. \\ & \left. \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^d \text{sign}(r_i - r_j) \times g(r_i, r_j) U(i, k) U(j, k) \right) = \\ & 2 \sum_{k=1}^d U^T(:, k) S \odot (Q - Z) U(:, k) = 2 \text{Tr}(U^T \zeta U), \end{aligned}$$

其中, $\zeta = S \odot (Q - Z)$ 为拉普拉斯矩阵; S 为社会等级差异的符号矩阵; Q 为对角矩阵, 每个元素为 $Q(i, i) = \sum_{j=1}^n Z(j, i)$.

然后, 将式(21)表示为第 k 次迭代的 Lagrangian 函数形式, 形式如下:

$$L_k = \|W \odot (G - UHU^T)\|_F^2 + 2\lambda \text{Tr}(U^T \zeta U) + \alpha \|H\|_F^2 + \beta \|U\|_F^2 + 2\lambda \text{Tr}(U^T \zeta U). \quad (22)$$

通过移除常量, L_k 可被重写为

$$\begin{aligned} L_k = & -2\text{Tr}((W^T \odot W^T \odot G^T) UHU^T) + \\ & \alpha \text{Tr}(H^T H) + \beta \text{Tr}(U^T U) + \text{Tr}((W^T \odot W^T \odot \\ & (UHU^T)) UHU^T) + 2\lambda \text{Tr}(U^T \zeta U). \end{aligned} \quad (23)$$

L_k 关于 U 和 H 的偏导数定义如下:

$$\begin{aligned} \frac{\partial L_k}{\partial U} = & (W^T \odot W^T \odot (UHU^T)) UH + (W \odot W \odot \\ & (UHU^T)) UH^T + (W^T \odot W^T \odot (UHU^T)) UH^T + \\ & (W \odot W \odot (UHU^T)) UH + 2\alpha U + 4\lambda S \odot Q - \\ & 2(W^T \odot W^T \odot G^T) UH - 2(W \odot W \odot G) UH^T - \\ & 4\lambda S \odot ZU. \end{aligned} \quad (24)$$

$$\begin{aligned} \frac{\partial L_k}{\partial H} = & U^T (W \odot W \odot (UHU^T)) U + U^T (W^T \odot W^T \odot \\ & (UHU^T)) + 2\alpha H - 2U^T (W \odot W \odot G) U. \end{aligned} \quad (25)$$

对于 H 和 U 的偏导数, 目标函数(22)的优化解决方法可通过梯度下降法进行求解, 具体如算法 1 所示:

算法 1. MF-SI 模型.

输入: 符号网络 G 、用户-评论关联矩阵 E 、用户-评论的帮助评级矩阵 R ;

输出: 正向链接关系对集合 T' 和负向链接关系对集合 D' 的排序列表.

- ① 利用矩阵 E 和矩阵 R 构建交互意见矩阵;
- ② 依据式(15)计算矩阵 W ;

- ③ 依据符号网络 G 计算用户社会地位 $\{r_i\}_{i=1}^n$;
- ④ 随机初始化 U 和 H ;
- ⑤ while 不收敛 do
- ⑥ 计算 $\frac{\partial L_k}{\partial U}$ 和 $\frac{\partial L_k}{\partial H}$;
- ⑦ 更新 $U \leftarrow U - \delta_U \frac{\partial L_k}{\partial U}$;
- ⑧ 更新 $H \leftarrow H - \delta_H \frac{\partial L_k}{\partial H}$;
- ⑨ end while
- ⑩ 设置 $\hat{G} = UHU^T$;
- ⑪ 设置 $T' = \{\langle u_i, u_j \rangle \mid \text{sign}(\hat{G}_{ij}) = 1\}$;
- ⑫ 设置 $D' = \{\langle u_i, u_j \rangle \mid \text{sign}(\hat{G}_{ij}) = -1\}$;
- ⑬ 按照 $\hat{G}_{ij}$ 中值的降序排序 T' 和 D' 的用户对.

