软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2014,25(12):2893−2904 [doi: 10.13328/j.cnki.jos.004731]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于社会学理论的信任关系预测模型*

王 英[1,2,3], 王 鑫[1,2,4], 左万利[1,2]

[1](吉林大学 计算机科学与技术学院,吉林 长春 130012)
[2](符号计算与知识工程教育部重点实验室(吉林大学),吉林 长春 130012)
[3](吉林大学 数学学院,吉林 长春 130012)
[4](长春工程学院 计算机技术与工程学院,吉林 长春 130012)
通讯作者: 左万利, E-mail: zuowl@jlu.edu.cn

摘 要: 随着社会网络的盛行,信任作为用户之间交互的基础,在信息共享、经验交流和社会舆论方面发挥着重要作用.然而,信任是一个复杂而抽象的概念,受多种因素影响,很难识别信任形成的诱因以及其形成机制.由于来自社会科学的社会学理论有助于解释社会现象,而社会网络反映了现实世界中用户之间的联系,因此,从社会学角度出发,通过研究社会等级理论和同质性理论获取信任关系的发展规律,进而构建信任关系预测模型.首先,对社会等级理论和同质性理论进行阐述,并验证了社会等级理论和同质性理论在社会网络中的存在;然后,分别针对社会等级理论和同质性理论对信任关系的影响提出社会等级正则化方法和同质性正则化方法;最后,利用非负矩阵的三维分解方法并结合社会等级理论和同质性理论实现对信任关系预测的建模,并提出 SocialTrust 模型用于信任关系预测.实验结果表明,相比于其他方法,该方法在信任关系预测方面具有较高的精度.

关键词: 信任关系预测;非负矩阵的三维分解;社会等级理论;同质性理论;SocialTrust
中图法分类号: TP311

中文引用格式: 王英,王鑫,左万利.基于社会学理论的信任关系预测模型.软件学报,2014,25(12):2893−2904. http://www.jos.org.cn/1000-9825/4731.htm
英文引用格式: Wang Y, Wang X, Zuo WL. Trust prediction modeling based on social theories. Ruan Jian Xue Bao/ Journal of Software, 2014,25(12):2893−2904 (in Chinese). http://www.jos.org.cn/1000-9825/4731.htm

Trust Prediction Modeling Based on Social Theories

WANG Ying[1,2,3], WANG Xin[1,2,4], ZUO Wan-Li[1,2]

[1](College of Computer Science and Technology, Jilin University, Changchun 130012, China)
[2](Key Laboratory of Symbolic Computation and Knowledge Engineering (Jilin University), Ministry of Education, Changchun 130012, China)
[3](College of Mathematics, Jilin University, Changchun 130012, China)
[4](College of Computer Technology and Engineering, Changchun Institute of Technology, Changchun 130012, China)
Corresponding author: ZUO Wan-Li, E-mail: zuowl@jlu.edu.cn

Abstract: With the pervasion of social media, trust, as the basis of human interactions, has been playing an important role in addressing information sharing, experience communication, and public opinions. However, trust is a complex and abstract concept influenced by many factors, and it is difficult to identify the inducing factors and analyze the formation mechanism. Recognizing that social theories from social sciences are helpful to explain social phenomena, and social networks reflect user correlations in real world, this paper investigatethes trust prediction problem from the perspective of social science, and constructs trust prediction model by studying the

* 基金项目: 国家自然科学基金(61300148); 吉林省科技发展计划(20130206051GX); 吉林省科技计划(20130522112JH); 中国博士后基金(2012M510879); 吉林大学基本科研业务费科学前沿与交叉项目(201103129)

收稿时间: 2014-05-06; 修改时间: 2014-08-21; 定稿时间: 2014-09-30

disciplines of trust relations occurring and developing based on social status theory and homophily theory. Firstly, it gives a brief introduction to social status theory and homophily theory, and verifies the existence of social status theory and homophily theory in social networks. Then, it proposes social status regularization and homophily regularization according to the effects of social status theory and homophily theory in predicting trust relations. Lastly, it models trust prediction by incorporating non-negative matrix tri-factorization, social status theory and homophily theory, and establishes trust prediction model SocialTrust. Experimental results demonstratethe effectiveness of the proposed method in trust prediction with a higher accuracy than other baseline methods.

Key words: trust prediction; non-negative matrix tri-factorization; social status theory; homophily theory; SocialTrust

信任反映了一个用户对另一个用户行为及能力的综合评价,通常可以看作一种决策支持工具,用于解析相关可靠的信息源,尤其是从可信源中寻求建议从而做出决策[1,2].由于不同用户之间的信任关系不同,相同用户在不同领域内的信任关系强度也存在差异,使得信任关系的不确定性成为信任预测的最大挑战.在社会网络中,面对如此多的用户交互以及用户产生内容,建立信任模型的需求逐渐成为一个越来越重要的研究课题.当前,一些学者开展了信任关系挖掘的研究来反映信任关系的复杂性、动态性和不确定性[3−6].毋庸置疑,这些研究成果有效地推动了信任模型研究的发展,极大地丰富了人们对信任关系内涵的进一步认识.但通过深入分析,现有方法中大多没有考虑社会学理论对信任关系的影响.来自社会科学的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,将社会学理论的原理、原则、观点及方法运用于对社会网络中用户之间信任关系的研究,成为信任关系预测研究的新视角.本文基于社会学理论对信任关系的预测进行建模,提出 SocialTrust 模型用于信任关系预测.主要贡献体现在:

(1) 验证了社会学理论在社会网络中的存在,例如:社会等级理论认为,低等级用户更易与高等级用户建立信任关系;同质性理论认为,相似的用户更易建立信任关系;

(2) 为了克服信任网络数据的稀疏性特点,提出新颖的信任关系预测模型 SocialTrust.该模型通过非负矩阵的三维分解方法,并结合社会学理论,实现对信任关系预测的建模,有效地解决了数据稀疏性问题;

(3) 针对用户所属的不同领域对信任关系预测模型 SocialTrust 进行验证,实验结果表明,该方法相比于其他传统方法具有较高精度.

本文第 1 节为相关性研究.第 2 节为社会学理论介绍.第 3 节为基于社会学理论的信任模型构建.第 4 节为实验及结果分析.第 5 节为结论和展望.

1 相关性研究

社会网络的日益流行,吸引了国内外众多学者对信任关系展开相关研究,用户间的信任关系主要依靠建立信任模型来解决.

