P2P 网络信任数据存储机制综述*)

方 群[1,2,3] 吴国新[1,2] 吴 鹏[1,2] 钱 宁[1,2] 赵生慧[1,2]
(东南大学计算机科学与工程学院 南京 210096)[1]
(东南大学计算机网络和信息集成教育部重点实验室 南京 210096)[2]
(安徽师范大学数学计算机科学学院 芜湖 241000)[3]

摘 要 P2P 网络以其在动态、自组织、伸缩性、鲁棒性、资源利用率等方面的优势,已经发展成为当前互联网络的重要分支并获得广泛应用。但由于 P2P 支持用户匿名访问,其中包含大量恶意节点,因而存在安全隐患,故须依赖信任机制保证对等节点间的协作,其中信任数据存储成为一个关键问题。首先介绍 P2P 信任模型的基本构成,进而对目前 P2P 信任模型中常用的信任数据存储技术作了深入研究,同时通过分析对比总结各自的优缺点,最后提出理想的信任数据存储方案应具备的特征。

关键词 P2P,信任,存储,分布 Hash 表

Survey of Trust Data Storage Mechanism in Peer-to-Peer Network

FANG Qun[1,2,3] WU Guo-xin[1,2] WU Peng[1,2] QIAN Nin[1,2] ZHAO Sheng-hui[1,2]
(School of Computer Science and Engineering, Southeast University, Nanjing 210096, China)[1]
(Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China)[2]
(College of Math & Computer Science, Anhui Normal University, Wuhu 241000, China)[3]

Abstract Because of many good features, such as dynamic, auto organized, scalability, robust and resource utility, peer-to-peer (P2P) network becomes to an important branch of Internet and has been widely applied. But because P2P network permits anonymous access including many malicious peers, which bring many security problems to P2P network, cooperation between peer and peer must be supported by trust relation established between peer and peer. This paper focused on the issue of trust data storage. Firstly the components of general trust model were introduced. Next several typical trust data storage models of P2P were discussed. After careful analyzed, these models display their advantages and disadvantages. Finally this paper puts forward the essential features of the ideal trust data storage model.

Keywords Peer-to-peer, Trust, Storage, DHT

1 引言

随着网络技术的飞速发展和网络应用领域不断拓展,传统 Client/Server 系统逐渐显露出其固有的缺陷。主要因为系统大多以服务器为中心,资源集中存储和管理,虽然管理效率较高,但服务器业已成为系统瓶颈,严重制约着整个系统的性能。若系统中用户数目较多,则系统可能因服务器过载而导致性能严重下降,另外还存在着单点失效和 DoS 攻击等安全隐患。另外处于网络边缘的资源在集中模式下得不到有效使用,利用率不理想。

为了改进 C/S 系统的性能,增强其健壮性和伸缩性,提高资源的利用率,一种新的技术——P2P (peer-to-peer, 对等网络)[1]被重新利用。在 P2P 系统中,不存在客户机和服务器的角色区别,每个节点身兼服务的提供者和消费者两种角色,既可以做为服务器为其他节点提供服务,也可以消费其他节点提供的服务。通过 P2P 网络将分布于网络边缘的资源组织起来,现有资源得以充分利用,其综合计算能力将远远超过现有中心服务器。正因为 P2P 网络具有动态性、自组织、伸缩性好、扩展性好、适应性强、健壮性强、可靠性好等优点,经过多年的研究和开发,P2P 被广泛应用于多种领域,如文件共享、即时通信、网络协作等。目前已经有大量成熟的应用系统出现,例如 Napster, Gnutella, BT 和 eDonkey 等。

虽然 P2P 应用日益广泛,但目前 P2P 网络中存在着大量的自私行为,例如"搭便车"(Free Riding)问题[2]和"公共悲剧"(Tragedy of Common)[3]问题,以及大量的欺诈、伪造等恶意行为,严重影响系统可用性。究其原因,主要是由于用户片面追求个体利益的最大化而损害其他个体或公共利益,只想获取而不愿意贡献所致。因而总的来看,目前整个 P2P 系统资源可用性不高,服务质量有待改善。

