

复杂网络环境下基于推荐链分类的动态信任模型

张琳^{1,2}, 邢欢¹, 王汝传^{1,3,4}, 吴超杰¹

(1. 南京邮电大学 计算机学院, 江苏 南京 210003; 2. 南京邮电大学 计算机技术研究所, 江苏 南京 210003;
3. 江苏省无线传感网高技术研究重点实验室, 江苏 南京 210003;
4. 南京邮电大学 宽带无线通信与传感网技术教育部重点实验室, 江苏 南京 210003)

摘要: 针对网络环境中复杂的推荐信息处理问题, 提出了一种基于推荐链分类的信任模型。该分类方法基于节点间的诚实属性, 在实际经验数据的基础之上能选择出有效的推荐链。针对推荐信息的传播使用了以信息增益为基础的参数, 使推荐信息更精准, 考虑了时间的影响并且能把交互能力与诚实属性清楚地区分开。在最终的直接信任与推荐信息的聚合计算过程中采用了信息论中熵的概念, 摆脱了以往主观设定参数的模糊性。模型中主要的聚合参数能随着交互的进行而不断地修正, 达到了最贴近真实值的情形。仿真实验验证了新模型分类的有效性以及参数设置的合理性。

关键词: 推荐链分类; 信任模型; 信任传播; 信任聚合

中图分类号: TP393

文献标识码: A

Dynamic trust model based on recommendation chain classification in complex network environment

ZHANG Lin^{1,2}, XING Huan¹, WANG Ru-chuan^{1,3,4}, WU Chao-jie¹

(1. College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
2. Institute of Computer Technology, Nanjing University of Posts and Telecommunication, Nanjing 210003, China;
3. Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210003, China;
4. Key Lab of Broadband Wireless Communication and Sensor Network Technology,
Nanjing University of Posts and Telecommunications, Ministry of Education Jiangsu Province, Nanjing 210003, China)

Abstract: According to the recommendation information processing problem in complex network environment, a trust model based on the recommendation chain classification was proposed. The classification method was based on honesty attribute of nodes, which could choose an effective recommendation chain on the basis of practical experience data. The recommendation information dissemination parameters were based on the information gain, which made recommendation information be more accurate. The factor of time was also considered in this model. The ability of interaction and the one of honesty were distinguished clearly. The concept of information entropy in information theory was used in the final aggregation calculation of direct trust and recommendation trust, which could get rid of the ambiguity of the previous subjective parameter settings. The main polymerization parameters could be continuously corrected with the interactions in order to achieve the situation being closest to the reality. Simulation results show the validity of recommendation chain classification and the rationality of the parameter settings in the proposed model.

Key words: recommendation chain classification; trust model; trust propagation; trust aggregation

收稿日期: 2014-06-24; 修回日期: 2014-08-07

基金项目: 国家自然科学基金资助项目(61402241, 61170065, 61373017, 61171053, 61103195, 61203217, 61201163, 61202004, 61202354); 江苏省自然科学基金资助项目(BK2012436); 江苏省科技支撑计划(工业)基金资助项目(BE2012183、BE2012755); 省属高校自然科学研究重大基金资助项目(11KJA520001, 12KJA520002); 江苏省高校自然科学基金资助项目(13KJB520017); 南京邮电大学科研基金资助项目(NY213155); 高校科研成果产业化推进工程基金资助项目(JHB2012-7); 江苏高校优势学科建设工程基金资助项目(yx002001)

Foundation Items: The National Natural Science Foundation of China (61402241, 61170065, 61373017, 61171053, 61103195, 61203217, 61201163, 61202004, 61202354); The Natural Science Foundation of Jiangsu Province(BK2012436); Scientific & Technological Support Project (Industry) of Jiangsu Province(BE2012183, BE2012755); The Natural Science Key Fund for Colleges and Universities of Jiangsu Province (11KJA520001, 12KJA520002); The Natural Science Fund for Colleges and Universities of Jiangsu Province (13KJB520017); The Science Foundation of Nanjing University of Posts and Telecommunications (NY213155); Scientific Research & Industry Promotion Project for Higher Education Institutions(JHB2012-7); The Priority Academic Program Development of Jiangsu Higher Education Institutions(PAPD) (yx002001)

1 引言

当今的网络技术发展迅猛，安全问题层出不穷，越来越多的人投入到网络安全的研究中去。可信计算受到越来越多的人的关注。目前已经有很多学者提出了各式各样的信任模型以及可信机制，有基于模糊理论的^[1,2]，基于概率论的^[3]，基于上下文的^[4,5]，基于 D-S 证据理论的^[6]，基于云理论的^[7,8]等，它们都能在某个方向表现出特定的优势，但或多或少都存在一些问题。

复杂网络环境下网络节点间的信任关系极其复杂并且是动态变化的。很多网络节点直接交互过的节点并不多，所以当它们在试图与某个未曾交互过的节点交互时就需要推荐信息，由于网络节点交互关系的复杂性导致存在大量繁杂的推荐关系。很多学者尝试了一些推荐信任关系传播的计算方法，但很多文献对交互能力与诚实属性未做区分，当需要判断能否与一个节点交互时，其交互能力很重要。然而在使用推荐信息时，需要注重的是这个节点有没有提供了真实的信息，所以本文关注的是节点的诚信值，而非交互能力或者模糊不清的信任值。文献[9]在云计算的背景下使用了粗集理论和 IOWA 算子相结合来进行信任的计算。文献[10]将逻辑符号根据概率论重新定义，提出了一种基于概率逻辑和模糊逻辑相结合的信任模型，虽然该模型考虑了推荐信息，但并没有对推荐信息做深入的细分。文献[11]针对社交网络中搜索最佳信任路径这个 NP 完全问题，引入了新的概念信任质量(QOT)，提出了一种新的基于路径的多预计算法 MFPB-HOSTP。文献[12]将信任分成了独立信任和相关信任，主要研究了相关信任，并使用了一种参数估算技术，最终验证表明这种模型针对相关信任能很好地对信任进行预测。文献[13]提出了一种细粒度描述的信任模型，并对交互能力与诚实属性作了清楚地区分，但在推荐信任的计算中只在聚合时考虑了推荐链的长度，而没有考虑信任在推荐链中的衰减问题。在使用推荐信息时，很少有人对推荐链的选取提出合适的方法，即使提到了也仅仅只是简单地或者主观地做出推荐链的选取，然后再对这些推荐链进行加权合并。文献[14]给出了一种基于推荐链的信任模型，该模型考虑了时间衰减、信任传播中的衰减以及路径的权重分配问题，虽然该模型在选择推荐链时使用了几条