当前,对信任模型的研究主要可以分为两类:基于统计分析方法和基于机器学习方法.

基于统计分析方法通过集成多个与信任相关的维度,如历史信息、情境信息、声誉信息等实现高精度的信任关系预测.UgurKuter 等人[7]提出基于概率网络推理信任关系,并提出信任推理算法 SUNNY 计算社会网络中的信任值.唐杰等人[8]通过学习异质网络提出社会关系推测框架,该框架将社会理论合成为一个因子图模型,有效地改善了在目标网络中推理社会关系种类的准确性.AchimRettinger 等人[9]依据过去的观测和上下文信息解决信任推理问题,并提出信任模型 IHRTM,该模型利用统计关系学习获取情境敏感信息,情境敏感信息包括受信者的个体特征和情境状态,但该模型由于缺乏自适应学习策略使其在实际应用中存在一定的局限性.Khiabani 等人[10]提出通过集成历史、推荐、情境等信息构建信任模型 UTM 用于计算个体间的信任分值,实验表明,UTM 模型考虑了行为因素,能够高效地用于低交互环境中.Lee 等人[11]针对电子商务网站提出信任模型,该模型主要考虑了 4 个因素:店主的信任程度、所在网站的信任程度、基础设施因素(安全、第三方验证)以及公司规模因素.Nepal 等人[12]提出信任模型 STrust,该模型考虑了两个方面的信任:PopTrust 和 EngTrust,其中,PopTrust 反映了个体被社区中其他个体的信任程度,EngTrust 反映了个体信任社区中其他个体的程度.

基于机器学习的方法又可以分为两类:监督信任关系预测方法和非监督信任关系预测方法.监督信任关系预测方法首先从源数据中抽取特征,然后基于这些特征训练二值分类器.Liu 等人[13]提出使用分类方法从用户属性和交互属性中获取相关特征集合以实现信任关系的预测,但评估结果具有绝对性,即信与不信,忽略了信任的不确定性.Zolfaghar 等人[14]提出信任诱因的形成框架,并使用数据挖掘和分类方法研究信任形成过程,实验结果表明,决策树和神经网络方法优于其他方法.非监督信任关系预测方法大多基于信任演化,即依赖于用户间已存在的信任关系,但当已存在的信任关系非常稀疏时,信任演化可能会失败.Borzymek 等人[15]提出评级相似度以丰富传统的信任演化方法,实验结果表明:如果结合含有评级相似度的信任网络,那么越相似的用户,信任关系预测越准确.Xiang 等人[16]提出一种非监督模型,从交互活动(交流、加标签等)和用户之间的相似度两个方面自动地区分强关系与弱关系,但忽略了用户的领域信息,不同领域的相同用户之间的关系强度可能有显著的区别.Liu 等人[17]基于 Monte Carlo 方法和优化策略提出一种情境感知的信任网络发现算法 SCAN,实验结果表明,该方法在执行时间和信任网络质量方面优于其他算法.Jin 等人[18]认为,用户通常根据少许因素建立朋友关系,因此信任矩阵是低秩的,进而提出基于低秩矩阵分解的信任关系预测模型 RRMC,并通过实验证明该方法相比于传统方法具有稳定的性能.Tang 等人[19]通过探索同质效应研究信任关系预测问题,并利用低秩矩阵分解技术构建信任关系预测模型 hTrust,实验证明,该方法具有较好的性能.此外,Tang 等人[20]还提出多维度信任的概念,并将多维度信任应用到社会推荐中.Fang 等人[21]提出将单方面的信任信息按 benevolence,integrity, competence 和 predictability 分解为 4 个方面,并集成这 4 个方面的信任到概率矩阵分解模型中,从而实现高效的预测.

综上所述,虽然对信任模型的研究已有很多,但是大部分工作都没有考虑信任的所属领域.信任具有领域相关性,人们对不同的人在不同的领域会有不同的信任.如果粗略地认为信任具有唯一性,那么会引起信任关系过度泛化,使用这些泛化的信任关系进行信任关系预测,将会导致不精确的信任结果.此外,大多工作也没有考虑社会学理论对信任关系预测的影响.社会学理论有助于解释社会现象,将社会学理论引入信任关系预测的研究,将会在很大程度上提高信任关系预测的精度.

2 社会学理论

社会学理论有助于解释社会现象,其对理解社会网络起到关键作用.社会等级理论和同质性理论是比较重要的社会学理论,因此,结合社会等级理论和同质性理论实现信任关系的预测.

2.1 社会等级理论

社会等级理论主要用于有向网络,其反映了一个用户在网络中的地位,也反映了网络中个体对其他个体的依附程度[22].在社会等级理论中,通常:使用一个正的有向链接表示链接的创建者认为链接的接收者具有较高的等级,标记为“+”;而使用一个负向的有向链接表示链接的创建者认为链接的接收者具有较低的等级,标记为“–”.用户的社会等级可以通过有向网络的多步路径进行演化.

用户的社会等级在社会网络中是非常重要的概念,体现了一个用户在网络中的重要程度.在信任关系预测中,社会等级理论用于解释用户的社会等级如何影响用户之间信任关系的建立,即低等级用户更易与比其高等级的用户建立信任关系.例如在社会网络中,高等级用户通常属于权威用户,低等级用户更易与高等级用户建立信任关系,而且低等级用户与高等级用户建立的信任关系数量要远多于与低等级用户建立的信任关系数量.

2.2 同质性理论

同质性是个体关联到相似个体的趋势,用于解释现实社会中具有相似特征的用户之间更容易建立联系,即用户倾向于选择在某些方面与自己相似的用户交互.同质性因素主要由个体特性的相似以及所处社会环境的相似两类构成.个体特征的相似包括种族、性别、年龄、宗教、信仰、教育等,社会环境特征包括职位、社会地位、网络位置、行为、能力、愿望等.

在社会网络研究中,同质性假定了结点间属性的相似性,导致了结点之间联系的产生.同质性是试图解释人们彼此建立信任关系的最重要的理论依据,同质效应表明,相似用户具有高的可能性去建立信任关系.例如在商

品评价网站,对于商品具有相似品味的用户更可能彼此信任对方.同质性是最重要的社会关系理论之一,在很多社会网络中都有体现,例如 Twitter,Epinions 等.信任关系中的同质性可以通过研究信任关系和用户相似性之间的相互联系确定.

