

P2P 环境下局部可信度的神经网络识别方法

胡和平, 黄保华, 姚寒冰, 卢正鼎, 李瑞轩

(华中科技大学 计算机科学与技术学院, 湖北 武汉 430074)

E-mail: hphu@andin.net

摘要: P2P (Peer-to-Peer) 环境下对等实体的全局可信度得到了广泛重视和研究, 但计算全局可信度的基础——局部可信度却没有受到应有的重视。现有模型只给出了基于交易成功与失败次数统计比例的简单方法, 不能描述交易成功失败分布特性。首次将神经网络引入局部可信度的识别, 将能够反映分布特性的交易成功与失败序列作为神经网络输入来识别局部可信度, 给出了神经网络结构、输入规范化和训练样本构造方法。通过分析和实验可以看出, 该方法是有效和可行的。

关键词: P2P; 信任; 神经网络

中图分类号: TP393

文献标识码: A

文章编号: 1000-1220(2006)08-1503-03

Identifying Local Trust Value with Neural Network in P2P Environment

HU He-ping, HUANG Bao-hua, YAO Han-bing, LU Zheng-ding, LI Rui-xuan

(College of Computer Science & Technology, Huazhong University of Science & Technology, Wuhan 430074, China)

Abstract: Global trust value of P2P (Peer-to-Peer) has been studied in detail, but the base of it, local trust value, has not been explored in depth. The existent models only adopt simple methods to calculate it. These methods are based on count of success and failure times of transaction, so it cannot represent the distribution of success and failure in transaction history. It is the first time to introduce neural network to identify the local trust value in P2P environment. Transaction result Sequence that can represent the transaction history is used as input of neural network to identify local trust value. The structure of neural network, method of input standardization and training sample constructing are presented. A nalysis and experiment show that it is feasible and effective to identify local trust value with neural network in P2P environment.

Key words: P2P; trust; neural network

1 引言

P2P 与目前盛行的C/S (Client/Server) 模式不同, 它通过直接交换来共享计算机资源和服务^[1]。P2P 中各节点高度分布、高度自治、平等交互, 这一方面有利于增强可靠性和扩展性, 有利于资源聚合并减少费用开销, 使P2P 在存储资源、计算周期、内容共享等方面得到广泛应用^[2], 但另一方面却给P2P 系统的安全带来了严峻的挑战。

信任管理 (Trust management) 是为满足分布式系统安全要求提出的^[3], 经过多年研究, 信任管理途径已经成为解决分布式系统中传统授权机制不足的一个重要方法^[4]。P2P 领域借鉴信任管理的思想, 对P2P 环境下的信任管理进行了大量研究。这些工作大致可分为: (1) 基于PKI的信任模型; (2) 基于局部推荐; (3) 数据签名; (4) 全局可信模型^[5]。在这4类信任模型中, (2)和(4)可以不需要可信第三方和密码技术等条件, 是最符合P2P 基本理念的, 比较便于在P2P 环境部署。(2)和(4)都需要首先计算局部可信度, 而现有的模型对局部可信度的计算是十分简单的, 不能准确反映被描述对象的可信状态。

本文接下来首先分析现有模型局部可信度计算方法的不足并提出神经网络识别的思想, 然后依次给出局部可信度神经网络识别的基本原理和神经网络结构、输入规范化、样本获取与神经网络训练、局部可信度识别以及实验结果, 最后总结并提出进一步研究的方向。

2 相关工作

Poblan^[6] 定义了3个基本量, 即代码数据可信度 (Code Confidence)、对等实体可信度 (Peer Confidence) 和风险评估值 (Risk)。Risk 是通过诸如QoS (Quality of Service) 这样的评估或计算方法得来的。[6]给出了Code Confidence 和 Peer Confidence 的全局传递计算公式, 也给出了对等实体提供或拒绝服务请求的判别公式, 但没有非常明确地给出局部Code Confidence 和 Peer Confidence 的计算方法和公式。EigenRep^[7] 给出的节点 (与对等实体含义相同) i 对节点 j 的局部可信度计算公式为 $C_{ij} = \frac{S_{atij} - UnS_{atij}}{(S_{atij} - UnS_{atij})}$ (式1), S_{atij} 和 UnS_{atij} 分别为节点 i 与 j 在历史交易中积累的满意次数和满意次数。[5]改进了EigenRep, 给出的局部可信度公式为

