

P2P网络中基于动态推荐的信任模型

张景安, 郭显娥

(山西大同大学数学与计算机科学学院, 大同 037009)

摘要: 针对P2P网络中节点交易风险较大的问题, 模拟社会网络的人际交互过程, 提出一种基于动态推荐的信任管理模型。采用模糊聚类方法, 结合交互的上下文动态地选择推荐节点, 在推荐因子的计算上融入聚类分析结果, 提高了推荐的可靠性。分析和模拟实验表明, 该模型能有效提高推荐的准确性, 增强P2P网络的可用性。

关键词: P2P网络; 动态推荐; 信任模型; 模糊聚类

Trust Model Based on Dynamic Recommendation in P2P Network

ZHANG Jing-an, GUO Xian-e

(School of Mathematics and Computer Science, Shanxi Datong University, Datong 037009)

【Abstract】 To resolve the risk problem of transaction in P2P network, by simulating interpersonal interactive process in society network, a trust management model based on dynamic recommendation is proposed. Based on fuzzy cluster algorithms and the interactive context, recommendation nodes are selected dynamically. According to results of cluster analysis, recommendation factors are computed and the reliability of recommendation is enhanced. Analysis and simulation experiments show that the method can improve the accuracy of recommendation effectively and strengthen the availability of P2P network.

【Key words】 P2P network; dynamic recommendation; trust model; fuzzy cluster

1 概述

随着 Internet 的迅猛发展及网上应用的日益增多, P2P 网络以其灵活、便捷的工作特性逐渐成为计算机领域的研究热点。但由于 P2P 网络开放性、匿名性和动态性的本质, 因此节点间交易的风险较大。在 P2P 网络中, 节点参与网络是随机、自愿的, 每个节点可以随意中止服务, 而不同节点的能力和可靠性也不相同, 因此, 网络中存在大量的欺诈行为及不可靠的服务质量, 大大降低了网络的可用性。在 P2P 网络中建立有效的信任管理机制, 对 P2P 网络的健康发展十分必要。现有典型的信任管理机制主要有 2 类:

(1) 基于凭证的信任管理。典型的解决方案包括: Policy Maker 模型, KeyNote 模型等。基于凭证的信任管理技术提供标准、通用的机制建立安全策略, 通过对信任客体颁发信任凭证建立信任关系, 其可靠性通过访问控制信息的验证来保障。基于凭证的信任管理系统本质上使用了一种精确、静态的方式描述和处理复杂、动态的信任关系, 这种方法显然并不适合处理 P2P 网络中动态的信任关系。

(2) 基于信誉的信任管理。信誉一般定义为网络中的源实体对特定目标实体历史行为表现的评价。典型的解决方案包括 Beth 模型、Josang 模型和 Abdul-Rahman 模型等。为了度量实体的可信性, 需要根据直接交互经验(直接信任)或者其他实体的反馈(推荐信任)进行信任推导计算。基于信誉的信任管理主要基于信任的主观性及可度量性, 其信任关系类似于人际交互网络, 并可能受到多种因素的影响, 比较适合处理 P2P 网络中动态的信任关系。本文根据不同的应用需求, 借助模糊理论的聚类分析算法, 基于动态推荐建立节点间的信任关系。

2 基于动态推荐的信任模型

2.1 直接信任与推荐信任

一些成功的案例(如淘宝网)和已有的研究显示, 建立有效的信任管理模型能够降低风险, 提高交互的成功率。通常对目标节点从两方面进行信任评估: (1) 节点可依据自身的历史经验和观察, 通过总结对目标节点过去的直接交互经验, 对节点未来的可能行为作出判断, 以此建立相应的信任关系, 这种方式称为直接信任。(2) 通过目标节点与其他节点进行交互所积累的信誉值判断目标节点未来可能的行为特征, 以此建立信任关系, 这种方式称为推荐信任。将直接信任与推荐信任按某种方式综合求得如式(1)所示的节点最终信任度, 依据信任度的大小决定源节点与目标节点交互的行为特征。

$$T_{ij} = \beta DT_{ij} + (1 - \beta) RT_{ij} \quad (1)$$

其中, T_{ij} 表示节点 i 对节点 j 的最终信任度; DT_{ij} 表示节点 i 对节点 j 的直接信任度; RT_{ij} 表示节点 i 对节点 j 的推荐信任度; $\beta (0 < \beta < 1)$ 为权值, 表示节点的自信程度。