3 信任关系预测模型 SocialTrust

非负矩阵分解(non-negative matrix factorization,简称 NMF)[23]是近年发展起来的一种矩阵分解方法,其充分考虑了各种因素的影响,具有非常好的扩展性,被广泛应用于人脸识别、数据挖掘、生物信息学、遥感图像处理等方面.在不同的实际应用中,由于被分解矩阵过于稀疏,通常会根据先验知识对分解矩阵添加一些正则化(regularization)项,即约束条件,使求解结果向实际问题所需要的解靠近.正则化是为了解决过度拟合(overfitting)问题而引入的一种方法,一般来说有两种克服过度拟合的方法:一是补偿模型的某些部分(如 regularization);二是根据潜在的问题提供额外的数据来训练.本文采用第 1 种方法,通过添加社会等级正则化项和同质性正则化项作为信任关系预测的约束条件,然后使用迭代进行矩阵分解,以实现更准确的信任关系预测,从而消除重构预测关系中数值的不稳定性.

3.1 社会等级正则化

用户的社会等级体现了一个用户在网络中的重要程度,针对不同领域的信任网络,分别通过 Pagerank 算法求取各个领域中不同用户的社会等级.假设两个用户 u_i 和 u_j,u_i 的社会等级为 r_i,$1\leqslant r_i\leqslant n$,其向 u_j 建立信任关系的可能性为 $TrustPossible_{ij}$;u_j 的社会等级为 r_i,其向 u_i 建立信任关系的可能性为 $TrustPossible_{ji}$.假设 $r_i>r_j$,且 $i<j$,即 u_i 相比于 u_j 具有较高的社会等级,那么存在以下 3 种情况:

情况 1:$TrustPossible_{ij}>TrustPossible_{ji}$;

情况 2:$TrustPossible_{ij}<TrustPossible_{ji}$;

情况 3:$TrustPossible_{ij}=TrustPossible_{ji}$.

可以看出,情况 2 满足社会等级理论,即,同一领域内低等级用户 u_j 向高等级用户 u_i 建立信任关系的可能性 $TrustPossible_{ji}$ 高于高等级用户 u_i 向低等级用户 u_j 建立信任关系的可能性 $TrustPossible_{ij}$.如果信任关系满足社会等级理论,则必有公式(1)成立:

$$(r_i-r_j)(TrustPossible_{ij}-TrustPossible_{ji})<0 \tag{1}$$

那么,可以根据社会等级理论构造社会等级正则化项,其可以看作解决如下优化问题:

$$\min\sum_{i}^{n}\sum_{j=i+1}^{n}(\max\{0,(r_i-r_j)(TrustPossible_{ij}-TrustPossible_{ji})\}) \tag{2}$$

由于规定 $i<j$ 且 $1\leqslant r_j<r_i\leqslant n$,那么 $1\leqslant r_i-r_j\leqslant n-1$,$r_i-r_j$ 的值变化范围较大,为了使其变化在一个合理的范围,定义函数 $f(r_i-r_j)$ 替代 r_i-r_j,并且始终与 r_i-r_j 同号.通常,$0\leqslant f(r_i-r_j)\leqslant 1$,根据经验,定义 $f(r_i-r_j)$ 函数如下:

$$f(r_i-r_j)=\sqrt{\frac{1}{1+\log(r_j)}-\frac{1}{1+\log(r_i)}} \tag{3}$$

那么,社会等级正则化项(公式(2))可以重写为

$$\min\sum_{i}^{n}\sum_{j=i+1}^{n}(\max\{0,f(r_i-r_j)(TrustPossible_{ij}-TrustPossible_{ji})\}) \tag{4}$$

其含义为:如果 $f(r_i-r_j)(TrustPossible_{ij}-TrustPossible_{ji})<0$,这种情况满足社会等级理论,用户 u_i 和用户 u_j 之间的社会等级正则化值应为 0,不需要添加任何惩罚;相反,如果 $f(r_i-r_j)(TrustPossible_{ij}-TrustPossible_{ji})>0$,这种情况不满足社会等级理论,需要增加惩罚项,即用户 u_i 和用户 u_j 之间的社会等级正则化值应为

$$f(r_i-r_j)(TrustPossible_{ij}-TrustPossible_{ji}).$$

最后,对所有用户对间的社会等级正则化值求和并使其值最小.这样,$TrustPossible_{ij}$ 的值就会更逼近 $TrustPossible_{ji}$,以尽可能地避免违背社会等级理论的情况发生,使得高等级用户向低等级用户建立信任关系的

可能性 $TrustPossible_{ij}$ 更趋近于低等级用户向高等级用户建立信任关系的可能性 $TrustPossible_{ji}$.

3.2 同质性正则化

同质性理论认为,高相似的用户对相比于低相似的用户对更易建立信任关系.定义$\xi(i,j)$表示两个用户 u_i 和 u_j 在同一领域内的同质系数,其可以看作为用户之间的相似程度,满足 3 个条件:

条件 1:$0\leqslant\xi(i,j)\leqslant1$;

条件 2:$\xi(i,j)=\xi(j,i)$;

条件 3:同质系数越高,用户间的相似度越大,建立信任关系的可能性越高.

由于高相似度用户更易建立信任关系,那么针对商品评价网站,在同一领域内具有相近偏好的用户更易建立信任关系,定义同质性正则化项如下:

$$\min\sum_{i=1}^{n}\sum_{j\neq i}^{n}\zeta(i,j)\,\|Preference_i-Preference_j\|_F^2 \tag{5}$$