规则，但其本质还是简单地通过与阈值的比较选取推荐链，摆脱不了阈值选取的主观性，而在最终的直接信任与推荐信任的聚合时使用的参数还是以主观地分配 2 个参数，其中直接信任的参数较大。文献[15]主要针对推荐信任中的依赖问题，提出了一种基于证据理论信任模型，然而该模型并没有考虑证据冲突时的情况，在推荐信任聚合时，也没有考虑推荐链的选取问题。文献[16]也没有对推荐信息进行分类。

针对上述文献中的问题，急需一种高效的推荐链分类方法，因为这能够减少很多没必要的推荐信息，能够有效去除推荐信息中的噪声。

2 模型构架

本文模型主要包括直接信任的计算、推荐链分类、推荐信任的计算、推荐信任的聚合以及直接信任和推荐信任的聚合，模型如图 1 所示。

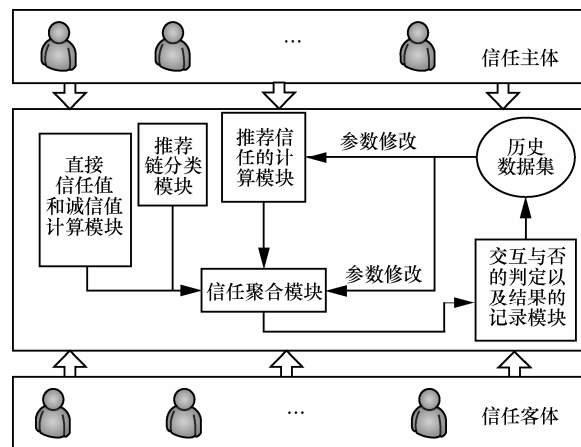


图 1 复杂网络环境下基于推荐链分类的动态信任模型

具体流程如下，如图 2 所示。

Step1 信任客体 C 请求信任主体 B 的某项服务，首先观察他们的交互次数是否大于 α 并且计算直接信任值 T 是否大于 β ，若都满足，表示这 2 个节点是熟识的，信任主体 B 是可信的，可以直接交互，并且把交互结果存放于历史记录中，执行 Step4。否则执行 Step2。

Step2 在交互次数和直接信任值之间如有一个不满足条件，就需要推荐信息。获取推荐链并用分类法加以分类，通过推荐信任计算方法和历史数据集提供的参数 $G(S, i)$ 计算得到推荐信任值。最后通过历史数据集提供的参数 ϵ 计算得到最后的信誉值。

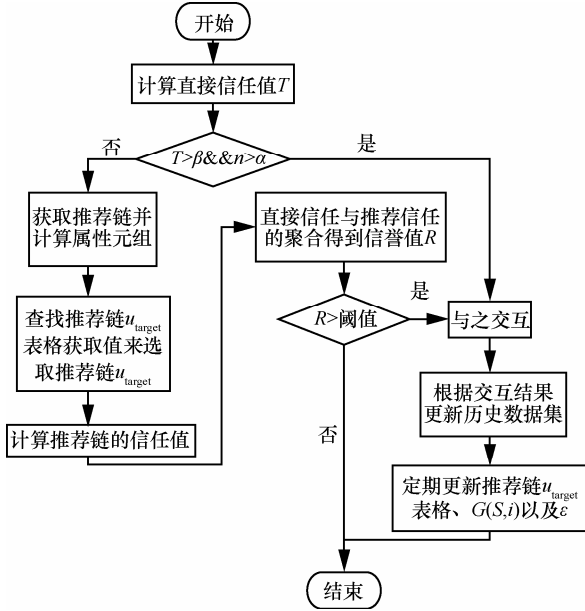


图2 模型流程

Step3 判断信誉值是否大于阈值，若大于则交互，否则寻找其他提供此项服务的信任主体。将交互结果存放于历史记录中。

Step4 在一个固定的时间周期之后，根据历史记录刷新推荐链 u_{target} 表格、参数 $G(S,i)$ 和参数 ϵ 。

3 信任的计算

3.1 直接信任的计算

定义 1 (评价函数) 节点 C_i 与节点 C_j 直接交互后， C_i 对 C_j 的评价函数如下

$$f_{C_i \rightarrow C_j}^k = \begin{cases} y_1, & \text{非常满意} \\ y_2, & \text{较为满意} \\ y_3, & \text{一般满意} \\ y_4, & \text{不满意} \end{cases} \quad (1)$$

其中， y_1 、 y_2 、 y_3 、 y_4 根据实际经验设定。

由于信任受到时间的影响，这里将时间分成每段大小相等的时间段，时间段大小为 t ，第 x 个时间段内节点 C_i 与节点 C_j 交互 n 次，则第 x 个时间段

内节点 C_i 对节点 C_j 直接信任值为 $f_x = \frac{\sum_{k=1}^n f_{C_i \rightarrow C_j}^k}{n}$ 。