按照社会网络人际交互的特征, 在对目标节点进行信任评价时, 如果源节点有多次与目标节点的直接交互经验(特别是近期), 则很容易建立起与目标节点的信任关系。但实际上往往存在 2 种情况: (1) P2P 网络规模很大, 节点间重复多次交互的可能性较低或下一次交互距上一次交互的时间间隔太长, 这样源节点在对目标节点进行信任评估时会产生不自信,

基金项目: 山西省教育厅高科技开发基金资助项目(20051256)

作者简介: 张景安(1970—), 男, 副教授、硕士, 主研方向: 网络安全, 计算机网络; 郭显娥, 副教授

收稿日期: 2009-06-11 **E-mail:** zhangjingan@126.com

从而希望得到更多的证据；(2)源节点需要与陌生的目标节点进行交互，根本没有交互的历史经验可查，这时可信第三方的推荐就尤为重要。因此，在P2P网络中，获取可靠的推荐信任非常重要，许多学者也对这一问题进行了专门研究^[1-2]。

2.2 模型的建立

由上文可知，P2P网络中节点最终信任度的计算通过直接信任和推荐信任的有效综合完成，为了专注于本文的研究重点，只对直接信任的获取作简单的讨论。

2.2.1 直接信任评价

在P2P网络环境中，主体的信任度是由多方面因素决定的，因此，在确定各主体的信任向量时必须考虑多方面的因素^[3]，借助文献^[4]的研究成果，通过引进模糊集合论的思想实现对信任的动态建模，给出P2P网络中评估对象的定义。

定义1 P2P网络评估对象(节点)定义为一个多元组 $O=(Ba, Pr, Dc, Re, \dots)$ 。其中， Ba 为节点的有效带宽向量； Pr 为节点提供的资源数向量； Dc 为节点的数据处理能力向量； Re 为节点的可靠性向量……，其中，属性向量的多少根据实际应用需求动态地确定。

然后借助模糊子集和隶属度的概念，可以对信任进行动态分组描述，设 $L=\{l_1, l_2, l_3, \dots\}$ 为信任等级集合， l_i 为第 i 个信任等级，信任等级的多少可依据评价的粒度粗细动态确定，为便于理解，可采用自然语言对 l_i 进行定性命名。例如，用 l_1 表示非常可信， l_2 表示可信， l_3 表示一般可信， l_4 表示不可信。在实际进行评价时，可以通过节点各属性对模糊集合 L 的隶属度向量来表述评价的结果。假设某节点 Y 要对节点 X 进行直接信任评估，节点 $X(X \in O)$ 共有 n 个属性，信任等级集合有 m 个元素，若第 i 个属性的评价结果为： $T_i=\{t_{i1}, t_{i2}, \dots, t_{ij}, \dots, t_{im}\}$ ，其中， t_{ij} 表示节点第 i 个属性对第 j 个信任等级的隶属度，则所有属性的评价结果构成一个 $n \times m$ 的矩阵 T_{nm} 。最后，依据模糊综合评判的方法得出节点 X 的最终信任向量 $DT_{yx}=(t_1, t_2, \dots, t_n)$ ，再通过等级量化向量 $F=(f_1, f_2, \dots, f_n)$ 按式(2)进行反模糊化计算，得出最终的量化评价结果。

$$DT_{yx}^{final} = \sum_{i=1}^n t_i f_i \quad (2)$$

2.2.2 推荐信任评价

本文对推荐信任的研究首先基于以下假设：在P2P网络中，当源节点发出推荐请求时，响应请求的推荐路径有 n 条 ($n \geq 2$)，见图1。图中， $Re_i (i=1, 2, \dots, n)$ 为推荐节点，源节点 X 与 Re_i 具有直接信任关系，目标节点 Y 与 Re_i 有直接或推荐信任关系(这时可能有路径传递衰减)。