收稿日期: 2005-05-31 收修改稿日期: 2005-09-02 基金项目: 国家自然科学基金项目 (60403027) 资助 作者简介: 胡和平, 男, 1952年生, 教授, 主要研究方向为软件工程、智能决策系统、信息安全等; 黄保华, 男, 1973年生, 博士研究生, 主要研究方向为P2P 安全与应用等; 姚寒冰, 男, 1976年生, 博士研究生, 主要研究方向为Grid 安全等。

$P_{ij} = \frac{S_{ij}}{I_{ij}}$ (式2), 其中 I_{ij} 为节点 i 与 j 在最近某个固定时间 t 内 (t 随具体应用而定, 例如 1 个月) 实际交互的次数, S_{ij} 为在节点 i 看来交易成功的次数, 如果 $I_{ij} = 0$, 则设 $P_{ij} = 0$

从式 (1) 和 (2) 可以看出, 局部可信度只是交易成功和失败次数统计量的一个比例关系, 不能反映交易序列的一些其它特征, 比如交易成功和失败的时间分布等, 而这些特征对评价可信度是非常必要的

为使可信度全面反映交易的情况, 可以从交易成功与失败的记录序列直接获取可信度, 这可以借鉴模式识别的原理, 利用神经网络 (Neural network) 来实现这样的识别功能^[8]

3 基本原理与神经网络选择

设有交易成功与失败记录序列 $A = (a_1, a_2, \dots, a_n)$, 有一复杂系统 S , 能够以 A 为输入并输出能够反映 A 各种特性的可信度 T , 表示为 $T = S(A)$. 用神经网络识别可信度就是要设计一个以 A 为输入并输出可信度 T 的神经网络 NN , 即 $T' = NN(A)$, 并使 T' 以需要的精度 ϵ 逼近 T , 即 $\frac{1}{2}(T' - T)^2 < \epsilon$. 设计 NN 的过程称为系统辨识^[9], 原理如图 1 所示, 过程为: (1) 获取 S 的一组输入和输出, 即训练样本; (2) 设计一个 NN 并通过样本训练它; (3) 如果训练结果满足要求, NN 设计完成, 否则转 (2) 继续执行. 通过训练满足精度要求的神经网络就可以用来识别可信度

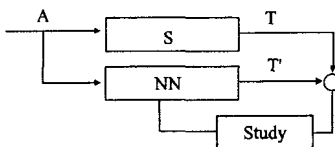


图 1 神经网络辨识原理

识别可信度的神经网络是一个多输入单输出的具有模式识别能力的网络, 可以采用 BP 神经网络 BP 网络结构如图 2 所示

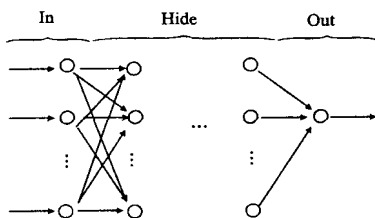


图 2 BP 神经网络结构图

各神经元有多个输入一个输出, 输入和输出的计算关系为: $y = f(\sum_{i=1}^n \omega x_i - \theta)$, y 是输出, f 称为作用函数, θ 称为阈值, ω 是第 i 个输入的连接权重, x_i 是第 i 个输入. 网络从输入层开始依次对每层进行计算, 最后一层的输出就是网络的输出 f 可以选择阈值型、线性型和 S 型函数, 如 hardlim、purelin、tansig 等^[10]. ω 通过训练过程确定

隐含层节点数和层数可调, 在训练过程中根据需要选

定输入层节点个数 k 代表网络可以处理的交易序列的长度. 为在精度和复杂性间平衡, 输入的个数 k 不可能是无限的, 而是一个相对合理的值. k 的大小通过实验选定

因为神经网络输入个数 k 是定的, 而交易记录序列 A 的长度 n 是不定的, 因此, 首先要对 A 进行规范化处理

4 输入规范化

定义 1 交易序列是某一节点记录的它与另一节点的交易成功与失败序列, 可以用一个有序 n 元组表示, 即 $A = (a_1, a_2, \dots, a_n)$.

定义 2 活动交易序列是最近一段时间以来的交易序列, 可以用一个有序 m 元组表示. 设交易序列 $A = (a_1, a_2, \dots, a_n)$, 则活动交易序列 $B = (a_{nm+1}, a_{nm+2}, \dots, a_n)$, 记为 $B = (A, m)$.