定义 2 时间衰减函数 $g_x = \alpha^{N_c - x}$ ， N_c 为当前的时间段， x 为节点 C_i 与节点 C_j 存在交互的时间段。

定义 3 (直接信任值) 考虑时间衰减的影响得到直接信任值如下

$$D_{C_i \rightarrow C_j} = \frac{\sum_x g_x f_x}{\sum_x g_x} \quad (2)$$

3.2 诚信值的计算

当没有足够的直接交互经验时就需要邻居节点的推荐信息，这里需要的是邻居节点的诚实属性而非邻居节点的交互能力属性。诚信值来源于节点每次评价是否属实。

节点 C_i 与提供服务的节点 C_j 直接交互后， C_i 对 C_j 的评价函数 $f_{C_i \rightarrow C_j}^k$ ，提供服务的节点 C_j 对于本次服务的预判评价值为 $h_{C_j \rightarrow C_i}$ ，这里的预判评价值即提供服务的节点对本次交互给出的评价，使用评价函数与预判评价值的差值作为诚实函数

$$honesty_{C_i \rightarrow C_j}^k = \begin{cases} 1, & |f_{C_i \rightarrow C_j}^k - h_{C_j \rightarrow C_i}| \leq \theta \\ 0, & \text{未作评价} \\ -1, & |f_{C_i \rightarrow C_j}^k - h_{C_j \rightarrow C_i}| > \theta \end{cases} \quad (3)$$

其中， θ 为诚实函数阈值。

第 x 个时间段内节点 C_i 与提供服务的节点 C_j 交互 n_x 次，则第 x 个时间段内节点 C_i 的诚实函数

$$honesty_{C_i \rightarrow C_j}(x) = \frac{\sum_{k=1}^{n_x} honesty_{C_i \rightarrow C_j}^k}{n_x} \quad (4)$$

诚信值同样受到时间的影响，所以需要时间衰减函数 $g_x = \alpha^{N_c - x}$ ， N_c 为当前的时间段， x 为节点 C_i 与节点 C_j 存在交互的第 x 个时间段，所以节点 C_i 与节点 C_j 交互所得 C_i 的诚信值为

$$honesty_{C_i \rightarrow C_j} = \frac{\sum_x g_x honesty_{C_i \rightarrow C_j}(x)}{\sum_x g_x} \quad (5)$$

定义 4 活跃度是用来描述节点交互行为的频繁程度。节点 i 关于节点 j 的活跃度定义如下

$$h_i^j = 1 - \frac{1}{n_x + \partial} \quad (6)$$

其中， n_x 表示节点 i 与节点 j 的在第 x 个时间段内的交互次数，其中参数 $\partial > 0$ 。

在计算诚信值时，假设节点 i 与节点 j 只在最近的时间段内交互过一次，并且 $honesty_{C_j \rightarrow C_i}^k = 1$ ，

可得到最后的 $honesty_{C_j \rightarrow C_i} = 1$ ，显然只交互过一次却得到最大的诚信值是不可靠的。因此需要一种基于活跃度的诚信值，这里使用的是活跃度与诚信值的乘积。第 x 个时间段的基于活跃度的诚信值

$$honesty_{C_i \rightarrow C_j}^h(x) = \left(1 - \frac{1}{n_x + \delta}\right) \frac{\sum_{k=1}^{n_x} honesty_{C_i \rightarrow C_j}^k}{n_x} \quad (7)$$

所以，节点 C_i 与节点 C_j 交互所得 C_i 基于活跃度的诚信值

$$honesty_{C_i \rightarrow C_j}^h = \frac{\sum_x g_x honesty_{C_i \rightarrow C_j}^h(x)}{\sum_x g_x} \quad (8)$$

4 推荐链的分类

把每个节点根据诚信值分为 5 个级别，由下列映射函数确定。这里的 $v1, v2, v3, v4$ 根据实际情况确定。

$$\varphi(honesty_{C_i \rightarrow C_j}^h) = \begin{cases} A, & v4 < honesty_{C_i \rightarrow C_j}^h \leq 1 \\ B, & v3 < honesty_{C_i \rightarrow C_j}^h \leq v4 \\ C, & v2 < honesty_{C_i \rightarrow C_j}^h \leq v3 \\ D, & v1 < honesty_{C_i \rightarrow C_j}^h \leq v2 \\ E, & 0 \leq honesty_{C_i \rightarrow C_j}^h \leq v1 \end{cases} \quad (9)$$

当缺少直接交互信息时，就需要准确的推荐信息，可靠的推荐信息来源于推荐节点较高的诚信度，所以，此时需要的是节点的诚实属性而非交互能力。

推荐链：由申请服务节点和提供服务节点之间的节点组成。假设有如图 3 所示的交互关系。

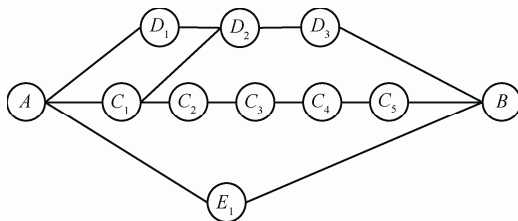


图 3 推荐关系

服务申请者 A 申请有关 B 的服务时，存在 4 条推荐链：(1) D_1, D_2, D_3 ；(2) C_1, C_2, C_3, C_4, C_5 ；(3) E_1 ；(4) C_1, D_2, D_3 。

在实际的社交网络中，由于交互环境极其复杂，所以存在大量的推荐链。由六度分离理论，

即每个人最多通过 5 个人就能认识世界上任何一个人。所以推荐链的长度小于等于 5 即可。在很多的推荐链中，由于每个节点的诚信值有 5 个等级，所以每条推荐链有 5^5 即 3 125 种。很多文献直接把多条推荐链加权平均得到一个推荐的结果，而很多推荐链可能由于其中某个或多个节点是恶意的而导致整条推荐链是不可信的，这样做的后果是在推荐信息中加了很多噪声，导致最后的推荐信息的不准确。如果仅凭主观判断，很模糊而且误差大，所以是很困难的。针对以上问题必须对推荐链做出精确选择，选择的唯一办法是对这些推荐链进行分类，然后再选择可靠的推荐链。

为了区分每一类推荐链，这里用每条推荐链的每个节点的诚信值的级别表示成属性值的一个元组 $\langle honesty_1, honesty_2, honesty_3, honesty_4, honesty_5 \rangle$ 。用图 3 中的推荐链(2)，则推荐关系如图 4 所示。