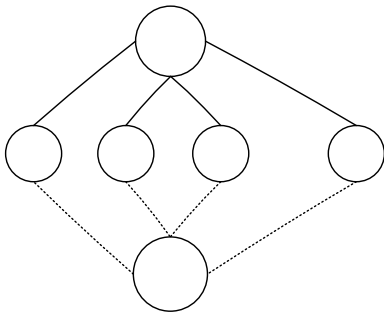


图1 推荐信任示意图

由于P2P网络规模巨大，因此这种假设在实际应用中很容易实现；否则，如果 $n=0$ 或 1 ，对推荐信任再展开研究就失去了意义。节点加入P2P网络的时间越长，它可能获得的

推荐路径就越多。当源节点获得多个目标节点的推荐信任后，一般的方法是采用式(3)进行加权综合，以求得最终的信任向量，然后采用式(2)进行量化，得出最终的推荐信任值：

$$RT_{yx} = \frac{\sum_{k=1}^m RT_{kx} \times w_{kx}}{\sum_{k=1}^m w_{kx}} \quad (3)$$

其中， RT_{yx} 表示源节点 Y 获得的目标节点 X 的推荐信任向量； RT_{kx} 表示第 k 条路径的推荐信任向量； $w_{kx} (0 < w_{kx} < 1)$ 为第 k 条路径的推荐因子。

基于这种方式将所有的推荐都进行合成会产生可靠性的问题。在众多的推荐节点中，很难保证每一个节点都是可靠的，而且如果存在大量的推荐节点，对每一个推荐都进行合成是否必要也需进一步研究。实际上在社会网络的人际交互中，当需要对一个不了解或陌生的人建立信任时，并不是向所有与之打过交道的人都进行询问，而是根据交互的重要性，向不同可信度的人进行咨询(如果交互的事情很重要，则人们更愿意相信可信度高的人的推荐，通常信誉高的人，其推荐也更可靠)，再综合判断得出结论。基于这种思想，文献^[5]虽然提出了最强信任路径的思想，但没有考虑交互的动态性，为此，本文提出了基于模糊聚类分析的动态推荐方法。

本方法的基本思想是，当源节点需要获取目标节点的推荐信任时，首先在网络中广播推荐请求；在一定的周期内，当收到推荐节点返回的应答信息后，源节点统计推荐节点的数量，如果节点数量满足需求或所有直接信任的节点均返回了推荐应答，则停止广播推荐请求；然后，源节点依据自己直接交互所获得的推荐节点信任向量的相关属性，采用模糊聚类的方法对推荐节点进行动态分类；最后，根据交互的重要性，动态地选择分类中的部分可信节点作为推荐节点进行推荐信任综合。具体步骤如下：

(1) 标定与数据标准化

定义2 设论域 $O=\{X_1, X_2, \dots, X_n\}$ 是待分类的推荐节点全体，每一个节点具有 m 个属性，则其直接信任向量可用多元组 $(x_{i1}, x_{i2}, \dots, x_{im})$ 表征，这个过程称为标定。

标定完成后可得到一个 $n \times m$ 的原始资料矩阵 $S_{n \times m}=(x_{ij})_{n \times m}$ 。为了使具有不同量纲的数据能进行比较，需要对数据作适当的变换，并根据模糊矩阵的要求，将数据压缩到区间 $[0, 1]$ 上，这个过程称为数据标准化，经过数据标准化的矩阵为 $S'_{n \times m}=(x'_{ij})_{n \times m}$ 。

(2) 建立模糊相似矩阵

定义3 设论域 $O=\{X_1, X_2, \dots, X_n\}$ ，其中， $X_i=(x_{i1}, x_{i2}, \dots, x_{im})$ ， $X_j=(x_{j1}, x_{j2}, \dots, x_{jm})$ ，用 r_{ij} 表示对象 x_i 与 x_j 间的相似度，称为相似系数。

