第 43 卷 第 4 期 计 算 机 工 程 2017 年 4 月
Vol. 43 No. 4 Computer Engineering April 2017

·先进计算与数据处理· 文章编号:1000-3428(2017)04-0034-05 文献标志码:A 中图分类号:TP393

面向社交大数据的个体行为信任评价

黎梨苗[1,2],陈志刚[2],刘志雄[1],叶 晖[1]

(1. 长沙学院 数学与计算机科学系,长沙 410003; 2. 中南大学 软件学院,长沙 410075)

摘 要: 由于大数据环境下个体行为具有多样性的特点,使得基于局部信息的一般个体行为信任评价模型考虑因素不全面,导致个体面临信任危机。为此,提出一种改进的个体行为信任评价模型。采用多数据融合获得信任评价结果,利用 D-S 理论对关联信任评价的个体信任 mass 函数值与评估结果进行整合,计算个体出现不信任情况的概率。融合个体信任态势求出关联个体的不信任态势,获得个体参与信任评价的权重,得出个体行为信任评价。实验结果表明,与基于局部信息的一般个体行为信任评价模型相比,该模型具有更高的可靠性和安全性。

关键词: 大数据;社交网络;信任评价;个体行为;信任态势

中文引用格式: 黎梨苗,陈志刚,刘志雄,等. 面向社交大数据的个体行为信任评价[J]. 计算机工程,2017,43(4):34-38.

英文引用格式: Li Limiao, Chen Zhigang, Liu Zhixiong, et al. Individual Behavior Trust Evaluation for Social Big Data[J]. Computer Engineering,2017,43(4):34-38.

Individual Behavior Trust Evaluation for Social Big Data

LI Limiao[1,2], CHEN Zhigang[2], LIU Zhixiong[1], YE Hui[1]

(1. Department of Mathematics and Computer Science, Changsha University, Changsha 410003, China;
2. School of Software, Central South University, Changsha 410075, China)

【Abstract】 Because the individual behavior has diversity characteristics under the big data environment, it results that the consideration of the general individual behavior trust evaluation model based on local information is not comprehensive, causing serious trust crisis of individual. Therefor, this paper presents an improved trust evaluation model of individual behavior. Firstly, it uses multi-data fusion to get the result of trust evaluation, and then fuses the mass function of the individual trust of relative trust evaluation and the evaluation results by using the theory of D-S. It gets the probability of the individual appearing distrust. Each related individual distrust situation is obtained by the fusion with individual trust situation factor, and it can get the fusion of weight about each individual participating in the individual behavior trust evaluation and the individual behavior trust evaluation. Experimental result shows the proposed model has higher reliability and security than the general trust evaluation model based on local information.

【Key words】 big data; social network; trust evaluation; individual behavior; trust situation

DOI: 10.3969/j.issn.1000-3428.2017.04.006

0 概述

大数据时代的到来进一步丰富了人们以计算机网络为媒介的个体网络行为,包括使用网络选择的行为、信息交换的行为、个体之间交往的行为[1]。本文所提的社交指的是个体在网络上实现的人与人之间的交往。目前,主要流行的社交网络有 Google Circle, Facebook 等,即时通信有 Wechat、微信、QQ 等,微博有新浪微博、Twitter 等,不同类型的博客有科技博客、新浪博客等,视频共享有 Youtube、优酷等,图片分享有 Flickr, Photovine 等,网络书签有 Digg, Reddit, Delicious 等。个体在社交网上的行为各种各样,有浏览、点赞、评论、留言、发布图文、转发等,研究个体社交行为需要对行为数据进行处理来分析个体行为的特征[2]。本文采用网络爬虫来获得大数据,从而对个体的信任进行计算。社交大数据环境下个体信任评价模型和关键技术已倍受学者的关注,已成为了他们研究的重点。因此,本文提出一