定义 3 紧缩交易序列是交易序列的分段平均序列. 设 m 元交易序列 $B = (a_1, a_2, \dots, a_m)$, 则 B 的 k 元紧缩交易序列 $C = (c_1, c_2, \dots, c_k)$, 记为 $C = (B, k)$. 设 $s = \lfloor \frac{m}{k} \rfloor$,

$$\text{当 } i \times s \leq m \text{ 时, } c_i = \frac{1}{s} \sum_{j=1}^s c((i-1) * s + j)$$

$$\text{当 } i \times s > m \text{ 时, } c_i = \frac{1}{r} \sum_{j=(i-1) * s}^m c_j, r \text{ 是 } m \text{ 对 } k \text{ 的余数}$$

为计算局部可信度, 各节点需要记录活动交易序列 B .

5 样本获取及神经网络训练

训练样本有两个来源, 即 P2P 环境实际数据和专门设计的样本. 从 P2P 环境获取训练样本的过程如下:

- (1) 获取节点与 x 的交易序列 A ;
- (2) 由 A 获取长为 m 的活动交易记录 $B, B = (A, m)$;
- (3) 按照选择的 BP 网络输入个数 k 将 B 紧缩为 $C, C = (B, k)$;
- (4) C 和 x 的全局可信度 T 构成一个训练样本 (C, T) .

从 P2P 环境获取的训练样本能够使神经网络识别出的可信度尽量接近全局可信度. 训练样本中再加入专门设计的样本可以实现人为的奖励和惩罚节点的一些行为模式. 专门设计训练样本的过程就是按照神经网络的输入个数 k 和一定的奖励规则给定 A 和对应的可信度 T .

有了训练样本后, 就可以设计 BP 网络并对其进行训练. 样本给出后, BP 网络输入的个数已经确定, 用试探的方法给出隐含层数和各层神经元个数、作用函数、训练方法等, 通过训练确定一个合适的网络

6 局部可信度识别

有了训练好的神经网络后, 局部可信度的计算就十分简单. 过程如下:

- (1) 取出与该节点的交易序列 A ;
- (2) 由 A 生成长为 m 的活动交易序列 $B, B = (A, m)$;
- (3) 求 B 的 k 元紧缩 $C, C = (B, k)$;
- (4) 将 C 作为神经网络的输入, 网络输出就是可信度. 可以看出局部可信度识别过程计算的时间复杂度为 $O(k)$.

7 实验及结果分析

需要通过实验来回答如下问题: (1) 需要多复杂的神经网络才能够实现可信度识别功能; (2) 训练的计算量有多大; (3) 局部可信度识别计算量有多大; (4) 可信度神经网络识别的效果如何

为回答这些问题, 从一个基于 JXTA^[11] 的 P2P 数据备份原型系统实验获取 200 个训练原本, 并专门构造了 40 个样本奖励最近交易成功率增加的节点并惩罚成功率下降的节点。BP 网络选择 20 个输入, 经过试探, 隐含层数选择 1, 隐层节点数选择 20, 隐层选择双曲正切作用函数 $\tan\text{sig}$, 输出层选择全线性作用函数 purelin , 允许误差为 0.00001, 网络经过 500 次左右训练后可达到精度要求。在训练好的网络上计算 220 个可信度需要不到 1 秒时间。神经网络的训练和可信度计算在一台普通 PC 上进行, 配置为 CPU: P4 1.7G, RAM: 512M。

在训练完成的神经网络上, 设计了 4 个具有代表性的测试样本进行识别, 结果如图 3 所示

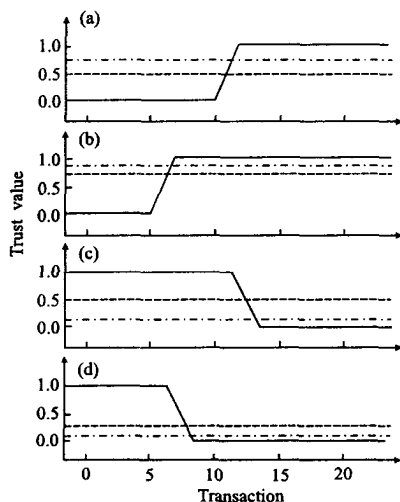


图 3 神经网络识别结果

图中实线表示交易系列; 水平的虚线表示采用简单的局部可信度计算公式 $P = \frac{S}{T}$ 计算的结果, S 表示成功次数, T 表示交易总次数; 水平的点划线表示神经网络识别出的局部可信度。图 3 中 (a) 和 (b) 是可信度有增加趋势的样本, 神经网络识别的可信度比简单方法计算的结果高, (c) 和 (d) 是可信度有降低趋势的样本, 神经网络识别的可信度比简单方法计算的结果低。