当采用距离法时，相似系数 $r_{ij}=1-ad(x_i, x_j)$ ，其中， a 为适当选取的参数，它使得 $0 \leq r_{ij} \leq 1$ ，而 $d(x_i, x_j)$ 可以是海明距离或欧氏距离等，最终可求得模糊相似矩阵 $R_{n \times n}=(r_{ij})_{n \times n}$ 。

(3) 求解模糊相似矩阵 R 的传递闭包 $t(R)$

若模糊相似矩阵 R 不是模糊等价矩阵，可通过逐次平方方法求得传递闭包，具体算法为：从模糊相似矩阵 R 出发，依次求平方： $R \rightarrow R^2 \rightarrow R^4 \rightarrow \dots$ ，当第1次出现 $R^k \circ R^k = R^k$ 时， R^k 就是所求的传递闭包 $t(R)$ 。

(4) 聚类分析

选取合适阈值 $\lambda \in [0, 1]$ ，按照过程(3)中建立的模糊等价矩阵 $t(R)$ ，求得 λ 截矩阵 R_λ 便可得到动态聚类，即将各推荐

节点按信任相似度进行模糊分类。

(5) 确定推荐节点

在实际应用时,可依据交互类型的不同,动态地选择分类阈值 λ ,从而得到满足应用需求的分类结果。求得分类结果后,可通过特征向量属性值的加权和均值(式(4)和式(5))确定不同类群所属的信任类型(如前所述,可用语言变量刻画)。

$$p_i = \sum_{k=1}^m \omega_k x_{ik} \quad (4)$$

其中, $i \in [1, n], \omega_k \in [0, 1]$ 为权值,且 $\sum_{k=1}^m \omega_k = 1, x_{ik}$ 表示特征向量的属性值;假设共分 c 个类群,某一类群的主体有 h_c 个,则:

$$\bar{p}_x = \frac{1}{h_c} \sum_{k=1}^{h_c} p_k \quad (5)$$

其中, $x=1, 2, \dots, c$, 然后依据 \bar{p}_x 的值决定类群所属的信任类型。最后选择合适类群中的节点作为最终的推荐节点。

(6) 确定推荐因子

确定推荐节点后,假设某类推荐节点的特征向量属性值的加权和均值为 \bar{p}_k , 则其推荐因子可按式(6)计算。

$$w_{kx} = \gamma(t, n) \bar{p}_k \quad (6)$$

其中, w_{kx} 表示第 k 个节点的推荐因子; $\gamma(t, n)$ 表示推荐传递衰减因子(它是时间和路径长度的函数)。最后,按式(3)进行推荐信任合成。使用式(1)进行最终信任的综合,通过量化的结果决定最终的交互类型。

3 实验仿真与分析

3.1 实验 1

实验 1 主要验证本模型的可行性与合理性。假设某 P2P 节点 A 要与另一节点 B 进行网上交易,设在时间 t 内共获得了 5 个推荐节点(X_1, X_2, \dots, X_5)对 B 的推荐值,每个推荐节点用 3 个属性(Ba, Re, Rt)刻画其主体信任特征(其中, Ba 表示节点的有效带宽; Re 表示节点的可靠性; Rt 表示节点的历史推荐信誉度),通过数据采集与综合,得到原始资料矩阵:

$$S_{5 \times 3} = \begin{bmatrix} 0.9 & 0.9 & 0.8 \\ 0.6 & 0.5 & 0.6 \\ 0.8 & 1 & 0.9 \\ 0.3 & 0.5 & 0.4 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$$