由于 P2P 网络与人类社会具有较高的相似性,因此可以仿造人类社会中存在的信用体系,在 P2P 系统中建立信任机制,使用信任机制来规范用户的行为,惩恶扬善,构建一个可信网络[4]。在可信的网络中,用户之间存在协作关系,协作是否能够成功,则取决于协作双方的可信程度。信任关系实际上是网状分布的,每个用户的信任度取决于其他用户的评价,而任一用户的行为和评价也将直接或间接地影响其他部分用

*)基金项目:国家 863 计划(2007AA01Z422);安徽省高校教师青年基金(2007jq1061);安徽省教育厅自然科学研究项目(2005kj088)。方 群博士生,副教授,硕士生导师,主要研究领域为分布式网络信任与安全技术、可信计算等。

户的信任度。

本文着重研究 P2P 信任模型中常见的信任数据的存储模式。第 2 节介绍 P2P 信任模型的分类，第 3 节通过研究现有信任模型，抽象出具有一般性信任模型的框架结构。第 4 节重点总结各种不同的信任模型中常用的信任数据的存储方式，第 5 节根据信任需求，总结得出理想信任数据存储模式应具有的特点，最后总结并展望。

2 信任模型分类

自 1994 年 Marsh 率先提出信任概念以来，信任管理理论与技术的研究已经获得了较大的发展。目前有多种信任模型分类方法，按照文献[13]的方法可将 P2P 信任模型分为以下 4 种：

(1)集中控制模型。此类系统中存在少数领袖节点(Leader Peer)负责监控网络中的其他节点的行为，定期发布违规节点。领袖节点的合法性依赖于 CA 证书保证。由于此类系统中存在可扩展性、单点失效等问题。此类系统中典型的是采用 PKI[5]，如 eDonkey 系统。

(2) 数据签名。目前较为流行的文件共享应用 Kazaa，采用该方法(Sig2Dat[6])，在每次下载完毕以后，用户对经过核准的数据进行签名表达自己的信任，因而数据获得的签名越多也就意味着其可信度越高。

(3) 基于局部推荐的信任模型。此类模型中，节点的信任数据来源于局部节点，因而存在片面性，但通过广播方式可以在短时间内获得信任数据。如 Cornelli 对 Gnutella 的改进建议[7]中采用的就是这种方法。

(4) 基于全局推荐的信任模型。此类系统中，为了获得某个节点的信任度，需要综合整个网络中所有节点的推荐信息。Stanford 的 eigenRep[8] 和文献[13]中采用此种信任模型，此类模型在计算信任度时开销较大，还存在迭代算法的收敛性问题。

3 P2P 信任模型

3.1 信任模型框架

一个信任模型 TM(Trust Model)可以形式化定义如下：

$TM=\{Ents, Vals, Funcs, Evts, Prots, Nets\}$

其中：

Ents——实体(Entry)集合，实体包括节点、用户、链路以及各种其他资源；

Vals——属性(Value)集合，即与信任计算相关的各实体的属性集合，如 ID、信任度、公钥等属性；

Funcs——映射或函数(Function)集合，即各实体—属性间的映射关系；

Evts——事件(Event)集合，即网络中发生的各种事件，如节点下/上载资源、节点恶意行为等；

Prots——协议(Protocol)集合，即节点交互活动中为了维护信任管理体系所遵循的协议的集合；

Nets——网络(Network)模型，即代表信任模型所依赖的网络模型。

网络模型 Nets 又可以简单抽象为一个有向图 $G=\langle V, E\rangle$，其中 V 表示节点集合，E 表示有向边的集合，不同的网络模型中节点之间的关系不同，因而有不同的边集合。

因而，可以给出信任模型的一般性框架如图 1 所示。

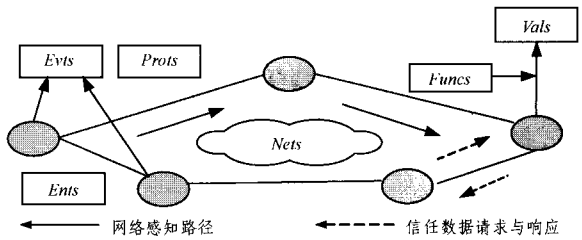

图 1 通用信任模型的框架

从图 1 中可以看出，下层网络(如 Gnutella、Chord 等)是信任模型赖以生存的基础，但信任模型并非仅局限于该种网络，在不同的网络平台之上可能构建相似的信任管理系统。引起实体 e(节点或资源)的信任度值改变的原因是与 e 相关的某些事件(下载或提供下载)的发生，e 可能(也可能没有)参与该事件。事件发生后产生的结果被网络感知(可能通过监视节点)，并将证据转送给 e 的档案节点 d(即存放 e 的信任数据的节点)，由档案节点按照一定算法更新信任数据。当发起节点需要获取实体 e 的信任数据时，就向 d 发送请求，然后根据 d 返回的信任数据决定(较复杂的计算和决策问题)是否值得信任实体 e。