其含义为:$\|Preference_i-Preference_j\|_F^2$ 表示用户在同一领域内的偏好差异,同质系数$\xi(i,j)$用于控制用户之间的偏好差异.同质性理论认为,相似用户更易建立信任关系.为了满足同质性理论,用户 u_i 和用户 u_j 在同一领域偏好 $Preference_i$ 与 $Preference_j$ 越接近,那么 u_i 和 u_j 更易建立信任关系.通过对所有用户对的偏好求和,再取最小,以增加相似用户对之间建立信任关系的可能性.在商品评价网站中,同质系数$\xi(i,j)$可以通过用户在同一领域的商品评级以及用户在该领域的信任网络获取.通常,用户 u_i 对商品 $item_i$ 的评级 $Rate_{ij}$ 在 0 分~10 分之间,其中,$Rate_{ij}=0$ 表示用户 u_i 没有对商品 $item_i$ 进行评级,$Rate_{ij}=1$ 表示用户 u_i 对商品 $item_i$ 具有较低的评级,$Rate_{ij}=10$ 表示用户 u_i 对商品 $item_i$ 具有最高的评级.此外,两个用户共同信任的用户数量越多,那么用户相似的可能性越大.假设信任网络中用户 u_i 信任的用户集合为 N_i,用户 u_j 信任的用户集合为 N_j,$|N_i|$表示用户 u_i 信任的用户数量,$|N_i\cap N_j|$表示用户 u_i 与用户 u_j 共同信任的用户数量.定义同质系数计算如下:

$$\zeta(i,j)=\alpha\frac{\sum_{k=1}^{m}Rate_{ik}\times Rate_{jk}}{\sqrt{\sum_{k=1}^{m}Rate_{ik}\times\sum_{k=1}^{m}Rate_{jk}}}+(1-\alpha)\frac{|N_i\cap N_j|}{|N_i\cup N_j|} \tag{6}$$

其中,$0\leqslant\alpha\leqslant1$,$\alpha$控制通过余弦定理获得的用户对间的评级相似度,$1-\alpha$控制两个用户共同信任的用户比重.

3.3 基于社会学理论的信任关系预测模型

分别针对不同领域构建信任网络,定义 G 表示信任网络,其以邻接矩阵形式表示.由于信任网络矩阵 G 非常稀疏并且低秩,非负矩阵的三维分解方法寻求一种低秩逼近的表示方式.本文将信任网络矩阵 G 分解为 3 个低秩矩阵,通过 3 个低秩矩阵的相乘可以重构信任网络矩阵 G,见公式(7):

$$G\approx UHV^T \tag{7}$$

其中,$G\in R_+^{n\times n}$,$U\in R_+^{n\times d}$,$H\in R_+^{d\times d}$,$V\in R_+^{n\times d}$ 且 $d<<n$.由于信任网络矩阵 G 中的行表示用户,列也表示用户,行和列的特征相同,导致 $V=U$,即

$$G\approx UHU^T \tag{8}$$

其中,矩阵 U 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.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.U^T 是矩阵 U 的转置矩阵.新预测的信任网络矩阵 G 可以通过 3 个矩阵的乘积求取,乘积后得到的大矩阵中的每一个元素 $G(i,j)=U(i,:)HU^T(j,:)$表示用户 u_i 与用户 u_j 建立信任关系的可能性.

非负矩阵的三维分解问题一般并不用数学上直接分解的办法,尽管分解出来的精度很高,但效率过低.通常将其转换为一个最优化问题,通过乘性迭代求取局部最优解,本文中即求取低秩矩阵 U 和 H.然而,由于信任关系矩阵 G 过于稀疏,分解很容易产生过度拟合问题,通常添加一些正则化项来避免过度拟合.其可以看作如下优化

问题:

$$\min_{U,H} \| G - UHU^T \|_F^2 + \alpha(\| U \|_F^2) + \beta(\| H \|_F^2) \\ \text{s.t. } U \geqslant 0, H \geqslant 0 \tag{9}$$

其中,$\|\cdot\|_F^2$ 表示矩阵的范数,α,β为正则项系数,$0\leqslant\alpha\leqslant1$,$0\leqslant\beta\leqslant1$,分别对矩阵 U 和 H 添加正则项以避免过度拟合.已有研究已经证明:上述模型能够用于信任关系预测,并且具有较好的性能[24],但是这些方法并没有考虑社会学理论对信任关系预测的影响.本文综合考虑了社会学理论中的社会等级理论和同质性理论,通过添加社会等级正则化项和同质性正则化项作为约束条件,构建新颖的模型 SocialTrust 实现更精确的信任关系预测.SocialTrust 模型见公式(11),其可以看作是解决如下优化问题:

$$\begin{aligned} &\min_{U,H} \| G - UHU^T \|_F^2 + \alpha(\| U \|_F^2) + \beta(\| H \|_F^2) + \\ &\lambda_1 \min \sum_i^n \sum_{j=i+1}^n (\max\{0, f(r_i - r_j)(TrustPossible_{ij} - TrustPossible_{ji})\}) + \\ &\lambda_2 \min \sum_{i=1}^n \sum_{j\neq i}^n \zeta(i,j) \| Preference_i - Preference_j \|_F^2 \\ &\text{s.t. } U \geqslant 0, H \geqslant 0 \end{aligned} \tag{10}$$

其中,λ_1 和 λ_2 为非负参数,λ_1 用于控制社会等级理论的贡献程度,λ_2 用于控制同质性理论的贡献程度.

$U(i,:)HU^T(j,:)$表示用户 u_i 与用户 u_j 建立信任关系的可能性,$U(j,:)HU^T(i,:)$表示用户 u_j 与用户 u_i 建立信任关系的可能性,那么 $TrustPossible_{ij}=U(i,:)HU^T(j,:)$,$TrustPossible_{ji}=U(j,:)HU^T(i,:)$.此外,矩阵 U 表示用户-偏好矩阵,$U(i,:)$表示用户 u_i 的偏好,$U(j,:)$表示用户 u_j 的偏好,那么 $Preference_i=U(i,:)$,$Preference_j=U(j,:)$.因此,公式(10)可以重写为

$$\begin{aligned} &\min_{U,H} (\| G - UHU^T \|_F^2 + \alpha(\| U \|_F^2) + \beta(\| H \|_F^2)) + \\ &\lambda_1 \min \sum_i^n \sum_{j=i+1}^n (\max\{0, f(r_i - r_j)(U_i HU_j^T - U_j HU_i^T)\}) + \\ &\lambda_2 \min \sum_{i=1}^n \sum_{j\neq i}^n \zeta(i,j) \| U_i - U_j \|_F^2 \\ &\text{s.t. } U \geqslant 0, H \geqslant 0 \end{aligned} \tag{11}$$

定义矩阵 R,使得:

$$R_{ij} = \begin{cases} f(r_i - r_j), & \text{if } (U_i HU_j^T - U_j HU_i^T) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

那么可以得出:

$$\sum_i^n \sum_{j=i+1}^n (\max\{0, f(r_i - r_j)(U_i HU_j^T - U_j HU_i^T)\}) = \| R \odot (UH^T U^T - UHU^T) \|_F^2 \tag{13}$$

同样,定义矩阵 Z 表示同质系数矩阵,那么可以得出:

$$\sum_{i=1}^n \sum_{j\neq i}^n \zeta(i,j) \| U_i - U_j \|_F^2 = 2Tr(U^T(Q - Z)U) \tag{14}$$

其中,Q 是一个对角矩阵,第 i 个对角元素为同质系数矩阵 Z 中第 i 列值之和.