基金项目: 国家自然科学基金(61379057,61502057);长沙市科技局计划项目(ZD1601035,ZD1601038)。

作者简介: 黎梨苗(1979—),女,博士,主研方向为大数据、网络安全;陈志刚,教授、博士、博士生导师;刘志雄、叶 晖,博士。

收稿日期: 2016-03-31 **修回日期:** 2016-06-12 **E-mail:** 305209431@qq.com

种面向社交大数据的个体行为信任评价模型。采用多信息融合获得信任评价结果并结合 D-S 理论对关联信任评价的个体信任 mass 函数值与评估结果进行融合[3],求出节点出现不信任情况的概率。然后融合态势要素求出每个关联节点的不信任态势,再根据每个关联节点的信任态势获得每个节点参与信任评价的权重融合,在上述基础上得出个体行为信任评价模型。本文提到的态势是指在社交网络中个体所面临的信任危险,造成信任危险的原因包括个体本身在网络中的行为产生的影响及评价个体给出的评价对个体信任的影响。

1　融合个体行为评价数据的信任评价

在社交大数据环境下,依据个体行为评价数据融合的个体信任问题,态势评估主要包括个体行为评价数据融合、个体行为评价态势融合及个体信任态势融合。个体行为评价数据融合主要是将各种各样相关的行为数据进行信息融合,从而获得社交大数据环境下个体信任出现问题的概率[4-5];个体行为评价态势融合主要是通过个体不信任发生的概率、个体信任的概率以及信任受影响的概率求出个体节点的信任态势;节点信任态势融合是根据各关联节点的信任态势与其权重求出个体信任态势。

1.1　个体行为评价数据融合

个体行为数据融合是对个体行为(如浏览、点赞、评论、留言、发布图文、转发等)数据进行汇总,获得个体由此产生不信任的概率,因此,识别框架定义为 $\Theta=\{m,\bar{m}\}$,幂集为 $2^{\Theta}=\{\phi,\{m\},\{\bar{m}\},\{M\}\}$,其中,$\phi$ 指个体行为的异常数据导致不信任产生但又没产生;M 指异常数据可能导致不信任产生但又没产生。对个体行为数据汇总结果 i 的 mass 函数表示为:$f_i(\phi)=0$,$f_i(m)=g_i(m)$,$f_i(\bar{m})=g_i(\bar{m})$,$f_i(M)=0$。由于本文使用 D-S 理论来处理冲突数据时,要采用组合原则的归一化理论进行处理,在此过程中导致了不符合推理的结论,因此本文采用优化的 D-S 证据理论来合成个体行为数据汇总结果的 mass 函数值,表示如下:

$$\begin{cases} f_i(\phi)=0 \\ f_i(m)=\prod_{i=1}^{n} f_i(m)+eu(m) \\ f_i(\bar{m})=1-f(m) \\ f_i(M)=0 \end{cases} \tag{1}$$

其中:

$$e=1-\prod_{i=1}^{n} f_i(m)-\prod_{i=1}^{n} f_i(\bar{m})$$

$$u(m)=\frac{1}{n}\left(\sum_{i=1}^{n} f_i(m)+\sum_{i=1}^{n} f_i(\bar{m})\right)$$

本文采用 D-S 证据理论完成社交大数据环境下个体行为数据的融合,获得个体行为不信任产生的概率 $f(m)$,此函数能够描述出相关参与评价的个体受到非参与评价个体的行为影响信息,然后根据被评价个体行为态势要素融合来获得参与评价的每个个体的不信任态势。

1.2　个体行为评价态势融合

在社交大数据环境下,个体行为不信任产生的机率能够通过个体行为评价体现出来,可是,个体行为不信任的产生对个体的信任影响主要取决于参与评价的个体评价数据、个体本身产生的消极数据及恶意个体给出的消极评价数据 3 个因素,因此,本文需根据个体行为不信任产生概率 $f(m)$、不信任产生成功率 $c(m)$ 及恶意个体给出的消极评价影响 q 这 3 个因素,通过态势融合来获得个体在社交大数据环境下个体信任影响的态势[6]。

依据个体行为判断是否含有不信任产生的数据,如果没有,那么不信任产生成功的概率为 0;如果含有,且每一项行为都含有,那么不信任产生成功的概率为 1;否则一一判断哪一项个体行为评价存在不信任产生的数据,将此数据进行权重累加,得到不信任产生成功的概率 $c(m)$。