采用海明距离法(取 $\alpha=0.5$),按上述步骤得到动态聚类图见图 2。

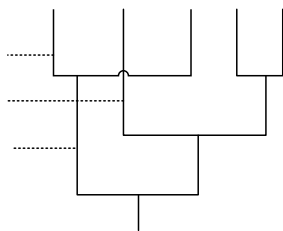


图 2 动态聚类图

取不同的 λ , 分类结果不同,这时要按照实际情况,选取合适的分类,这里取 $\lambda=0.85$,可将推荐节点分为 3 类: $1\{X_1, X_3\}$; $2\{X_2\}$; $3\{X_4, X_5\}$ 。然后,设刻画推荐节点属性的 3 个特征参数的权重集为 $\{0.2, 0.3, 0.5\}$,按上面的分析可求得 3 类推荐节点信任向量值的综合加权和均值分别为: $\bar{p}_1=0.88$, $\bar{p}_2=0.57$, $\bar{p}_3=0.39$,用对应的语言变量来描述,则可认为 3 类推荐节点的可信度分别为非常可信、一般可信、不可信。这时,可根据不同的交互目的动态地选择推荐节点。

如果非常强调推荐的可靠性,则可选择第 1 类节点作为推荐节点;如果更注重推荐的全面性,则可选择第 1 类和第 2 类节点作为推荐节点;第 3 类节点由于可信度低,因此为避免恶意推荐,原则上应舍弃(当然也可事先设定一个推荐节点信任度阈值,低于该阈值的节点首先舍弃,然后进行选择)。最后,即可按上述方法求得各类推荐节点的推荐因子并通过综合求得推荐信任值。

3.2 实验 2

实验 2 主要验证模型动态推荐的准确性。为了简化计算,专注于本文的研究内容,忽略了推荐过程中的传递衰减,将聚类分析后得到的推荐节点信任向量值的综合加权和均值直接作为推荐因子。推荐节点总数为 1000,仍按实验 1 的方式考察节点信任属性,并按其信任度分为如实验 1 的 3 类,为方便计算,规定非常可信的节点占 30%,一般可信节点占 50%,不可信节点占 20%。每次随机选取不同比例的节点作为恶意节点,提供恶意推荐,通过 VC++ 实现本文的算法。实验 2 对比了选择不同可信度的节点作为推荐节点时推荐的准确率。推荐准确率通过公式 $s=1-|T_{推荐}-T_{真实}|$ 计算,其中, $T_{推荐}$ 表示推荐信任值; $T_{真实}$ 表示实际信任值,结果如图 3 所示。由图 3 可见,当选择第 1 类节点作为推荐节点时(第 1 类推荐),推荐的准确率很高,基本保持在 0.95 以上,即使恶意节点比例达到 0.7,其准确率仍在 0.9 以上,显示了较高的推荐可靠性。当选择第 1 类和第 2 类节点作为推荐节点时(第 2 类推荐),推荐准确率有所降低,由于未引进惩罚机制,因此当恶意节点比例较大时(超过 0.5),推荐的准确率下降较快。主要原因是节点可信度越低,其恶意推荐的可能性越大。但当恶意节点的比例小于 0.4 时,该类推荐仍具有较高的可靠性,其准确率在 0.85 以上,反映了在恶意节点比例较小时,该类推荐具有较高的可用性。当选择所有的节点均作为推荐节点时(即不对节点作分类,第 3 类推荐),可以看到其推荐准确率基本上随恶意节点比例的增加呈线性降低,从而体现了信任模型的必要性。

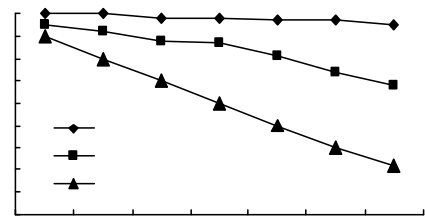


图 3 推荐准确率比较

3.3 分析与讨论

(1)采用动态推荐的方法,小于信任阈值的低可信度节点被事先被排除,而对大于信任阈值的节点,可依据交互的情况动态选择其全部或部分作为推荐节点。从网络的综合表现来看,高信任度的节点具有高的推荐可信度,仿真结果也验证了这一假设。这样,一方面提高了推荐选择的灵活性,另一方面提高了推荐的可靠性。

(2)在对推荐节点分类时,为简化计算,本文仅选择节点的 3 个属性来评价,原则上属性选择得越多,评价越准确,但过多的属性又会增加节点的计算量和网络负担,因此,在实际应用时,要依据网络与节点情况平衡选择最关键的属性。

(下转第 180 页)