在数学最优化问题中,拉格朗日乘数法是一种寻找变量受一个或多个条件所限制的多元函数的极值的方法.这种方法将有 n 个变量与 k 个约束条件的最优化问题转换为一个有 $n+k$ 个变量的方程组的极值问题,其变量不受任何约束.这种方法引入了新的标量未知数,即拉格朗日乘数——约束方程的梯度(gradient)的线性组合里每个向量的系数.由于公式(11)所求为最小值,那么其拉格朗日公式为

$$\begin{aligned} F(U,H) = &\| G - UHU^T \|_F^2 + \alpha(\| U \|_F^2) + \beta(\| H \|_F^2) + \lambda_1 \| R \odot (UH^T U^T - UHU^T) \|_F^2 + \\ &2\lambda_2 Tr(U^T(D - Z)U) - Tr(\wedge_1 U) - Tr(\wedge_2 H) \end{aligned} \tag{15}$$

其中,$\wedge_1$ 和 $\wedge_2$ 为拉格朗日乘子.然后,分别对 U 和 H 求偏导,并根据 KKT 约束条件[25,26],应用梯度下降法可得:

$$U_{ik} \leftarrow U_{ik}\sqrt{\frac{A_{ik}}{B_{ik}}}, H_{ik} \leftarrow H_{ik}\sqrt{\frac{C_{ik}}{D_{ik}}}.$$

其中,

$$\begin{aligned}
A &= G^TUH + GUH^T + \\
&\quad \lambda_1(R^T \odot R^T \odot UHU^TUH^T + R \odot R \odot UH^TU^TUH + R \odot R \odot UHU^TUH^T + R^T \odot R^T \odot UH^TU^T) + 2\lambda_2 ZU, \\
B &= UHU^TUH^T + UH^TU^TUH + 2\lambda_1(R^T \odot R^T \odot UHU^TUH + R \odot R \odot UH^TU^TUH^T) + \\
&\quad 2\lambda_2 QU + \alpha U, \\
C &= U^TGU + \lambda_1 U^T(R^T \odot R^T \odot UHU^T + R \odot R \odot UHU^T)U, \\
D &= U^TUHU^TU + 2\lambda_1 U^T(R \odot R \odot UH^TU^T)U + \beta H.
\end{aligned}$$

在获取 U 和 H 后,用户 u_i 和用户 u_j 建立信任关系的可能性为 $U_iHU_j^T$,即原信任网络矩阵 G 中的缺省值可以通过分解后得到的矩阵乘积获取,从而实现信任关系的预测.

4 实 验

4.1 数据集

Epinions 是著名的商品评价网站,其最重要的特征是用户之间存在信任网络,这样的网站提供了较好的平台研究信任关系预测.本文采用 Epinions 作为实验的数据集,获取用户评级(rating)和信任网络(trust network).对于用户评级,包含商品名称、商品分类、评级分数、评级产生时间以及评级的帮助程度(the helpfulness of this rating);对于信任网络,包括信任关系建立的施信者(trustor)与信任关系指向的受信者(trustee)以及信任关系建立的时间.

由于信任是领域相关的,用户在不同领域对相同用户有着不同的信任强度.本文针对 Epinions 中 Movies 领域、Kids & Family 领域和 Books 领域分别进行信任关系预测,各领域的数据统计见表 1.

Table 1 Data statistics of different fields

表 1 各领域的数据统计

	Movies (3)	Kids & Family (19)	Books (4)
用户数量	3 055	2 696	3 254
商品数量	16 024	13 379	34 412
用户评级数量	26 552	36 736	50 230
信任关系数量	40 039	35 195	47 855
最大施信者数量	235	337	555
最大受信者数量	381	214	434

4.2 评估方法

信任模型是指建立量化的评价体系,以信任值度量节点的可信任程度.在实验评估过程中,使用非监督预测中的常用评估方法[27].首先,将所有用户对分为 A 和 B 两个部分:$A=\{\langle u_i,u_j\rangle|G(i,j)=1\}$表示已存在信任关系的用户对集合,其按照建立信任关系的时间先后顺序进行排序;$B=\{\langle u_i,u_j\rangle|G(i,j)=0\}$表示不存在信任关系的用户对集合.然后,从 A 中选择前 x%的用户对作为已知信任关系的用户对,标记为 C,剩下的 1–x%的用户对标记为 D;同时,设置 $D=\{\langle u_i,u_j\rangle|G(i,j)=0\}$清除 D 中已存在的信任关系.L 表示通过本方法所预测的信任关系集合,例如在图 3 的例子中,$L=O+P+Q$,其中,$O\in D,P\in B,Q\in B,|L|=|D|$.那么,信任关系预测准确率计算公式为

$$TP = \frac{|L \cap D|}{|L|} \tag{16}$$

其中,TP 表示信任关系预测准确率,$|L\cap D|$表示通过本方法所预测的信任关系与 D 中交集的个数.

4.3 社会等级理论和同质性理论在社会网络中的验证

4.3.1 社会等级理论的验证

社会等级理论认为,低等级的用户更易与比高等级的用户建立信任关系.为了证明社会等级理论在社会网络中的存在,需验证如下问题:具有低等级的用户与低等级用户建立了更多的信任关系,还是与高等级用户建立了更多的信任关系?

本实验中,用户的社会等级通过 Pagerank 算法获得.然后,将所有用户分为 10 个等级,标记为

$$\{Level_1,Level_2,\ldots,Level_{10}\},$$

其中,$Level_1$ 具有最高的等级级别,$Level_{10}$ 具有最低的等级级别.那么,各个级别之间建立信任关系的比率如图 1 所示.