3.2 信任模型中的基本概念

信任模型中的基本概念定义了节点间信任关系的基本属性，一般使用连续的或离散的形式表达(满足可测性)。各种属性都由监控节点报告而得(即采样)，不同的属性的采集和统计方法可能不同，最终由特定算法得到信任度值：

$T=f(x_1, x_2, \cdots, x_m)$($x_i$ 是影响信任度因素，m 是因素的数目)

在计算信任度时考虑的因素越多，对节点的评价也就会越准确，但也会增加计算复杂程度和存储开销。下面我们以最简单的信任模型 EigenRep[8] 为例，来介绍信任模型中常用的基本概念，在更为复杂的信任模型中，涉及的属性较多，但它们都与以下属性有一定关系。

(1)交易

所谓的交易(Interaction)是指节点 i 与节点 j 进行的一次交互，这种交互可能包含多种内容，如查询、下载、计算等。每个交易有一个发起者和一个提供者，如果交易的结果使得发起者满意(即对交易结果的评价是正面的，与其预期一致)，则交易成功(Success)，否则交易失败(Failure)。一般用 S_{ij} 和 F_{ij} 分别表示成功数和失败数。

(2)局部看法和评价

局部看法一般指节点间交易历史中的成功比例，即成功次数与交易总次数的比值，用 P_{ij} 表示。评价(Evaluation)一般可以定义为两个节点间在所有交易中成功次数和失败次数的数值对，即$\langle Success, Failure\rangle$。这种定义完全根据节点间的交易历史得出，不考虑被评价者的信任度。评价一般用 E_{ij} 表示。

(3)信任和信任域

信任(Honesty)表达节点间的一种关系，本来无法量化，但可用某种方法将它转换为数量，这样信任就可以定义为节点 i 对节点 j 的评价值，其中节点 i 是发表评价者，而节点 j 是被评价者，评价者可以根据与被评价者的历史交易情况及其全局信任度得出评价结果。

信任一般表示成一个连续实数值或逻辑值，常用的数值

表示方法将信任值限定于一个数域中，如在[0,1]或[−1,1]上。如果使用逻辑值表示信任，则信任域为{T,F}或{0,1}。因此也可以定义信任是节点间信任关系在信任域上的投影，信任一般用 O_{ij} 表示。

(4)推荐度和信任关系矩阵

推荐度(Recommendation)是指节点 i 在全局中对节点 j 的正向积极评价中的贡献程度，一般用 R_{ij} 表示。

信任关系矩阵 $R=(R_{ij})_n$

(5)可信度

可信度定义为全部参与评价的节点对某个节点评价的加权和，表示为 T_j 网络中所有节点的可信度构成整个网络的可信向量 $T=(T_1,T_2,\cdots,T_n)^T$。

3.3 信任度的计算

信任数据更新由交易事件驱动。在网络中的每一次交易结果传送至相关节点，并按照既定的计算方法重新计算各节点的信任度，某个节点信任度的变化会影响到网络中其他部分节点的信任度值，这是一个迭代的过程，仅当这种迭代是收敛的时候，整个网络才会在多次迭代后稳定下来。

信任关系矩阵 R 决定了迭代过程的收敛性。表达这一迭代过程可以使用信任方程，即 $R^TT=T$。

通过以上数据可以计算出节点 r 的全局可信度 $T_r=\sum_k(R_{kr}\times T_k)$。

信任度是信任模型中最为核心的概念，通常是全局概念，即 P2P 网络中某个节点在整个网络中的被信任的程度，它的取值将会影响节点的访问权限和行为约束。例如，如果节点不断上载好的资源，则它的信任度会逐渐增加，那么作为奖励，它可以从其它节点那里获得更多的资源或服务。但信任度的值不是由节点本身决定的，而是由与节点相关的其他节点共同评价的综合的结果。除此之外，有的系统还为某些资源如文件、链路等赋予信任度的属性。