6 实 验

通过实验, 验证 3 个问题:

- 1) 通过对比不同方法, 所提模型是否有效地实现正向链接和负向链接预测?
- 2) 交互意见和地位理论作为可靠度因子和正则化项在所提链接预测模型中的作用如何?
- 3) 通过对比不同量化策略, 哪些量化方法对提高模型预测质量更加有效?

6.1 实验设置和评价策略

实验设置: 1) 将所有的链接关系对表示为集合 A , 并按照链接的创建时间降序排列所有链接关系对, 从 A 中选取时间 T 前的 $x\%$ 的链接关系对作为已知部分的符号网络信息; 2) 对于每个 x , 收集正向链接、负向链接、用户-评论关系信息和用户-评论帮助(helpfulness)评级信息等相关内容作为已知的网络拓扑结构数据和用户交互行为数据; 3) 在实验数据集 A 中的 x 分别取值 10, 30, 50, 70, 100 形成已知数据集 D_x , 其余 $1-x\%$ 作为待评估数据集.

采用通用评价标准^[24]以衡量所提模型的预测效果. 对于每个待评估数据集, 正向链接关系对集合和负向链接关系对集合分别用 T 和 D 表示, 并作为评价标准. 然而, 通过所提模型的预测结果依据模型所得置信度将所有用户形成的链接关系对按照降序排列, 选取前 $|T|$ 链接关系对作为预测的正向链接集合 T' , 后 $|D|$ 链接关系对作为预测的负向链接集合 D' . 最终, 模型的预测质量(prediction quality, PQ)被形式化表示为

$$PQ = \frac{|T \cap T'| + |D \cap D'|}{|T| + |D|}, \quad (26)$$

其中, $|\cdot|$ 表示集合的数量. 依据文献[23], 预测质量 PQ 值通常非常小, 但同随机预测结果相比, 模型预测结果更具有价值和意义. PQ 值小的原因是考虑到所有用户都有可能建立链接关系, 即 $n \times n$ 作为基数, 由此产生的预测结果值就会很小. 为了确保实验结果的可靠性, 所有实验构建 5 次, 取平均值作为实验最终结果.

6.2 链接预测模型对比效果

为了回答第 6 节开始部分的第 1 个问题, 对 4 个基本(baseline)方法进行比较:

1) MF 方法. 该方法是将正负链接关系以矩阵形式表示, 并使用矩阵分解方法获得预测值. 将越高的预测值看作是正向链接, 相反, 越低的预测值看作是负向链接.

2) Inter 方法. 该方法是基于交互意见与符号网络链接属性的强相关性而定义, 依据正向和负向交互数量的排列链接关系对. 正向交互数量越多, 建立正向链接关系的可能性就越大; 负向交互数量越多, 建立负向链接关系的可能性就越大.

3) Status 方法. 该方法是基于地位理论与符号网络链接属性的强相关性而定义, 依据链接关系 2 个用户的等级差异值排列链接关系对. 正向等级差异越大, 建立正向链接关系的可能性就越大; 负向等级差异越大, 建立负向链接关系的可能性就越大.

4) Random 方法. 该方法是依据数据集中正向和负向链接比例随机选取链接关系对. 一般来说, 在符号属性预测中, 随机方法的结果就能达到约 50% 的准确率. 因此, 该方法在文献[24]中被建议作为基本方法, 同随机方法相比, 更能体现预测模型提高的效果和意义.

在本组实验中, 由于 MF-I 模型与 MF-SI 模型所使用的待分解矩阵不同, 因此依据交互的数量降序排列并选取同其他模型具有相同链接关系数量的交互数据构建伪符号网络矩阵. 通过交叉验证方法调谐 MF-I 和 MF-SI 模型参数. 为了达到一般性实验目的, 将所有模型分解后矩阵 H 和 U 的正则化项参数设置为 $\alpha = \beta = 0.1$. MF-SI 模型中, 设置参数 $l = 0.1$. 实验对比结果如图 4 所示.