结合个体行为不信任产生的概率 $f(m)$、不信任产生成功的概率 $c(m)$ 以及恶意个体给出的消极评价影响 q,得到不信任产生对个体的信任影响,表示如下:

$$l=f(m)c(m)q \tag{2}$$

假设个体行为评价在相同时间内同时受到许多参与评价个体的恶意评价,将所有恶意评价对个体不信任产生起到的效果进行累加,将其视为个体行为评价在相同时间内所受到不信任产生的影响,即个体行为的不信任产生态势 L,表示如下:

$$L=\frac{1}{t}\sum_{i=1}^{n} l_i \tag{3}$$

其中,t 为评价时间段;n 指个体节点被恶意节点评价的次数;l_i 指各恶意个体评价对个体信任造成的影响。

1.3　个体信任态势融合

在社交大数据环境下,依据各个参与评价的个体信任态势权重的融合计算出个体不信任产生的态势值[7]。根据个体行为数据和各行为所占的权重情况,求出个体信任的态势权重,表示如下:

$$k_m=\sum_{i=1}^{s} k_i \tag{4}$$

其中,s 指个体参与社交行为的次数;k_i 指个体各行为所占权重。如果社交大数据环境下,个体各行为的权重之和为 1,假设个体只进行一种社交行为(如浏览),那么把该个体的行为权重平均分配到上述行为,因此,个体信任的权重之和也是 1。然后根据参与评价的各个个体的不信任态势 L 与个体信任态势权重 k_m,可以求出个体信任态势值 CZ。

$$CZ = \sum_{i=1}^{o} L_i k_{Mi} \tag{5}$$

其中，o 指参与评价的个体数；L_i 指参与评价的个体信任态势；k_{Mi} 指个体各行为所占权重。综上所述，可以利用个体行为数据融合、个体行为评价态势融合与个体信任态势融合得到社交大数据环境中个体信任态势值 CZ，得到个体的信任情况，在上述基础上分析个体信任评价模型[8-9]。

2 面向社交大数据的个体信任评价模型

在时间段 $t-1$ 到 t 内，个体获得一个恶意评价个体，该恶意个体的不信任态势通过式(6)进行描述：

$$F_{v,j}c(t) = \delta_1 + \delta_2 F_{v,j}c(t-1) \tag{6}$$

其中，δ_1 指个体信任初始降低度，且 $\delta_1 > 0$；δ_2 是虚拟的一个奖励系数，且 $\delta_2 > 0$，主要用来描述连续受到恶意个体评价的个体。从式(6)可以看出，该不信任态势呈逐渐增长趋势，假设生命周期为 $num = 0$ 且个体获得多个恶意个体评价，那么可以依据式(6)求出社交大数据环境下个体信任存在危险的累加值，由此表明个体信任在不断受到威胁。

假设个体在时间段中没有受到恶意个体的评价，那么个体信任威胁根据式(7)所示的关系降低 $\frac{1}{\xi}$。

$$F_{v,j}c(t) = \begin{cases} F_{v,j}c(t-1)\left(1 - \frac{1}{\xi}F_{v,j}num(t-1)\right), \\ F_{v,j}num(t-1) < \xi \\ 0, F_{v,j}num(t-1) \geqslant \xi \end{cases} \tag{7}$$

由式(7)可知，如果不信任态势生命周期 num 增长至 ξ，那么个体不信任受到的影响将降为 0，即在 $t-1$ 到 t 时间段内，如果没有恶意个体参与评价，那么认为恶意个体不会参与评价。

采用从信任威胁中收集重要信息的方法对个体行为评价数据进行融合，依据融合的结果对社交大数据环境下的信任威胁进行分类，此方法在实际生活中已得到广泛应用，且具有一定的准确性[10]。

假设信任威胁指标 $0 \leqslant R_b(t) \leqslant 1$ 为个体 b 在时刻 t 所遭受的信任威胁，如果 $R_b(t) = 1$，那么此时个体信任置于严重危险当中；如果 $R_b(t) = 0$，那么此时个体信任没有危险。但是，随着 $R_b(t)$ 值的不断增加，社交大数据环境下个体信任所面临的威胁也会不断增加，由于个体不同行为的权重以及不同恶意个体评价的威胁不同[11]，因此该类恶意个体评价的威胁程度用 λ_i 来表示，该个体不同行为的权重用 θ_i 表示。个体 b 在时刻 t 遭受第 $i(1 \leqslant i \leqslant A)$ 类 $H_i(t)$ 累加的信任风险将通过式(8)来求得：

$$R_{b,i}(t) = 1 - \frac{1}{1 + In(\lambda_i(\sum_{F_{v,j} \in H_i(t)} F_{v,j}c(t)) + 1)} \tag{8}$$