信任度的提供者在不同的模型中有很大差别，这主要取决于 P2P 网络模型、拓扑结构以及采用什么样的协议。信任度数据本身也存在着可信度的问题，即要防止恶意节点破坏信任度的计算规则，将伪造虚假的数据提供给请求者。因此，在信任模型中，需要制订一套信任数据的安全传递协议，一方面为信任度数据的生成、传递、存储规定了原则，另一方面主要提供安全机制，保证信任度数值本身的可靠性。

在 P2P 网络中，一个节点的信任度既不是由它本身计算，也不应由它自身保存，而是在需要时计算，在异地保存，采取不同的信任度保存策略会给信任模型的效率带来一定的影响，进而也会影响整个系统的性能。

4 P2P 信任数据的存储

可信网络由信任机制保证其服务的可靠性，而信任数据是信任模型中最基本的组成部分，它反映了网络运行过程中各实体的历史信息，通过它可以预测实体在未来时间内按照预期执行的概率。

在信任度计算的过程中，需要大量相关数据，但保存所有实体历史数据的方案显然是不可行的，将造成存储空间的不足。但如果数据量过少，也会影响实体信任度的评估准确性，特别是无法保证信任评估的动态性。EigenRep 是一种平均值模型，其中保存了最少的信任数据，即若干关于实体信任的三元组〈S，F，T〉，它们分别表示交易成功、失败次数和提供方的信任度。如果要实现更为安全的信任机制，则可能还需要保留密钥等数据。图 2 显示的是常见的信任数据结构。

ID_i		$T_r^{(k+1)}$	
$S_{k_1}r$	$F_{k_1}r$	$T_{k_1}^{(k)}$	ID_{k_1}
$S_{k_2}r$	$F_{k_2}r$	$T_{k_2}^{(k)}$	ID_{k_2}
……	……	……	……
$S_{k_t}r$	$F_{k_t}r$	$T_{k_t}^{(k)}$	ID_{k_t}

图 2 信任数据记录结构[8,13]

图 2 中节点 r 的标识符用 ID_r 表示，T_r 是其全局信任度，右上角标$(k+1)$表示迭代的次数。

根据信任数据与实体对象的距离，可以将信任数据存储分为两类，即本地保存和远地保存，前者是节点保存自身的信任数据，一般仅限于保存交易历史。后者是指将信任数据保存在其他节点，该节点称为原节点的档案节点。根据信任数据保存地点的分布密度，可以将信任数据存储分为集中保存和分散保存。集中存放是指将信任数据存储在固定的一个或几个节点之上，分散保存则是有大量节点参与信任数据的保存，每个节点保存一个或多个节点的信任数据。

数据的存储模式中的核心问题是数据存储的位置、数据的分布及相互关系。为了简化计算，目前常用系统都是把一个实体的信任数据集中在一点存放，而不是分散存放，因此，后两个问题可不考虑。数据存储的位置在不同的信任模型中不同，就节点与其档案节点的对应关系来看，可以分为 $m:1$，$1:1$ 和 $m:n$ 三种情形。不管哪一种存储模式，都必须解决信任数据的读取和更新问题，不同模型在数据存取效率和开销方面有较大差异。

4.1 $m:1$ 模式

多对一的模式就是所有节点的信任数据都存储于单一节点，即是集中式信任数据存储模式。这种模式可以使用如下映射表示：

$f(node_i)=node_0,(i=1,2,\cdots)$

其中 $node_0$ 代表中心信任服务器，它是所有节点的档案节点。在传统的 C/S 模型网络中，一般都采取这种存储模式。也有 P2P 网络采用此种模型，如 JXTA[5] 网络采用的即是这种 $m:1$ 模式。JXTA 采用基于 PKI 机制的信任管理机制，较为典型，但开销太大，常需建立当前节点组中所有 Peer 的信任度表格，计算组与组之间的信任度算法也很复杂。

$m:1$ 模式的优点很明显，信任数据全部存放在同一个可信的中心信任服务器，存取方式较简单，更新也比较方便。但是它存在着所有集中式系统共有的缺陷，也就是负载不均衡，可伸缩性不强，另外由于所有节点都知道其档案节点的位置，因此容易遭受 DoS 攻击，还存在单点失效问题。因此采用集中式信任数据存储模式不是理想的方案，不能够体现 P2P 的优越性。