通过实验结果可以看出:

1) MF-I 模型和 MF-SI 模型在预测质量方面一致优于其他基本方法, 特别是远远高于 MF 方法和随机方法, 说明所提出模型涉及的 2 个链接构建诱

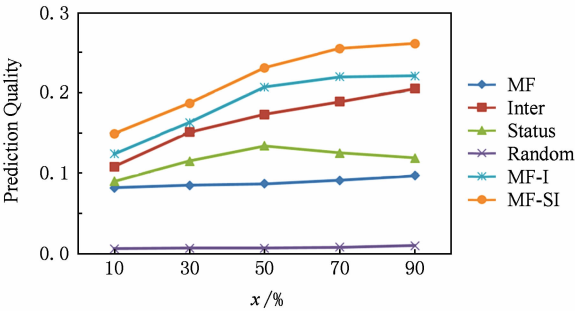


Fig. 4 Performance comparison of different link prediction methods.

图 4 不同链接预测方法对比结果

因对提高模型预测准确度发挥一定贡献. 另外, 同 MF-I 模型相比, MF-SI 模型具有最好的实验效果, 预测质量平均高于 MF-I 模型 0.025. 说明集成交互意见与地位理论的模型要优于其他模型.

2) 从 MF-I 模型和 MF 模型的对比结果可以发现, 将交互意见建模为增强可靠度因子能够提高矩阵分解方法的预测质量. 另外, 基于交互意见与链接符号属性强关联关系的 Inter 方法也具有较好的预测效果, 优于 Status 方法和 MF 方法, 充分说明交互意见在预测链接关系方面的重要作用.

3) MF-I 模型和 Status 方法, 随着 $x\%$ 的增加, 特别是当达到 50% 以上时, 实验效果不会显著提高, 甚至 Status 方法会出现下降, 其主要原因是随着已知链接关系数量的增加, 等级差异不大的用户之间建立链接的准确率会有较为明显的下降.

对所有的实验结果执行 T-test 检验, 其结果说明所提高的实验效果是显著的. 总而言之, 借助地位理论的正则化项所提出的 MF-SI 模型, 同一些基本方法相比, 特别是随机方法, 实验结果都有显著提高; MF-I 模型证明了交互意见在预测符号网络方面具有重要作用; 同样, 集成了交互意见和地位理论的 MF-SI 模型具有更好的效果, 并且能够实现准确地预测符号网络链接关系符号属性. 这也正好回答了第 6 节开始部分所提出的第 1 个问题.

6.3 模型的参数分析

MF-SI 模型有 2 个重要参数控制其实验效果, 分别是: 1) W 用于控制交互数据对待分解矩阵的可靠度; 2) λ 用于控制融入结构平衡理论或社会地位理论对模型的贡献程度. 在本组实验中, 尝试固定一个参数取值同时调谐另一个参数方式, 观测 MF-SI 模型的效果.

在本组实验中,通过设定参数取值去掉部分理论的影响产生模型的多种变体,首先固定参数 λ 取值,对于 MF-SI 模型,经验地设置参数 $\lambda=0.1$. 通过对比 3 种不同 $g(x)$ 取值以衡量 $\mathbf{W}$ 对模型的贡献,分别是:1) $g(x)=1$; 2) $g(x)=random, random \in [0,1]$; 3) $g(x)=1-\frac{1}{1+\ln(x+1)}$. 特别是,当 $g(x)=1$ 时,增强可靠度矩阵 $\mathbf{W}$ 中所有元素都为 1,相当于 MF-SI 模型去掉交互意见的影响,从而产生 MF-S 模型,即仅基于地位理论实现链接预测. 实验对比结果如表 2 所示:

Table 2 Different Measurements of Reliability Matrix W

表 2 可靠度矩阵 W 不同取值方式对比效果

$g(x)$	$x\%$				
	10%	30%	50%	70%	90%
1	0.113	0.151	0.196	0.199	0.204
random	0.088	0.129	0.166	0.172	0.177
$1-1/(1+\ln(x+1))$	0.149	0.187	0.231	0.235	0.239

从表 2 可以看出:

1) 同 $g(x)=1-1/(1+\ln(x+1))$ 效果相比, $g(x)=random$ 效果降低较为明显. 表明函数 $g(x)$ 不能设置为随机数,同时也证明了基于交互数量定义的函数 $g(x)$ 在模型中作用较为明显,能够提高模型的质量.

2) 通过设置 $g(x)=1$ 后,相当于 MF-SI 模型去掉了交互意见的影响,同 $g(x)=1-1/(1+\ln(x+1))$ 效果相比, $g(x)=1$ 的实验效果分别降低了 0.03~0.04.

实验结果进一步证明集成交互意见的重要性,其能够有助于解决社会地位理论自身的局限性,进而提高预测质量. 另外,以合理的方式融合多个社会学理论到稀疏学习模型中是能够提高预测质量的,这回答了第 6 节开始部分的第 2 个问题.

通过调谐参数 λ 控制地位理论正则化项的贡献程度, λ 分别取 $\{0, 0.01, 0.1, 0.5, 1, 5, 10\}$. 实验结果如图 5 所示.

通过参数 λ 不同取值的比较,可以看出:

1) 对于 MF-SI 模型,当参数 $\lambda=0.1$ 时, MF-SI 模型达到了峰值.

2) 参数 $\lambda=0$,相当于 MF-SI 模型去掉了地位理论正则化项部分.

随着参数 λ 从 0 开始增加,实验效果也随之得到了显著的提高. 这样的结果说明地位理论的正则化项能够提高模型的预测质量.

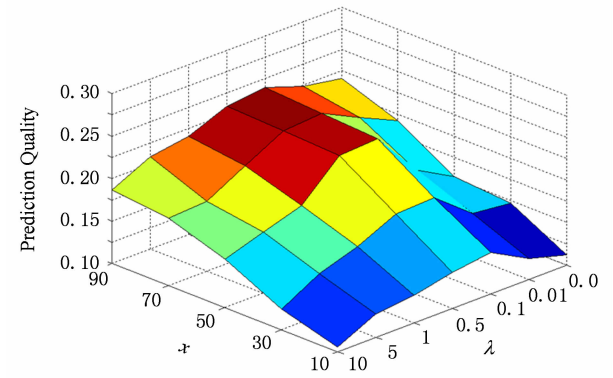


Fig. 5 Performance comparison for different parameter λ .
图 5 不同参数 λ 取值实验效果对比结果

6.4 MF-SI 等级量化策略对链接预测的影响

权重矩阵 $\mathbf{W}$ 的不同量化策略已经在实验 6.3 节中进行有效验证. 在本组实验中,重点衡量 MF-SI 模型中社会等级系数,其用于控制 2 个用户之间等级差异大小. 采用 3 种社会等级策略量化用户的社会等级,分别是 PageRank, PolarityRank 和 PolarityIndegree. 依据量化后的权重排序所有节点,再利用式(19)计算 2 个节点的权重值 $g(r_i, r_j)$. 除了以上 3 种策略之外,与 6.3 节实验相似,设置 $g(r_i, r_j)=1$ 和 $g(r_i, r_j)=random$. 其中, $i, j=1, 2, \dots, n$; $random \in [0,1]$. 表 3 表示采用不同社会等级量化策略的实验对比结果.