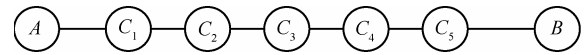


图 4 推荐关系链

推荐链中使用的信任值是推荐信息节点关于被推荐信息的节点的诚信值。在现实社会中，某些人的不诚信行为大多数只是针对某些固定的人，一个人的综合信誉不能代表这个人对于某些人的诚信程度。所以在推荐链中只需要推荐信息节点对于被推荐信息的节点的诚信值，而不是推荐节点的综合信任值。

为了表达简便，有

$honesty_1$ 表示 $\varphi(honesty_{C_1 \rightarrow A}^h)$ ，

$honesty_2$ 表示 $\varphi(honesty_{C_2 \rightarrow C_1}^h)$ ，

$honesty_3$ 表示 $\varphi(honesty_{C_3 \rightarrow C_2}^h)$ ，

$honesty_4$ 表示 $\varphi(honesty_{C_4 \rightarrow C_3}^h)$ ，

$honesty_5$ 表示 $\varphi(honesty_{C_5 \rightarrow C_4}^h)$ 。

这里设一个目标值集合 $U = \langle useful, unuseful \rangle$ ，而目标值用 u_j 表示，目标是在给定属性值的一个元组 $\langle honesty_1, honesty_2, honesty_3, honesty_4, honesty_5 \rangle$ ，得到最可能的目标值 u_{target} 。用概率公式表示成

$$u_{target} = \arg \max_{u_j \in U} P(u_j | honesty_1, honesty_2, honesty_3, honesty_4, honesty_5) \quad (10)$$

由贝叶斯公式可以改写成

$$\begin{aligned}
 u_{\text{target}} &= \arg \max_{u_j \in U} \frac{P(\text{honesty}_1, \text{honesty}_2, \text{honesty}_3, \text{honesty}_4, \text{honesty}_5 | u_j) P(u_j)}{P(\text{honesty}_1, \text{honesty}_2, \text{honesty}_3, \text{honesty}_4, \text{honesty}_5)} \\
 &= \arg \max_{u_j \in U} P(\text{honesty}_1, \text{honesty}_2, \text{honesty}_3, \text{honesty}_4, \text{honesty}_5 | u_j) P(u_j) \quad (11)
 \end{aligned}$$

这里，推荐链中每个推荐节点的诚信级别之间是相互独立的，所以联合概率是每个节点诚信级别的概率乘积。表达式如下

$$\begin{aligned}
 &P(\text{honesty}_1, \text{honesty}_2, \text{honesty}_3, \text{honesty}_4, \text{honesty}_5 | u_j) \\
 &= \prod_i P(\text{honesty}_i | u_j) \quad (12)
 \end{aligned}$$

将其代入式(9)得

$$u_{\text{target}} = \arg \max_{u_j \in U} P(u_j) \prod_i P(\text{honesty}_i | u_j) \quad (13)$$

显然，当 u_{target} 的值是 *useful* 时，该推荐链是需要的，反之亦然。例如，服务请求节点存在推荐链，其属性元组 $\langle B, C, A, D, A \rangle$ ，计算得到 $P(\text{useful})P(B | \text{useful})P(C | \text{useful})P(A | \text{useful})P(D | \text{useful})P(A | \text{useful}) = 2.3 \times 10^{-4}$; $P(\text{unuseful})P(B | \text{unuseful})P(C | \text{unuseful})P(A | \text{unuseful})P(D | \text{unuseful})P(A | \text{unuseful}) = 1.32 \times 10^{-6}$ 则该推荐链作为需要的推荐链。

由于有些推荐链的长度是小于 5 的，例如推荐链长度为 3，并且属性值的元组为 $\langle \text{honesty}_1, \text{honesty}_2, \text{honesty}_3 \rangle = \langle B, B, A \rangle$ ，这里把缺少的节点的诚信值级别默认为 *A*，则此推荐链的属性值元组为 $\langle B, B, A, A, A \rangle$ ，由于在实际生活中，推荐链的长度越长，则推荐的可靠性越低，所以将缺少的节点的诚信级别默认为 *A* 认为是合理的。

这里把用在全部事件的基础上某事件出现的比例作为估计概率。当使用 $\frac{x}{n}$ 估计 $P(\text{honesty}_3 = B | u_j = \text{useful})$ 时，其中 n 是训练集中 $u_j = \text{useful}$ 的次数， x 是其中 $\text{honesty}_3 = B$ 的次数。如果 $n = 1000, P(\text{honesty}_3 = B | u_j = \text{useful}) = 0.00002$ ，那么 x 只可能是 0，导致 $P(\text{honesty}_3 = B | u_j = \text{useful})$ 概率值过低。或者，如果当概率估计为 0，那么 $u_{\text{target}} = \arg \max_{u_j \in U} P(u_j) \prod_i P(\text{honesty}_i | u_j)$ 中如果存在 $u_j = \text{useful}$ 的条件下 $\text{honesty}_3 = B$ ，那么其他项都将乘以 0，结果将有很大的偏差。

为解决该问题这里采用估计概率的一种贝叶斯方法， m -估计： $\frac{x + mp}{n + m}$ ，若 $m = 0$ 即 $\frac{x}{n}$ 。选择 p

的一种典型的方法是假定均匀的先验概率，因此令 $p = 1/5$ ，而 m 是一个常量。

该分类方法的优势在于给定一个训练数据集之后，一个推荐链是 *useful* 或者 *unuseful* 就是确定的。由于一共只有 3 125 种推荐链，根据上述方法计算这 3 125 个推荐链的 u_{target} 值存入推荐链 u_{target} 表格中，推荐链 u_{target} 表格格式如图 5 所示。

honesty_1	honesty_2	honesty_3	honesty_4	honesty_5	u_{target}
--------------------	--------------------	--------------------	--------------------	--------------------	---------------------

图 5 推荐链 u_{target} 表格

在给定一条推荐链的诚信值级别元组之后可直接从统计表格中查找得到该链是 *useful* 或者 *unuseful*，再进行信任和信誉的计算判断是否与之交互，并且将交互结果记录到历史数据集中。在固定时间段之后（比如一个月后）再重新计算这 3 125 种推荐链 u_{target} 值，更新推荐链 u_{target} 表格。如此往复，推荐链 u_{target} 表格将越来越接近最真实的情况。