个体 b 在时刻 t 总的信任风险由式(9)求得：

$$R_b(t) = 1 - \frac{1}{1 + In(\sum_{i=1}^{A} \theta_i(\sum_{F_{v,j} \in H_i(t)} F_{v,j}c(t)) + 1)} \tag{9}$$

在社交大数据环境中，个体将受到第 $i(1 \leqslant i \leqslant A)$ 类 $H_i(t)$ 影响信任的风险 Q_i，表示如下：

$$Q_i(t) = 1 - \frac{1}{1 + In(\lambda_i(\sum_{i=1}^{A} \theta_i(\sum_{F_{v,j} \in H_i(t)} F_{v,j}c(t)) + 1))} \tag{10}$$

由上述方法可见，首先得到所有参与评价的每个个体面临的信任风险，然后利用个体不同行为权重加权的方法逐一得到从直接到间接参与评价的个体所面临的信任风险，这样大大减少了信任风险计算时由于不同信任影响分类带来的计算量[12]。综上所述，个体信任风险计算如下：

$$Q(t) = 1 - \frac{1}{\alpha + In(\sum_{b=i}^{b} R_b(t)\theta_b + 1)} \tag{11}$$

3 实验验证

为验证本文提出的社交大数据环境下个体行为信任危险模型的科学性及有效性，采取网络爬虫来获取个体在社交网上的数据[13-14]。本文对一个学生群体在一个时间段内在社交网络上(QQ、微信、博客、微博、E-mail)的行为数据进行收集[15]。采用本文原理对数据进行处理，从以下方面进行比较与分析。

1)模型科学性

采用本文提出模型与基于局部信息的一般信任计算模型和实际个体信任态势进行比较，得到结果如图 1 所示。

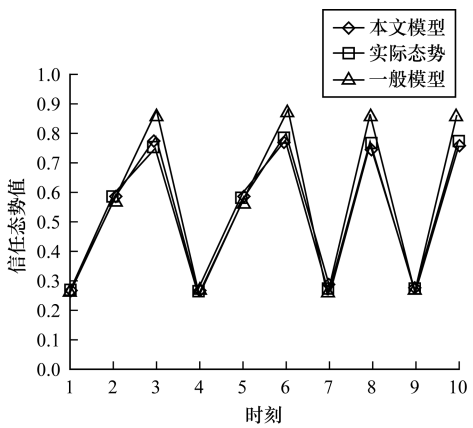

图 1 信任态势比较

由此可知,本文模型下个体信任态势与实际个体态势基本保持一致,而一般模型的个体信任态势与实际信任态势存在明显差异。这表明本文在大数据环境下的个体信任评价模型是科学的,能够通过计算获得个体的信任态势,从而有利于作为网上交易对象选择的判断标准。

2)模型有效性

在上述基础上进一步验证了模型有效性,得到结果如图 2 所示。由此可知,在同样条件下,将本文模型与一般信任模型进行个体信任态势测试,对得到的信任态势有效值进行比较,本文模型所得的值较稳定且有效性高于一般信任计算模型。这表明在本文模型下,对个体信任评价较为精确,符合实际情况,能够较好地把握大数据环境下个体信任存在的安全态势。

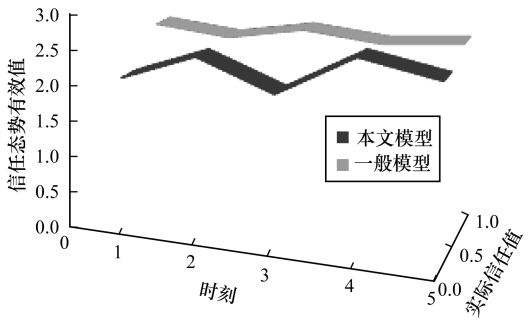

图 2　信任态势有效值比较

3)模型安全性

为验证本文模型的安全性,对恶意节点评价给个体造成的信任态势风险进行实验分析计算。将式(10)中的 λ_i 分别取 0.2,0.4,0.6,0.8,θ_i 取 0.3 进行实验,得到结果如图 3 所示。由此可知,随着恶意节点威胁的增大,个体信任受到的影响较大,这是因为当恶意节点参与评价时个体信任态势同样会受到影响。