4.2 1:1 模式

该模型是指把指定节点的信任数据存放在另一个唯一对应的节点，此模型中节点与其档案节点的对应关系如下：

$f(node_i)=node_j$，且 $f-1(node_j)=node_i(i\neq j,i,j=1,2,\cdots)$

文献[8]中 eigenRep 的信任数据存储基于 Berkeley 的 DHTs 协议 CAN[12]，CAN 从一个 d 维超环面空间中选择关键字，每个节点都与这个空间中的一个超立方体区域相关联，它的邻居是拥有相邻超立方体的节点，路由机制将查询请求消息转发至更加接近关键字的区域。CAN 具有一个区别于

其他算法的特性，即节点有 $O(d)$ 个邻居并且路径长度为 $O(dn^{1/d})$ 跳。因此，此种存储结构可以保证在 $O(dn^{1/d})$ 步内获取指定节点的信任数据。

文献[9]中采用 Chord[10] 协议，它使用一个一维环状关键字空间。负责保存关键字的节点是在数值上最接近关键字标识的后续节点，该节点称为关键字的后继。Chrod 维护两个邻居节点集合，每个都是包含有 k 个直接后继节点的后继列表，路由机制即依靠此列表，将请求转发给标识最接近但没有超过该关键字的节点，跳数为 $O(\log n)$。

文献[13]采用 Terrace P2P 网络构造信任数据存储网络，Terrace 是一种基于 d-tree 结构的 DHTs 构件，所有节点都被映射到一个逻辑 d-tree 上，并给每个节点赋予全局惟一的逻辑地址，它有如下特点：(1)任意节点 i 通过 Terrace 在 $O(\log N)$ 的消息复杂度内将节点 j 的属性(如评价 E_{ij})写入 $H(ID_j)$，同时保证节点 i,j 难以获知 $H(ID_j)$ 的具体位置(IP 地址)；(2)在不知 j 的 IP 地址的情况下，任意节点 i 可以按照 $O(\log N)$ 的消息复杂度从逻辑地址 $H(ID_j)$ 获取节点 j 的有关数据(如全局可信度 T_j)；(3) Terrace 具有较强的容错能力；(4)Terrace 具有较小的拓扑维护开销 $O(d)$。

1∶1 模式将信任数据分散在大量的其它节点中，使得单个节点用于维护信任数据的负载得以均衡，但其可靠性严重依赖于 DHT 等散列函数的可靠性，例如 Chord 和 CAN 一般是针对节点的 IP 地址进行 Hash，这样 IP 地址固定的节点其对应的档案节点也就固定，因而难以防止协同作弊。由于 P2P 网络的动态特性，这样的信任数据模型维护开销也容忽视。

4.3 m∶n 模式

此种模式中每个节点的信任数据保存于多个节点，因而可以避免因节点失效造成信任数据丢失。文献[10]中的 Chord 改进方法是将节点的映射由 1∶1 变为 1∶n，映射关系如下：

$$f(node_i)=\{node_{j1},node_{j2},\cdots,node_{jk}\},\ (i\neq j,\ k=1,2,\cdots,n)$$

节点的信任数据就被冗余存储在多个档案节点中，从任意一个档案节点都可以取到相同的信任数据，不但可以保证信任系统的可靠性，同时也可以用来鉴别信任数据的真伪。但是这种方法存储和更新开销过大，保持多个备份的同步也比较困难，给系统带来不必要的负担。

还有一种方案是预先路由表中存储各邻居节点的信任数据，在传递查询请求的过程中预先计算下一跳节点的信任度，选择可信的路由和目的节点，称为信任敏感的路由技术。这种方法的特点是在查询过程就可以进行信任数据的统计和计算，最终定位可信节点，而不需要获取全部信任数据后再一并计算，因而效率比较高，但节点的信任数据存储在邻居节点，同样存在安全性和同步困难问题，而在某些拓扑感知的 P2P 系统中，地理位置接近的节点往往 ID 也相近，因此无法防止协同作弊的出现。