从实验结果可以看出, 用包含一个隐含层, 隐层节点数与输入数相同的 BP 网络就可以实现可信度的识别; 训练的计算量和计算时间都是可以容忍的; 在训练好的神经网络上计算可信度速度非常快; 局部可信度神经网络识别法比现有的简单计算方法更能够反映实体的最新信任情况和信任的发展趋势, 具有更高的准确性和实用价值。

8 结论

信任管理是保证 P2P 安全的重要措施, 目前研究注重全局信任模型而忽视局部可信度的计算。简单的局部可信度计算公式只利用了实体间交易成功与失败次数的统计比例关系, 不能准确反映实体的可信任情况, 而基于神经网络的可信度识别方法将交易过程整体作为识别对象, 不但可以反映统计比例关系, 还可以全面反映交易情况的分布特性, 结果更加准确全面。神经网络训练样本中引入专门设计的样本还可以实现对特定交易模式的奖励和惩罚, 使辨识结果更具应用价值。

局部可信度的神经网络识别方法可能遇到的问题就是效率, 有三方面: (1) 需要记录交易系列; (2) 神经网络训练过程复杂且耗费计算资源较多; (3) 在训练好的神经网络上计算可信度也比简单公式计算要复杂许多。(1) 和 (3) 在很多环境中是可以容忍的, 而 (2) 并不需要所有节点都做这样的工作, 没有能力的节点完全可以使用其它节点提供的训练好的神经网络。

下一步将深入研究恶意节点的交易序列特征并设计专门样本来训练神经网络, 使辨识出的可信度对恶意节点有更加全面的惩罚作用, 抑制恶意节点对 P2P 环境的危害。

References

- [1] David Barkai. An introduction to peer-to-peer computing[J]. Intel Developer Update Magazine, February, 2000: 1-5.
- [2] Dejan S. Mijlojic, Vana Kalogeraki, Rajan Lukose, et al. Peer-to-peer computing [R]. Technical Report, HPL-2002-57R1, Hewlett-Packard Company, 2002.
- [3] Blaze M., Feigenbaum J., Lacy J. Decentralized trust management[C]. In: Proceedings of the 17th Symposium on Security and Privacy[J]. IEEE Computer Society Press, 1996: 164-173.
- [4] Blaze M., Feigenbaum J., Ioannidis J., et al. The role of trust management in distributed systems security[J]. In: Secure Internet Programming, Issues for Mobile and Distributed Objects. Berlin: Springer-Verlag, 1999: 185-210.
- [5] Dou Wen, Wang Huaimin, Jia Yan, et al. A recommendation-based peer-to-peer trust model[J]. Journal of Software, 2004, 15(4): 571-583.
- [6] Rita Chen, William Yeager. Poblano a distributed trust model for peer-to-peer networks[R]. Technical Report, TR-14-02-08, Palo Alto: Sun Microsystems, 2002.
- [7] Kamvar SD, Schlosser MT. EigenRep: reputation management in P2P networks[C]. In: Proceedings of the 12th World Wide Web Conference Budapest: ACM Press, 2003: 123-134.
- [8] Simon Haykin, Ye Shiwei, Shi Zhongzhi translate. Neural networks, a comprehensive foundation [M]. Beijing: China Machine Press, 2004.
- [9] Xu li'ha. Neural network control[M]. Beijing: Publish House of Electronics Industry, 2003.
- [10] Martin T. Hagan, Howard B. Demuth, Mark H. Beale, Dai kui translate. Neural network design [M]. Beijing: China Machine Press, 2002.
- [11] Li Gong. Project JXTA: a technology overview [Z]. White paper of Sun Microsystems, Inc. 2002.

附中文参考文献:

- [5] 窦文, 王怀民, 贾焰, 等. 构造基于推荐的 Peer-to-Peer 环境下的 Trust 模型[J]. 软件学报, 2004, 15(4): 571-583.
- [8] Simon Haykin 著, 叶世伟和史中植译[M]. 神经网络原理. 北京: 机械工业出版社, 2004.
- [9] 徐丽娜. 神经网络控制[M]. 北京: 电子工业出版社, 2003.
- [10] Martin T. Hagan, Howard B. Demuth, Mark H. Beale 著, 戴葵译. 神经网络设计[M]. 北京: 机械工业出版社, 2002.