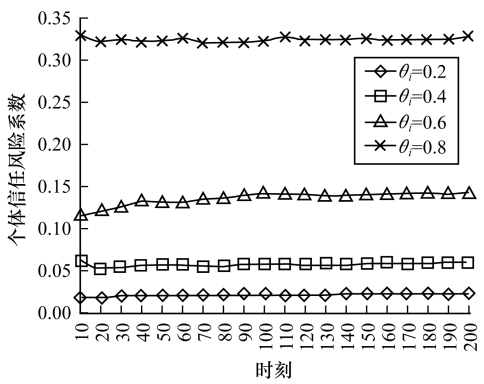

图 3　个体信任风险比较

在上述基础上,当 θ_i 的取值分别为 0.2,0.4,0.6(即不同行为参与评价)时所占权重不同,将节点行为分别对应于转发、评价、点赞。由图 4 分析可知,节点的 3 类行为参与信任计算时,对个体节点平均信任值在仿真周期的前期阶段影响较大。刚开始评价影响较转发大,当仿真周期不断增加时,三者影响区分趋于模糊,但总的趋势是点赞行为对个体信任风险影响较小,直接评价影响较大,这说明本文模型考虑节点行为对信任评价的影响符合实际情况。由此可见,在大多数环境下,本文模型评估结果较精确,能够用此方法很好地评估目前青少年的行为,并对其行为是否正常做出判断。

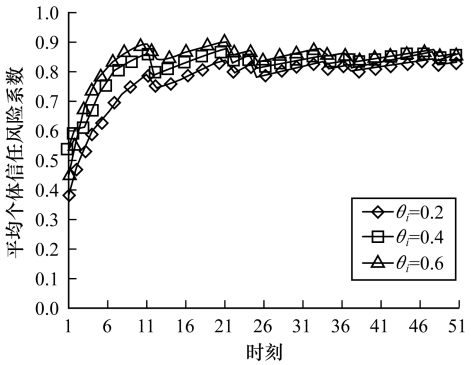

图 4　不同行为对个体信任风险的影响比较

4　结束语

本文在大数据环境下将多数据进行融合获得个体行为信任风险评估模型,将个体行为评价结果进行融合汇总,得到个体行为不信任产生的概率,采用 D-S 理论对关联信任评价的个体信任 mass 函数值与评估结果进行融合,求出节点出现不信任情况的概率。通过融合态势要素获得每个关联节点的不信任态势,得到每个关联节点的参与信任评价的权重融合。根据评价节点各行为和各行为所占的权重得到被评价节点的权重,依据评价节点的信任态势和节点权重得到节点本身的信任风险态势,融合个体行为数据源、态势要素及节点各行为信任态势求出个体的信任风险态势。在上述情况下,分析大数据环境下个体行为信任风险评估模型,仿真实验结果表明,本文模型可靠性及精确度高,符合实际情况,有利于对现实生活中的个体行为进行分析与评价。下一步将利用社会网络中的个体关系程度进行个体信任建模,使获得的节点信任值与节点实际行为相符。

参考文献

[1] 郭 强,郭耀煌,郭春香. 基于模糊相似度的群决策方案排序[J]. 西南交通大学学报,2010,45(2):307-311.

[2] 张 琰,吴 宜. 基于大数据的大学生网络社交行为研究[J]. 电子技术与软件工程,2014(23):30-31.

[3] 张新刚,王保平,程新党. 基于信息融合的层次化网络安全态势评估模型[J]. 网络安全技术与应用,2012,9(4):1072-1074.

[4] Deepa R,Swamynathan S. A Trust Model for Directory-based Service Discovery in Mobile Ad Hoc Networks[M]//Pérez G M,Thampi S M. Recent Trends in Computer Networks and Distributed Systems Security. Berlin,Germany:Springer-Verlag,2014:115-126.