Table 3 Different Measurements of Social Statuses

表 3 不同等级量化策略实验对比结果

Quantification Methods	$x\%$				
	10%	30%	50%	70%	90%
$g(r_i, r_j)=1$	0.129	0.142	0.164	0.189	0.205
$g(r_i, r_j)=random$	0.110	0.129	0.148	0.171	0.188
PageRank	0.124	0.162	0.205	0.211	0.213
PolarityRank	0.149	0.187	0.231	0.235	0.239
PolarityIndegree	0.135	0.176	0.219	0.227	0.230

从表 3 可以得出以下结论:

1) PolarityRank 量化策略明显优于其他 2 种量化策略 PageRank 和 PolarityIndegree, 同时, PolarityIndegree 量化策略优于 PageRank 方法, 说明当量化符号网络中节点的等级时, 传统 PageRank 方法没有考虑负向入度链接的作用, 从而影响了节点最终的等级权重.

2) 同 PolarityRank, PolarityIndegree 和 PageRank 量化策略相比, $g(r_i, r_j)=1$ 和 $g(r_i, r_j)=random$ 获得较差的效果, 说明社会等级不能是随机数.

综合以上 2 点,该实验结果有效地回答了第 6 节开始部分的第 3 个问题。

7 结束语

符号网络链接预测问题近几年吸引了越来越多的研究人员的关注,同在已知链接关系基础上预测符号属性问题不同,链接预测既要预测未知链接关系还要预测链接的符号属性.因此,符号网络中链接预测问题较为困难,并且预测质量也不理想.本文通过探索交互意见和地位理论实现符号网络链接关系预测,从而形成新颖模型 MF-I 和 MF-SI.实验结果表明所提模型同其他基本方法相比具有较好的预测质量.

在今后的工作中,我们会继续探索社会学理论以及其他相关理论,以更好地解析符号网络数据;同时,进一步探索符号网络中相关应用研究,比如节点排序、社区发现、信息传播及推荐等.

致谢 在此,衷心感谢亚利桑那州立大学 DMML 实验室刘欢(Huan Liu)教授、感谢 DMML 实验室的所有同行、感谢 Yahoo 实验室的汤继良(Jiliang Tang)博士!