5 理想的信任数据存储模式应具备的特征

由于 P2P 网络中信任模型使用场合较多，在保证系统安全性、可靠性、公平性等特性的同时，用户更关心的是它的开销问题，需要将其自身维护开销降至最低。为了降低信任模型的开销，首先需要降低信任评估算法的复杂度，但信任算法过于简单，则会抵消信任机制的效果。另外由于信任计算是一种分布计算方式，信任数据的存储模式对计算信任度的开销也有重要影响，如果数据存储过于分散，数据量过大，在计算信任度时网络传输开销也较大。那么信任数据的存储模式应该具有哪些特征，才能保证信任模型的安全、高效、快捷呢？

我们认为，理想的信任数据存储模式应当具有以下特性：

(1)全局性。应当存储整个网络对实体的信任数据，而不应只考虑自身或者局部的信任数据，以防止片面地评价。这一点主要取决于信任模型采用的信任计算模型。

(2)均衡性。即信任数据的存储不能过于集中，集中不符合 P2P 网络的设计初衷，而应当均匀地分布在 P2P 网络中，采取基于 DHT 算法的方式可以达到这种均衡性。

(3)安全性。信任数据本身的安全性是保证信任机制安全性的关键因素，因而信任数据必须以安全的方式存放。这就要求任意节点不应知道本身的信任数据存放的确切位置，档案节点也不了解存放的数据的具体情况，以防作弊，恶意篡改其信任数据。这就需要强大的加密机制和 DHT 机制，最好是将逻辑地址和物理地址分离的方法。

(4)可用性。P2P 环境具有动态性，节点会动态地加入和退出，因而要保证信任数据的可用性，就要保证档案节点随时可用，要求对信任数据冗余存放，在部分档案节点失效后，信任机制仍然能够正常运行。

(5)高效性。需要提高计算效率和传输效率，降低计算开销和传输开销。由于信任数据的分散存储，信任数据的读取和更新都需要借助网络传输，从而增加了带宽开销。因此要求存取信任数据所经过的网络跳数尽可能少，这样才能减少传送时间，降低传输开销。另外也可采取数据压缩方法进一步减少实际传递的数据量。

其中，某些特征之间是相互制约的，比如可用性和高效性，如果单考虑可用性，片面地增加档案节点的数量，可能在获取信任数据时效率比较高，但在更新信任数据时则效率较低。因此，在设计信任模型时，需要全面考虑，根据实际系统的需求作好取舍。

结束语 由于 P2P 网络存在潜在的安全隐患，为了保证整个系统的安全性、公平性、可靠性等需求，在 P2P 系统中引入信任机制来保障系统的可信性。P2P 信任数据的存放策略是信任模型中的重要设计内容，同时也决定着信任模型其他部分的设计与实现，需要认真研究。但目前的存储模式实际上都是在原有 P2P 系统之外，维护着一个新的结构化 P2P 存储系统，效率方面仍然存在问题，因而限制了信任模型优势的进一步发挥。未来的工作主要是设计存储性能更好的信任管理模式，全面提高信任模型的效率。

参 考 文 献

[1] P2P Group. http://www.p2p.org

[2] Adar E, Huberman B. Free Riding on Gnutella [R]. Xerox PARC, 2000

[3] Feldmany M, Laiz K. Quantifying disincentives in peer-to-peer networks[C]// Workshop on economics of peer-to-peer systems [M]. LNCS 2735. Berkeley. CA: Springer-Verlag, 2003: 117-122

[4] Caronni G. Walking the Web of trust[C]// Sriram RD, ed. Proc. of the IEEE 9th Int'l Workshops on Enabling Technologies. Infrastructure for Collaborative Enterprises. IEEE Press, 2000: 153-159

[5] Altman J. PKI Security for JXTA Overlay Networks[R]. TR-I2-03-06. Palo Alto: Sun Microsystem, 2003

[6] Sig2dat specification. http://www.geocities.com/vlaibb/,2002
[7] Cornelli F. Choosing reputable servents in a P2P network[C]// Lassner D, ed. Proc. of the 11th Int'l World Wide Web Conf. Hawaii:ACM Press, 2002:441-449
[8] Kamvar S D, Schlosser M T. EigenRep: Reputation management in P2P networks[C]// Proceedings of the 12th Int'l World Wide Web Conference. Budapest: ACM Press:123-134
[9] Zhang Zhen,et al. A P2P Global Trust Model Based on Recommendation[C]//Proceedings of the Fourth International Conference on Machine Learning and Cybernetics. Guangzhou, August 2005
[10] Stoica I, et al. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications[C]//SIGCOMM'01. San Diego, California, USA, August 2001
[11] Ratnasamy S. A Scalable Content-Addressable Network[C]// SIGCOMM'01. San Diego, California, USA, August 2001
[12] Joseph D, Kubiatowicz J D. Routing Algorithms for DHTs: Some Open Questions[C]// Electronic Proceedings for the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02) 2002. MIT Faculty Club, Cambridge, MA, USA,2002
[13] 窦文.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):571-580