[5] 蔡青松,牛建伟,刘明珠. 一种评估机会社会网络中节点消息传播能力的方法[J]. 软件学报,2012,23(1):49-58.

[6] Emani C K,Cullot N,Nicolle C. Understandable Big Data:A Survey[J]. Computer Science Review,2015,17:70-81.

[7] 田春岐,江建慧,胡治国,等. 一种基于聚集超级节点的 P2P 网络信任模型[J]. 计算机学报,2010,33(2):345-355.

[8] Chen Zhigang,Li Limiao,Gui Jingsong,et al. KFTrust:P2P Trust Model Based on Evaluation Rank Using Kalman Filter[J]. International Journal of Autonomous and Adaptive Communications Systems,2015,8(2/3):268-287.

[9] 钟 琪,戚 巍. 基于态势管理的区域弹性评估模型[J]. 经济管理,2010(8):32-37.

[10] Du Wei,Chen Junliang. The Bayesian Network and Trust Model Based Movie Recommendation System[M]// Du Zhenyu. Intelligence Computation and Evolutionary Computation. Berlin,Germany:Springer,2013:797-803.

[11] Pawar P S,Rajarajan M,Dimitrakos T,et al. Trust Model for Cloud Based on Cloud Characteristics[M]// Fernández-Gago G,Martinelli F,Pearson S,et al. Trust Management VII. Berlin,Germany:Springer-Verlag,2013:239-246.

[12] Zhou Runfang,Hwang K. PowerTrust:A Robust and Scalable Reputation System for Trusted Peer-to-Peer Computing[J]. IEEE Transactions on Parallel and Distributed Systems,2007,18(4):460-473.

[13] 黎梨苗,陈志刚,桂劲松,等. P2P 网络环境下优先权信任模型研究[J]. 计算机工程,2013,39(5):148-151.

[14] Liang Xiaohui,Lin Xiaodong. Enabling Trustworthy Service Evaluation in Service-oriented Mobile Social Networks[J]. IEEE Transactions on Parallel and Distributed Systems,2014,25(2):310-321.

[15] Aringhieri R,Damiani E. Assessing Efficiency of Trust Management in Peer-to-Peer Systems[C]//Proceedings of the 14th IEEE International Workshops on Enabling Technologies:Infrastructure for Collaborative Enterprise. Washington D. C.,USA:IEEE Press,2010:124-132.

编辑 陆燕菲

(上接第 33 页)

[11] Tian Pingfang,Zhu Zhonghua,Li Xiong,et al. A Recommendation Mechanism for Web Publishing Based on Sentiment Analysis of Microblog[J]. Wuhan University Journal of Natural Sciences,2015,20(2):146-152.

[12] Wang Yan,Li Lei,Liu Guanfeng. Social Context-aware Trust Inference for Trust Enhancement in Social Network Based Recommendations on Service Providers[J]. World Wide Web,2015,18(1):159-184.

[13] Ma Hua,Hu Zhigang. Recommend Trustworthy Services Using Interval Numbers of Four Parameters via Cloud Model for Potential Users[J]. Frontiers of Computer Science,2015,9(6):887-903.

[14] Christou I T,Amolochitis E,Tan Zhenghua. AMORE:Design and Implementation of a Commercial-strength Parallel Hybrid Movie Recommendation Engine[J]. Knowledge and Information Systems,2016,47(3):671-696.

[15] Wang Ping,Chao K,Lo C. Satisfaction Based Web Service Discovery and Selection Scheme Utilizing Vague Sets Theory[J]. Information Systems Frontiers,2015,17(4):827-844.

[16] Tang Mingdong,Jiang Yechun,Liu Jianxun,et al. Location-aware Collaborative Filtering for QoS-based Service Recommendation[C]//Proceedings of the 19th IEEE International Conference on Web Services. Washington D. C.,USA:IEEE Press,2012:24-29.

[17] Çelik D,Elçi A. A Broker-based Semantic Agent for Discovering Semantic Web Services Through Process Similarity Matching and Equivalence Considering Quality of Service[J]. Science China Information Sciences,2013,56(1):1-24.

编辑 金胡考