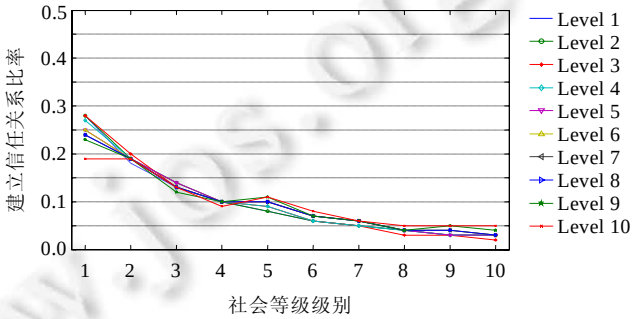


Fig.1 The ratio of establishing trust relations with different status levels

图 1 各个级别之间建立信任关系的比率

从图 1 可以看出:各个级别之间建立信任关系的比率不同,且随着等级级别的增加,比率逐渐降低.表明用户相比于低等级用户与高等级用户建立了更多的信任关系,即验证了社会等级理论在社会网络中的存在.从图 1 中还可以得出更深入的结论:无论哪一等级的用户,其规律都基本一致,即对方等级越高,越容易建立信任关系.

此外,如果将所有用户分为 10 个等级,$1\leqslant i<j\leqslant 10$,那么$\langle Level_i,Level_j\rangle$表示高等级用户向低等级用户建立信任关系;如果 $1\leqslant j<i\leqslant 10$,那么$\langle Level_i,Level_j\rangle$表示低等级用户向高等级用户建立信任关系.对每个$\langle Level_i,Level_j\rangle$,计算 $Level_i$ 中用户向 $Level_j$ 中用户建立信任关系的数量.那么,使用核心平滑密度估计(kernel smoothing density)方法,实验结果如图 2 所示,可以看出,低等级用户信任高等级用户的信任关系数量的中心数(concentrated numbers of trust relations)明显高于高等级用户信任低等级用户的信任关系数量的中心数,进一步验证了低等级用户向高等级用户建立了更多的信任关系.

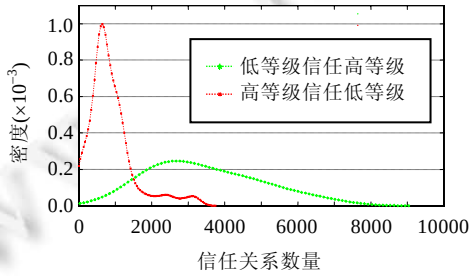


Fig.2 Density estimate of numbers of trust relations

图 2 信任关系数量的密度评估

4.3.2 同质性理论的验证

同质性理论认为,相似的用户更易于建立信任关系.为了证明同质性理论在社会网络中的存在,需要验证如

下问题:相似的用户间建立了更多的信任关系,还是不相似的用户间建立了更多的信任关系?

以 Music 领域为例,具有信任关系的用户对仅占 0.0043.将用户对按照相似度大小进行排序,每次从高相似对开始取出前 s%对作为高相似度对,再从尾部取出后 s%对作为低相似度对.由于信任关系非常稀疏,令 $s\in[0.0001,0.05]$,那么高相似对建立信任关系的数量与低相似度对建立信任关系数量如图 3 所示.

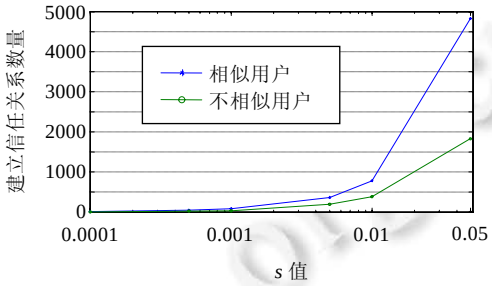


Fig.3 The correlations of user similarity and trust relations

图 3 用户相似度与信任关系的联系

从实验结果可以看出,相似的用户间建立了更多的信任关系.即验证了同质性理论在社会网络中的存在.

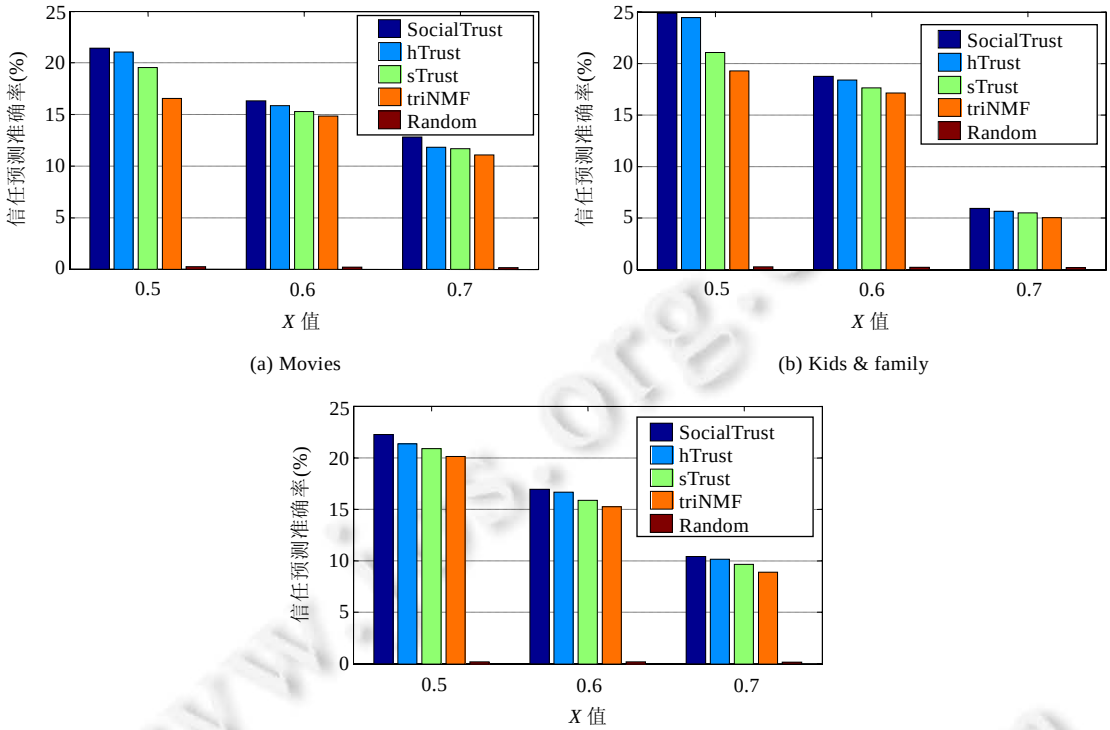
4.4 实验结果分析

由于信任具有领域相关性,分别针对不同领域的用户实现信任关系预测.对于每个领域,实验数据集 A 中的 x 分别取值 0.5,0.6,0.7,验证信任关系预测模型 SocialTrust 的有效性,并分别与以下 4 种方法进行对比:

- triNMF:仅使用非负矩阵的三维分解方法,而不考虑社会等级理论和同质性理论对社会网络的影响;
- hTrust:仅考虑同质性理论对信任关系的影响;
- sTrust:仅考虑社会等级理论对信任关系的影响;
- Random:随机预测信任关系,不考虑任何规则.