参 考 文 献

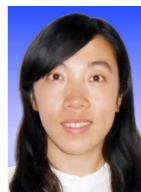
- [1] Cheng Suqi, Shen Huawei, Zhang Guoqing, et al. Survey of signed network research [J]. Journal of Software, 2014, 25(1): 1-15 (in Chinese)
(程苏琦, 沈华伟, 张国清, 等. 符号网络研究综述[J]. 软件学报, 2014, 25(1): 1-15)
- [2] Heider F. Attitudes and cognitive organization [J]. Journal of Psychology, 1946, 21(1): 107-112
- [3] Cartwright D, Harary F. Structural balance: A generalization of Heider's theory [J]. Psychological Review, 1956, 63(5): 277-293
- [4] Borzymek P, Sydow M. Trust and Distrust Prediction in Social Network with Combined Graphical and Review-Based Attributes [G] //Agent and Multi-Agent Systems: Technologies and Applications. Berlin: Springer, 2010: 122-131
- [5] Ye J, Cheng H, Zhu Z, et al. Predicting positive and negative links in signed social networks by transfer learning [C] //Proc of the 22nd Int Conf on World Wide Web. New York: ACM, 2013: 1477-1488
- [6] Zolfaghar K, Aghaie A. Mining trust and distrust relationships in social Web applications [C] //Proc of the 6th Int Conf on Intelligent Computer Communication and Processing. Piscataway, NJ: IEEE, 2010: 73-80
- [7] Javari A, Mahdi J. Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links [J]. ACM Trans on Intelligent Systems and Technology, 2014, 5(2): 1-24
- [8] Symeonidis P, Eleftherios T. Transitive node similarity: Predicting and recommending links in signed social networks [J]. World Wide Web, 2014, 17(4): 743-776
- [9] Lv L Y, Zhou T. Link prediction in complex networks: A survey [J]. Physica A: Statistical Mechanics and Its Applications, 2011, 390(6): 1150-1170
- [10] Hsieh C J, Chiang K Y, Inderjit S. Low rank modeling of signed networks [C] //Proc of the 18th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2012: 507-515
- [11] Yi C, Gu R T, Ji Y F. Sign inference for dynamic signed networks via dictionary learning [J]. Journal of Applied Mathematics, 2013, 2013(3): 701-710
- [12] Leskovec J, Huttenlocher D, Kleinberg J. Predicting positive and negative links in online social networks [C] //Proc of the 19th Int Conf on World Wide Web. New York: ACM, 2010: 641-650
- [13] Zheng X, Zeng D, Wang F. Social balance in signed networks [J]. Information Systems Frontiers, 2015, 17(5): 1077-1095
- [14] Chiang K Y, Natarajan N, Tewari A, et al. Exploiting longer cycles for link prediction in signed networks [C] //Proc of the 20th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2011: 1157-1162
- [15] Patidar A, Agarwal V, Bharadwaj K K. Predicting friends and foes in signed networks using inductive inference and social balance theory [C] //Proc of the 2012 Int Conf on Advances in Social Networks Analysis and Mining. Piscataway, NJ: IEEE, 2012: 384-388
- [16] Priyanka A, Garg V K, Narayanam R. Link label prediction in signed social networks [C] //Proc of the 23rd Int Joint Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2013: 2591-2597
- [17] Ramanathan G, Kumar R, Raghavan P, et al. Propagation of trust and distrust [C] //Proc of the 13th Int Conf on World Wide Web. New York: ACM, 2004: 403-412
- [18] Lan Mengwei, Li Cuiping, Wang Shaoqing, et al. Survey of sign prediction algorithms in signed social networks [J]. Journal of Computer Research and Development, 2015, 52(2): 410-422 (in Chinese)
(蓝梦微, 李翠平, 王绍卿, 等. 符号社会网络中正负关系预测算法研究综述[J]. 计算机研究与发展, 2015, 52(2): 410-422)

- [19] Wang Y, Wang X, Zuo W L. Research on trust prediction from a sociological perspective [J]. Journal of Computer Science and Technology, 2015, 30(4): 843-858
- [20] Wang Ying, Wang Xin, Zuo Wanli. Trust prediction modeling based on social theories [J]. Journal of Software, 2014, 25(12): 2893-2904 (in Chinese)
(王英, 王鑫, 左万利. 基于社会学理论的信任关系预测模型研究[J]. 软件学报, 2014, 25(12): 2893-2904)
- [21] Liu Ye, Zhu Weiheng, Pan Yan, et al. Multiple sources fusion for link prediction via low-rank and sparse matrix decomposition [J]. Journal of Computer Research and Development, 2015, 52(2): 423-436 (in Chinese)
(刘冶, 朱蔚恒, 潘炎, 等. 基于低秩和稀疏矩阵分解的多源融合链接预测算法[J]. 计算机研究与发展, 2015, 52(2): 423-436)
- [22] Leskovec J, Huttenlocher D, Kleinberg J. Signed networks in social media [C] //Proc of the SIGCHI Conf on Human Factors in Computing Systems. New York: ACM, 2010: 1361-1370
- [23] Wang Y, Wang X, Tang J, et al. Modeling status theory in trust prediction [C] //Proc of the 29th AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2015: 1875-1881
- [24] Nowell D L, 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



Wang Xin, born in 1981. PhD, lecturer. Member of China Computer Federation. His main research interests include data mining, machine learning, social network and recommendation system.



Wang Ying, born in 1981. PhD, associate professor. Senior member of China Computer Federation. Her main research interests include data mining, machine learning, social computing and search engine.



Zuo Wanli, born in 1957. Professor and PhD supervisor in the College of Computer Science and Technology, Jilin University. Senior member of China Computer Federation. His main research interests include information retrieval, Web mining and machine learning.