5 信任的聚合

5.1 推荐信任的计算

得到所需的推荐链之后，需要计算经过每条推荐链传播之后的信任值。每一层推荐节点提供的推荐信息所占的比重应该是不同的。通过每层节点提供推荐信息的信息增益所占的比重作为节点推荐信任的权重。

定义 5 熵函数

$$H(p) = -p \lg p - (1-p) \lg (1-p) \quad (14)$$

定义 6 第 i 个推荐节点推荐信息的信息增益 $G(S, i)$ 计算如下

$$G(S, i) = H(S) - \sum_{z \in \{A, B, C, D, E\}} \frac{|S_z|}{|S|} H(S_z) \quad (15)$$

其中， S 表示训练数据集， $H(S) = H(u_{\text{target}} = \text{useful}) = H(u_{\text{target}} = \text{unuseful})$ ， S_z 表示在训练数据集中第 i 个节点的诚信值为某个固定值的集合。 $|S|$ 和 $|S_z|$ 表示集合 S 和 S_z 中元素的个数。 $H(S_z)$ 表示诚信值为某个固定值的集合中占 *useful* 和 *unuseful* 比

例的熵。由于信息增益作为概率距离的度量方法，此处作为用来计算推荐节点的权重能够解决信任在推荐链中衰减的问题。例如，求第 2 个推荐节点推荐信息的信息增益，假设训练样例有 100 个， $u_{\text{target}} = \text{useful}$ 个数为 80， $u_{\text{target}} = \text{unuseful}$ 个数为 20，在 100 个样例中 $u_{\text{target}} = \text{useful}$ 条件下 $\text{honesty}_2 = A$ 有 35 个， $u_{\text{target}} = \text{unuseful}$ 条件下 $\text{honesty}_2 = A$ 有 1 个，表示为 $\text{honesty}_2 = A \leftarrow [35\text{useful}, 1\text{unuseful}]$ ，同理，依次有

$$\text{honesty}_2 = B \leftarrow [20\text{useful}, 3\text{unuseful}]$$

$$\text{honesty}_2 = C \leftarrow [15\text{useful}, 4\text{unuseful}]$$

$$\text{honesty}_2 = D \leftarrow [8\text{useful}, 6\text{unuseful}]$$

$$\text{honesty}_2 = E \leftarrow [2\text{useful}, 6\text{unuseful}]$$

得到

$$G(S, 2) = H(0.8) - \frac{36}{100} H\left(\frac{35}{36}\right) - \frac{23}{100} H\left(\frac{20}{23}\right) - \frac{19}{100} H\left(\frac{15}{19}\right) - \frac{14}{100} H\left(\frac{8}{14}\right) - \frac{8}{100} H\left(\frac{2}{8}\right) = 0.183$$

由此，经过归一化可以得到推荐节点的权重 G_i

$$G_i = \frac{G(S, i)}{\sum_{i=1}^5 G(S, i)} \quad (16)$$

由于训练数据集是固定的，所以推荐节点推荐信息的信息增益是固定的，并不会增加信任传播算法额外的复杂度。信任在推荐链中传播的计算方法借用文献[13]中的方法，并为它添加信息增益这个权重。这里以图 4 为例有

$$R_{C_{i-1} \leftarrow C_i}^B = W_{C_i}^a D_{C_i \rightarrow B} + (1 - W_{C_i}^a) \text{honesty}_{C_i \rightarrow C_{i+1}} R_{C_i \leftarrow C_{i+1}}^B$$

添加信息增益作为权重得到

$$RG_{C_{i-1} \leftarrow C_i}^B = G_i R_{C_i \leftarrow C_{i+1}}^B \quad (17)$$

其中， B 表示服务提供者。由此方法可以得到每条推荐链的信任值 $RL_j = RG_{A \leftarrow C_j}^B$ 。

5.2 直接信任与推荐信任的聚合

由以上推荐链分类获取有用推荐链的方法，还需要恰当的方法对各条推荐链的信任值进行合并。根据各条推荐链的诚信度分配合并的权重，需要如下 3 个参数的定义。

1) 推荐链平均推荐信任值 λ_j

$$\lambda_j = \frac{\sum_{i=1}^{h_j} \text{honesty}_i}{h_j} \quad (18)$$

其中， h_i 为第 j 条推荐链的长度。

2) 经过归一化得到第 j 条推荐链的权重

$$W_j = \frac{\lambda_j}{\sum_{j=1}^m \lambda_j} \quad (19)$$

3) 最终推荐信任值

$$R^{ob} = \sum_{j=1}^m W_j RL_j \quad (20)$$

当基于节点的相关信任值做出是否交互的决定时，这里根据节点的信誉值的大小来判断。在现实生活中， X 不了解 Y ，但 Y 有好的信誉，所以 X 愿意与 Y 交互。然而还有一种情况，虽然 Y 的信誉值不高，但 X 还是愿意与 Y 交互，原因就在于亲密的关系和很好的个人经验。所以当申请服务的节点 A 对于提供服务的节点 B 交互次数 n 大于 α 并且节点 A 对于提供服务的节点 B 的直接信任值 T_A^B 大于阈值 β 时，就不需要复杂的推荐信任值的计算。

综上可以得到信誉值为

$$RT_B^{C_i} = \begin{cases} D_{C_i \rightarrow B}, & \text{与 } B \text{ 交互次数大于 } \alpha \text{ 且 } D_{ob} > \beta \\ \varepsilon D_{C_i \rightarrow B} + (1 - \varepsilon) R^B, & \text{其他} \end{cases} \quad (21)$$

其中， ε 是系数，很多文献都是主观地选取一个 $\varepsilon \geq 0.5$ ，这并不科学。比如选取的 ε 较大，直接信任值很小并且推荐信任值较大，导致最后的信誉值很小，这是不妥当的。如果引入熵的概念，就能很好地解决这个问题。例如与节点 B 直接交互的成功率很高，那么 ε 取值应该偏大，因为交互成功的确定性相对较高，又若与节点 B 直接交互的成功率是 0.5，那么 ε 取值应该偏小，因为交互成功的确定性是最小的。