本文所提出的方法属于非监督方法,因此并不与已有的监督方法进行对比.所有的信任关系预测参数均通过交叉验证获得,其中,α=0.1,β=0.1,λ_1=0.7,λ_2=0.7,d=100.分别取实验数据集 A 中的 x%作为训练集,验证所识别的信任关系在后 1–x%中的比率.各个领域的实验结果如图 4 所示,通过比较不同方法的实验结果可以看出:针对不同领域,SocialTrust,hTrust,sTrust 和 triNMF 具有较好的性能,且本文所提出模型 SocialTrust 相比于 hTrust,sTrust,triNMF 方法具有较好的性能,进一步说明结合同质效应和社会等级效应在一定程度上提高了信任关系预测的准确率.Random 方法实验性能最差,更说明信任关系预测的必要性.此外,随着 x 的增加,所有方法的性能都逐渐降低;特别地,当 x=0.5 和 x=0.6 时,Kids & Family 领域的预测精度最高,Books 和 Movies 领域基本相近;当 x=0.7 时,在 Kids & Family 领域的预测精度最低,Movies 领域的预测精度最高.通常来说,训练集越大,预测效果应该更好.然而在本实验中,由于 A 和 B 数量固定,如果 x 增加,那么 1–x%则会减少,从 $D\cup B$ 中预测信任关系将会变得更困难,导致随着 x 的增加,所有方法的性能都会降低.

对于信任关系预测模型 SocialTrust,参数λ_1和λ_2分别控制着社会等级理论和同质性理论对 SocialTrust 的影响.图 5 给出了社会等级理论和同质性理论在不同参数控制下对信任关系预测的影响,其中,x 分别取值 0.5,0.6,0.7.λ_1,λ_2 分别取值为 0,0.01,0.1,0.5,0.7,1,10,且$\lambda_1=\lambda_2$(用λ代替).通过比较不同的λ值可以看出:随着λ的增加,在不同领域中,信任关系预测准确率具有相同的模式:首先增加,然后达到峰值,最后降低.当λ=0 时,即不考虑社会等级理论和同质性理论对信任关系预测的影响,信任关系预测准确率最低.随着λ的增大,信任关系预测准确率不断变化,在$\lambda\in[0.7,1]$时达到峰值.此时,如果λ继续增大,则会导致信任关系预测准确率急剧下降.因为当λ非常大时,社会等级正则化和同质性正则化控制整个学习过程,使得所学习得到的矩阵 U 和 H 失真,产生不精确的预测结果.

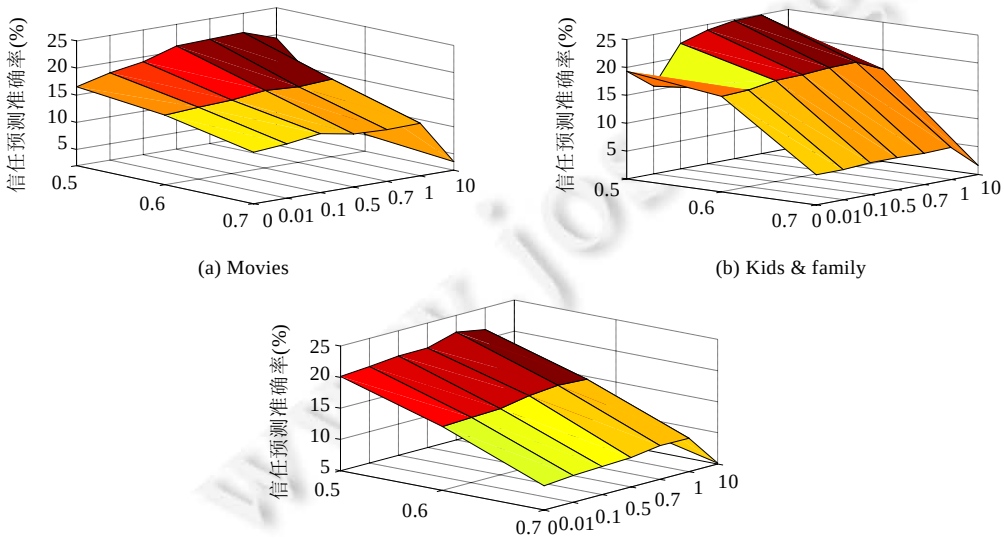


(a) Movies (b) Kids & family

(c) Books

Fig.4 The accuracy of trust prediction in different domains

图 4 不同领域信任关系预测准确率



(a) Movies (b) Kids & family

(c) Books

Fig.5 The impact of social status theory and homophily theory in SocialTrust

图 5 社会等级理论和同质性理论对信任模型 SocialTrust 的影响

5 结 论

本文通过研究社会学理论中的社会等级理论和同质性理论获取信任关系的发展规律,进而基于社会学理论构建信任关系预测模型.该方法不但有效地解决了数据稀疏性问题,还针对用户所属的不同领域实现了信任关系的预测.实验结果表明:相比于传统方法,所提出方法具有较高的精度.

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,以更好地解析社会网络数据;同时,进一步探索时间因素对信任关系预测的影响.

致谢 在此,我们向对本文工作给予支持和建议的评审老师表示感谢,同时,衷心感谢亚利桑那州立大学刘欢教授,感谢 DMML 实验室提供数据集,感谢 DMML 实验室的所有同行.

References:

[1] Mayayise T, Olusegun OI. E-Commerce assurance models and trustworthiness issues: An empirical study. Journal of Information Management & Computer Security, 2014,22(1):76–96. [doi: 10.1108/IMCS-01-2013-0001]

[2] Zhang J, Cohen R. A framework for trust modeling in multiagent electronic marketplaces with buying advisors to consider varying seller behavior and the limiting of seller bids. ACM Trans. on Intelligent Systems and Technology (TIST), 2013,4(2):1–33. [doi: 10.1145/2438653.2438659]

[3] Hoogendoorn M, Jaffry SW, van Maanen PP, Treur J. Design and validation of a relative trust model. Knowledge-Based Systems (KBS), 2014,57:81–94. [doi: 10.1016/j.knosys.2013.12.012]

[4] Kennedy WG, Krueger F. Building a cognitive model of social trust within ACT-R. In: Proc. of the AAAI Conf. on Artificial Intelligence (AAAI). Palo Alto: AI Access Foundation, 2013. 29–34.