定义 7 节点 B 的成功率如下

$$P_s^B = \frac{N_s^B}{M^B} \quad (22)$$

其中， M^B 表示节点 B 交互的总次数， N_s^B 表示成功的次数。

定义 8 推荐链的成功使用率 P_s^L ，即某类推荐链在被使用后并且最后促使节点与之交互的成功率

表 1 假定的历史数据

诚信值	$honesty_1$	$honesty_2$	$honesty_3$	$honesty_4$	$honesty_5$	描述
A	400	420	290	200	280	
B	300	270	260	450	320	
C	80	80	150	100	130	使用该推荐链并促使节点交互成功的推荐链
D	15	25	90	38	50	
E	5	5	10	12	20	
A	8	5	3	2	8	
B	12	13	6	22	16	
C	25	24	38	60	60	使用该推荐链并促使节点交互失败的推荐链
D	75	78	64	40	70	
E	80	80	89	76	46	

$$P_s^L = \frac{N_s^L}{M^L} \quad (23)$$

其中, M^L 表示该类推荐链被使用并且最后促使节点与之交互的总次数, N_s^L 表示其中成功的次数。

定义 9 确定度 AS

$$AS = 1 - H(P_s) \quad (24)$$

故节点 B 的确定度 $AS^B = 1 - H(P_s^B)$, 推荐链 L 的确定度 $AS^L = 1 - H(P_s^L)$ 。

由以上定义可得到最终的参数 ε

$$\varepsilon = \frac{AS^B}{AS^B + (\sum_{L \in C} AS^L) / |C|} \quad (25)$$

其中, C 表示申请服务节点 A 选用的所有推荐链的集合, $|C|$ 表示集合 C 中元素的个数。该参数的设置能很好地适应节点间动态变化的信任关系。

6 仿真实验与结果分析

实验主要从抵御恶意节点更快速、在相同的恶意节点比例下有更高的交易成功率、能有效识别多种恶意节点这些方面体现本文模型的有效性。本文以文献[14]为参考进行仿真实验, 为方便表示将文献[14]中模型简记为 trust 模型。

为了实现本文中推荐链的分类, 根据现实中的经验假定如下的 1 000 组推荐链历史数据集, 见表 1, 其中推荐成功的有 800 组, 推荐失败的有 200 组, 表中各数值表示每层推荐节点取各个值的数量。

假设节点 A 需要申请节点 B 的服务, 其中存在 10 条推荐链。根据表 1, 通过推荐链分类方法得到表 2。

表 2 推荐链分类

$honesty_1$	$honesty_2$	$honesty_3$	$honesty_4$	$honesty_5$	u_{target}
A	C	C	C	C	useful
B	D	E	C	A	unuseful
B	C	C	C	C	useful
C	B	C	A	C	useful
D	A	E	D	A	unuseful
B	D	C	E	A	unuseful
B	B	A	C	B	useful
C	C	B	A	D	useful
D	C	C	B	D	unuseful
D	D	C	C	A	unuseful

通过表 1 得到 $G(S,1) = 0.423$, $G(S,2) = 0.416$, $G(S,3) = 0.322$, $G(S,4) = 0.296$, $G(S,5) = 0.229$ 。从而 $G_1 = 0.25$, $G_2 = 0.25$, $G_3 = 0.19$, $G_4 = 0.18$, $G_5 = 0.14$, 这里也能看出推荐信息随着推荐链层数的增加而衰减。

表 3 仿真参数设置

参数	描述	参数值
y_1	评价函数: 非常满意	1
y_2	评价函数: 较为满意	0.8
y_3	评价函数: 一般满意	0.6
y_4	评价函数: 不满意	0
α	时间衰减函数参数	0.9
θ	诚信函数阈值	0.2
∂	活跃度函数参数	1.1
v_1	诚信值阈值	0.2
v_2	诚信值阈值	0.4
v_3	诚信值阈值	0.6
v_4	诚信值阈值	0.8
m	m -估计中 m 的设置	5

实验 1 为了观察使用本文模型后节点的信誉

值随交互次数不断增大的变化，实验将每个节点的初始信誉值赋为 0.5。为了便于观察各类节点信誉值的变化，做了同文献[14]中相同的处理，将节点分为 4 类：绝对信任即从不欺骗、一般信任即有时欺骗比绝对信任的信誉值稍差、临界信任即时好时坏、不可信任即每次交互都存在欺骗。图 6 中各类节点信誉值的变化趋势与使用文献[14]中模型的效果几乎相同，绝对信任的节点的信誉值随着交互次数的增加呈直线增加，一般信任的节点的信誉值也随着交互次数的增加而增加，但显然斜率小于绝对信任节点信誉值的变化趋势，临界信任的节点信誉值则在 0.5 附近起伏不定，不可信任的节点的信誉值则呈直线下降。

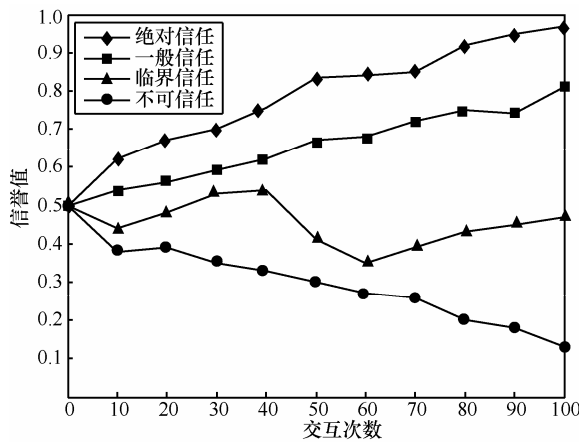


图 6 信誉值随交互次数的变化

实验 2 主要是为了体现出本文推荐链分类方法带来的能更快速地抵御恶意节点和更高的交易成功率的优势。文献[14]虽然也提出了推荐链的选择方法，但该方法主要还是依靠信任值与阈值的比较，关于阈值的设定依然是个主观的静态的过程。通过本实验可以看到本文模型较文献[14]的优势。

设定相同恶意节点的比例 50%，通过图 7 观察交易成功率随着时间的变化情况。本文模型的交易成功率在第 4 个时间段就已经达到 0.9，显然较 trust 模型有更快的速度。并且最后趋于 0.96 高于 trust 模型的 0.94。由于本文使用推荐链精确分类方法，并且本文中直接信任与推荐信任的聚合参数不会受到特殊节点以及节点交互动态变化的影响，因而本文模型体现出了更好的抗恶意节点的能力。

实验 3 在出现各种不同类型的恶意节点并且恶意节点的比例不断增大时，对比本文模型与

trust 模型的抵抗恶意节点的能力。图 8 中的恶意节点只是针对个别节点不提供真实服务。图 8 中虽然 2 种模型的交易成功率都随着恶意节点的比例增大而不断减小，但显然本文模型具有更好的抵抗此类恶意节点的能力。本文中的推荐链分类是基于两点之间的信任，所以针对此类恶意节点很容易识别。

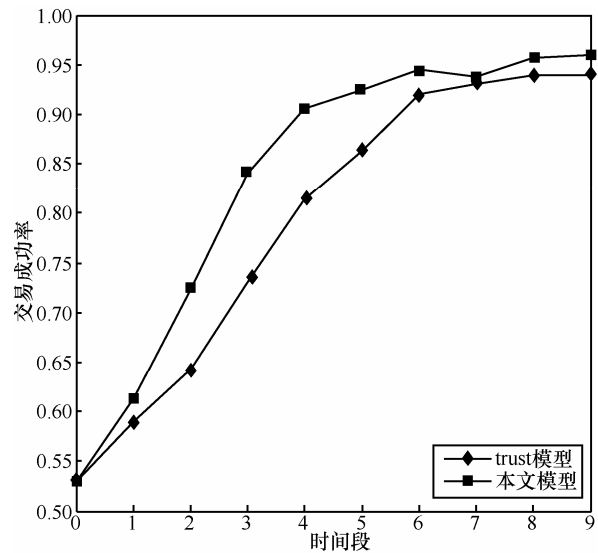


图 7 交易成功率随时间的变化

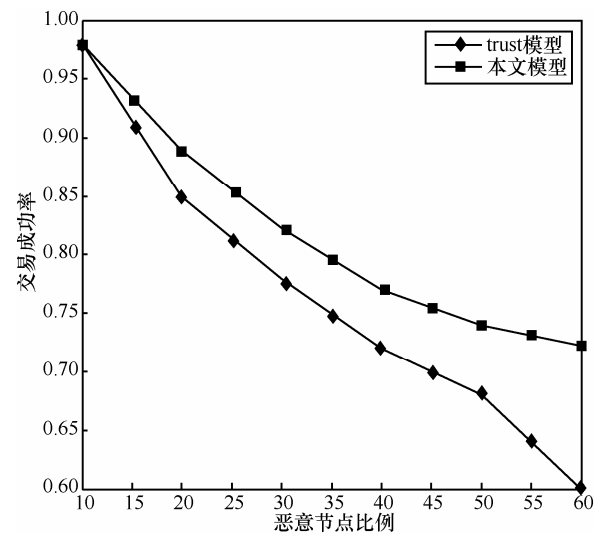


图 8 交易成功率随第一类恶意节点比例的变化

图 9 中的恶意节点是随机提供真实服务，即以 50% 的概率给其他节点提供真实服务。此类恶意节点较难处理，图 9 中 trust 模型和本文模型中交易成功率几乎都随着恶意节点比例增加成直线下降。但图 9 中本文模型较 trust 模型依然存在优势。再一次说明了本文模型中动态变化的聚合参数以及

推荐链分类方法针对复杂多变的网络环境有更强的优势。

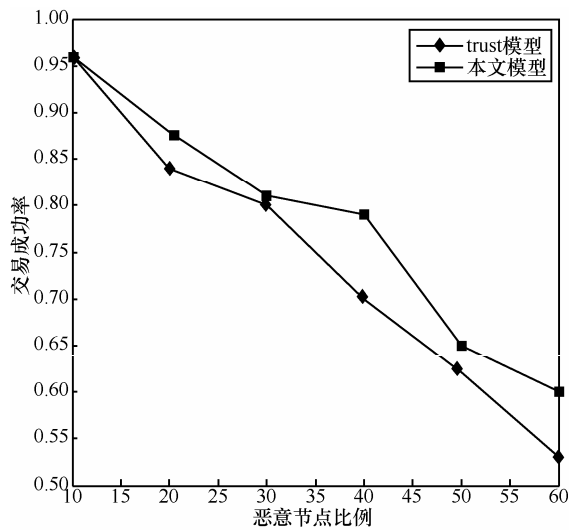


图9 交易成功率随第二类恶意节点比例的变化

实验4 信任计算的主要开销与信任计算过程中的考虑的节点数成正比。因此实验将本文模型和trust模型放在相同环境下并且交互成功率都在90%时，随着网络规模的变化进行信任计算开销的对比，其中信任计算开销即计算过程中需要的节点数。由图10可知，当网络规模不大时，2种模型的计算开销相差不大，但是随着网络规模的不断扩大信任计算开销的差距就显而易见了。文献[14]虽然也利用了六度分离理论将推荐链的长度大大缩减了，但在推荐链的选择上还是利用了与阈值的比较，这样并没有对推荐链做更进一步的细分，最终导致计算了很多不必要的节点的信任信息，这也是本文模型使用推荐链分类后的一个优势。