[5] Ivanov I, Vajda P, Korshunov P, Ebrahimi T. Comparative study of trust modeling for automatic landmark tagging. IEEE Trans. on Information Forensics and Security (TIFS), 2013,8(6):911–923. [doi: 10.1109/TIFS.2013.2242889]

[6] Liu YH, Sun Y, Liu SY, Kot AC. Securing online reputation systems through trust modeling and temporal analysis. IEEE Trans. on Information Forensics and Security (TIFS), 2013,8(6):936–948. [doi: 10.1109/TIFS.2013.2238929]

[7] Kuter U, Golbeck J. Using probabilistic confidence models for trust inference in Web-based social networks. ACM Trans. on Internet Technology, 2010,10(2):1–23. [doi: 10.1145/1754393.1754397]

[8] Tang J, Lou TC, Kleinberg J. Inferring social ties across heterogeneous networks. In: Proc. of the 5th ACM Int'l Conf. on Web Search and Data Mining (WSDM). Seattle: Association for Computing Machinery, 2012. 743–752. [doi: 10.1145/2124295.2124382]

[9] Rettinger A, Nickles M, Tresp V. Statistical relational learning of trust. Machine Learning, 2011,82(2):191–209. [doi: 10.1007/s10994-010-5211-x]

[10] Hamed K, Bashah IN, Jamalul-lail AM. A unified trust model for pervasive environments-simulation and analysis. KSII Trans. on Internet & Information Systems, 2013,7(7):1569–1584. [doi: 10.3837/tiis.2013.07.003]

[11] Lee MKO, Efraim T. A trust model for consumer Internet shopping. Int'l Journal of Electronic Commerce, 2001,6(1):75–92.

[12] Surya N, Paris C, Bista SK, Sherchan W. A trust model based analysis of social networks. Int'l Journal of Trust Management in Computing and Communications, 2013,1(1):3–22.

[13] Liu HF, Lim EP, Lauw HW, Le MT, Sun AX, Srivastava J, Kim YA. Predicting trusts among users of online communities: An epinions case study. In: Proc. of the 9th ACM Conf. on Electronic Commerce. Chicago: Association for Computing Machinery, 2008. 310–319. [doi: 10.1145/1386790.1386838]

[14] Kiyana Z, Abdollah A. A syntactical approach for interpersonal trust prediction in social Web applications: Combining contextual and structural data. Knowledge-Based Systems (KBS), 2012,26:93–102. [doi: 10.1016/j.knosys.2010.10.007]

[15] Borzymek P, Sydow M, Wierzbicki A. Enriching trust prediction model in social network with user rating similarity. In: Proc. of the Int'l Conf. on Computational Aspects of Social Networks. Fontainbleu: IEEE Computer Society, 2009. 40–47. [doi: 10.1109/CASoN.2009.30]

[16] Xiang RJ, Neville J, Rogati M. Modeling relationship strength in online social networks. In: Proc. of the Int'l Conf. on World Wide Web (WWW). Raleigh: Association for Computing Machinery, 2010. 981−990. [doi: 10.1145/1772690.1772790]

[17] Liu GF, Wang Y, Orgun MA. Social context-aware trust network discovery in complex contextual social networks. In: Proc. of the AAAI Conf. on Artificial Intelligence (AAAI). Toronto: AI Access Foundation, 2012. 101−107.

[18] Huang J, Nie F, Huang H, Lei Y, Ding C. Social trust prediction using rank-*k* matrix recovery. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence (IJCAI). Beijing: Int'l Joint Conf. on Artificial Intelligence, 2013. 2647−2653.

[19] Tang JL, Gao HJ, Hu X, Liu H. Exploiting homophily effect for trust prediction. In: Proc. of the Int'l Conf. on Web Search and Data Mining (WSDM). Rome: Association for Computing Machinery, 2013. 53−62. [doi: 10.1145/2433396.2433405]

[20] Tang JL, Gao HJ, Liu H. mTrust: Discerning multi-faceted trust in a connected world. In: Proc. of the Int'l Conf. on Web Search and Data Mining (WSDM). Seattle: Association for Computing Machinery, 2012. 93−102. [doi: 10.1145/2124295.2124309]

[21] Fang H, Bao Y, Zhang J. Leveraging decomposed trust in probabilistic matrix factorization for effective recommendation. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence (AAAI). Quebec: Association for the Advancement of Artificial Intelligence, 2014. 30−36.

[22] Leskovec J, Huttenlocher D, Kleinberg J. Signed networks in social media. In: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems. Atlanta: Association for Computing Machinery, 2010. 1361−1370. [doi: 10.1145/1753326.1753532]

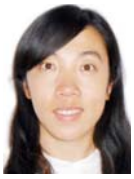
[23] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature, 1999,401(6755):788−791. [doi: 10.1038/44565]

[24] Huang J, Nie F, Huang H, Tu YC. Trust prediction via aggregating heterogeneous social networks. In: Proc. of the 21st ACM Int'l Conf. on Information and Knowledge Management. Maui: Association for Computing Machinery, 2012. 1774−1778. [doi: 10.1145/2396761.2398515]

[25] Ye JJ. Constraint qualifications and KKT conditions for bilevel programming problems. Mathematics of Operations Research, 2006, 31(4):811−824. [doi: 10.1287/moor.1060.0219]

[26] Lin CJ. Projected gradient methods for nonnegative matrix factorization. Neural Computation, 2007,19(10):2756−2779. [doi: 10.1162/neco.2007.19.10.2756]

[27] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology, 2007,58(7):1019−1031. [doi: 10.1002/asi.20591]



王英(1981－),女,吉林长春人,博士,讲师,CCF 高级会员,主要研究领域为社会网络,数据挖掘,搜索引擎.

E-mail: wangying2010@jlu.edu.cn

左万利(1957－),男,博士,教授,博士生导师,主要研究领域为机器学习,数据挖掘,搜索引擎,数据库.

E-mail: zuowl@jlu.edu.cn

王鑫(1981－),男,博士,CCF 学生会员,主要研究领域为社会网络,数据挖掘,搜索引擎.

E-mail: wangxccs@126.com