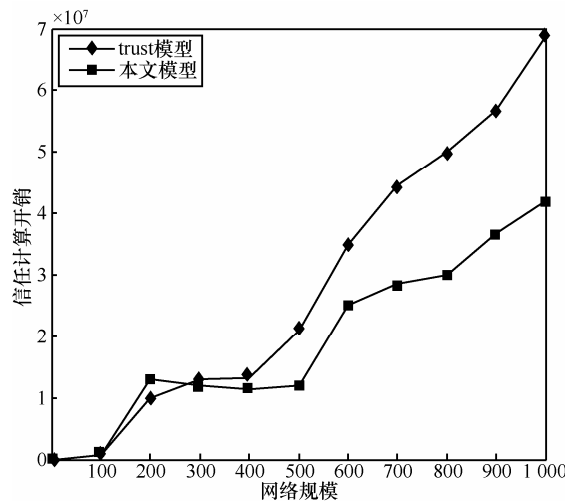


图10 信任计算开销随网络规模的变化

7 结束语

本文提出一种基于推荐链分类的动态信任模型，根据节点间的诚信情况利用概率论的知识将推荐链进行分类。从仿真结果可以看出使用这种分类方法比以往从主观上选择或简单地利用阈值的判断选择可信的推荐链来获取推荐信息要更准确。推荐信任传播使用了信息增益的原理，很好地说明了信任在推荐链中的衰减的问题更加贴近现实。在直接信任与推荐信任的聚合过程中引入熵的概念，根据历史经验数据获得的参数更准确，并且能随着交互的进行不断地修正，比以往主观模糊的参数设置有相当大的优势。本文中尚未考虑节点间的依赖关系，这也是下一步要继续研究的问题。

参考文献:

- [1] KAMAL K, BHARADWAJ, MOHAMMAD YAHYA H, AL-SHAMRI. Fuzzy computational models for trust and reputation systems[J]. Electronic Commerce Research and Applications, 2009, 8:37-47.
- [2] XIA H, JIA Z P, EDWIN H M S. Research of trust model based on fuzzy theory in mobile ad hoc networks[J]. IET Inf Secur, 2014,8: 88-103.
- [3] MIESO K D, SUN T, ISAAC W. Trust management in ubiquitous computing: a Bayesian approach[J]. Computer Communications, 2011, 34: 398-406.
- [4] CARMEN F G, ISAAC A, JAVIER L. Building trust from context similarity measures[J]. Computer Standards & Interfaces, 2014, 36: 792-800.
- [5] 王江涛, 陈志刚, 邓晓衡. 普通计算中一种上下文感知的自适应可信模型[J]. 通信学报, 2012, 33(11): 41-56.
WANG J T, CHEN Z G, DENG X H. Context-aware and self-adaptive trust model for pervasive computing[J]. Journal on Communications, 2012, 33(11): 41-56.
- [6] 秦艳琳, 吴晓平, 高键鑫. 分布式环境下信任路径选择性搜索及聚合研究[J]. 通信学报, 2012, 33(Z1): 148-156.
QIN Y L, WU X P, GAO J X. Research on selective trust-path search and aggregation in distributed environment[J]. Journal on Communications, 2012, 33(Z1): 148-156.
- [7] 李致远, 王汝传. P2P 电子商务环境下的动态安全信任管理模型[J]. 通信学报, 2011, 32(3): 50-59.
LI Z Y, WANG R C. Dynamic secure trust management model for P2P e-commerce environments[J]. Journal on Communications, 2011, 32(3): 50-59.
- [8] 顾鑫, 徐正全, 刘进. 基于云理论的可信研究及展望[J]. 通信学报, 2011, 32(7): 176-181.
GU X, XU Z Q, LIU J. Review of cloud based trust model[J]. Journal on Communications, 2011, 32(7): 176-181.
- [9] LI X Y, DU J P. Adaptive and attribute-based trust model for ser-

vicelevel agreement guarantee in cloud computing[J]. IET Inf Secur, 2013,7:39-50.

- [10] KAWSER W N, TONNY S K, AMJAD H, *et al.* A fuzzy and probabilistic logic based representational model of certain trust model[A]. International Conference on Informatics, Electronics & Vision[C].2012.1000-1005.
- [11] LIU G F, WANG Y, MEHMET A O, *et al.* Finding the optimal social trust path for the selection of trustworthy service providers in complex social networks[J]. Transactions on Services Computing, 2013, 6(2):152-167.
- [12] MARK H, SYED W J, PETER P V M, *et al.* Design and validation of a relative trust model[J]. Knowledge-Based Systems,2014,57:81-94.
- [13] 张琳, 王汝传, 王海艳. 基于多影响因素的网格信任传播算法[J]. 通信学报,2011,32(7):161-168.
ZHANG L, WANG R C, WANG H Y. Trust transitivity algorithm based on multiple influencing factors for grid environment[J]. Journal on Communications,2011,32(7):161-168.
- [14] 田俊峰, 鲁玉臻, 李宁. 基于推荐的信任链管理模型[J].通信学报, 2011,32(10):1-9.
TIAN J F, LU Y Z, LI N. Trust chain management model based on recommendation[J]. Journal on Communications, 2011,32(10):1-9.
- [15] 蒋黎明, 张琨, 徐建, 等. 证据信任模型中的信任传递与聚合研究[J].通信学报,2011,32(8):92-100.
JIANG L M, ZHANG K, XU J, *et al.* Research on trust transitivity and aggregation in evidential trust model[J]. Journal on Communications, 2011,32(8):92-100.
- [16] MAGDALINI E, MALAMATI D L, IRAKLIS V. A trust-aware system for personalized user recommendations in social networks[J]. IEEE Transactions on Systems, Man, and Cybernetics: System, 2014, 44(4):409-421.

作者简介:



张琳 (1980-), 女, 江苏丰县人, 南京邮电大学副教授、硕士生导师, 主要研究方向为分布式计算、网络安全、信任、可信计算等。



邢欢 (1991-), 男, 江苏启东人, 南京邮电大学硕士生, 主要研究方向为分布式计算、信息安全、信任、可信计算等。



王汝传 (1943-), 男, 安徽合肥人, 南京邮电大学教授、博士生导师, 主要研究方向为计算机软件、计算机网络和网络、信息安全、无线传感器网络、移动代理和虚拟现实技术等。

吴超杰 (1993-), 男, 江苏无锡人, 主要研究方向为信任计算、可信计算。