(上接第 18 页)

[36] 李存华,孙志挥,陈耿.基于网格上近似的大规模数据集离群点检测算法 GROUT.计算机应用研究, 2003, 20(9):34-136
[37] Aggarwal C C, Yu P. Outlier detection for high dimensional data// Proc. of the ACM SIGMOD International Conference on Management of Data. Santa Barbara, 2001:37-47
[38] Angiulli F, Pizzuti C. Outlier Mining in Large High Dimensional Data Sets. IEEE Trans. Knowledge and Data Eng., 2005, 2 (17):203-215
[39] Angiulli F, Basta S, Pizzuti C. Distance-based detection and prediction of outlier. IEEE Trans. Knowledge and Data Eng., 2006, 2(18): 145-160
[40] Aggarwal C C. Re-designing Distance Functions and Distance-based Applications for High Dimensional Data. SIGMOD Record Date, 2001, 30(1):13-18
[41] Yu Dantong, Gholamhosein S, Zhang Aidong. FindOut: Finding Outliers in Very Large Datasets. Knowledge and Information Systems, 2002,4(4):387-412
[42] Dutta H, Giannella C, Borne K, et al. Distributed top-k outlier detection in astronomy catalogs using the demac system// Proc. of 7th SIAM International Conference on Data Mining. Minneapolis,2007:208-215
[43] 许龙飞,熊君丽.基于粗糙集的高维空间离群点发现算法研究.计算机工程与应用,2004,40(7):58-60
[44] Knorr E M, Ng R T. Finding Intentional Knowledge of Distance-based Outliers// Proc. of the 25th VLDB. Edinburgh,1999:211-222
[45] Chen Zhixiang, Tang Jian, Fu Ada Wai-Chee. Modeling and Efficient Mining of Intentional Knowledge of Outliers// Proc. of the 7th International Database Engineering and Applications Symposium Conference. Hong Kong,2003:44-53
[46] Shekhar S, Lu C-T, Zhang P. A Unified Approach to Spatial Outliers Detection. GeoInformatica, 2003,7(2):139-166
[47] Shekhar S, Lu C-T, Zhang P. Detecting Graph-based Spatial Outliers. International Journal of Intelligent Data Analysis (IDA), 2002, 6(5):451-468
[48] Lu C-T, Chen Dechang, Kou Yufeng. Algorithms for Spatial Outlier Detection// Proc. of 3rd International Conference on Data Mining. Melbourne, 2003: 597-600
[49] Lu C-T, Chen Dechang, Kou Yufeng. Detecting Spatial Outliers with Multiple Attributes// Proc. of the 15th International Conference on Tools with Artificial Intelligence. Sacramento, 2003:122-128
[50] 文俊浩,吴中福,吴红艳. 空间孤立点检测. 计算机科学,2006, 33(5):185-187
[51] Sanjay C, Sun Pei. SLOM: a new measure for local spatial outliers. Knowledge and Information Systems, 2006, 9(4): 412-429
[52] Kou Y, Lu C-T, Chen D. Spatial Weighted Outlier Detection// Proc. of the SIAM Conference on Data Mining. Bethesda, 2006: 613-617
[53] Xue Anrong, Ju Shiguang. Algorithm for Spatial Outlier Detection Based on Outlying Degree// Proc. of the WCICA 2006. Dalian,12(7):6005-6009
[54] Jagadish H V, Koudas N, Muthukrishnan S. Mining deviants in a time series database// Proc. of the 25th VLDB. Edinburgh, 1999:341-350
[55] Choy K. Outlier detection for stationary time series. Journal of Statistical Planning and Inference, 2001, 99 (2):111-127
[56] Ma J, Perkins S. Time-series novelty detection using one-class support vector machines// Proc. of the International Joint Conference on Neural Networks. 2003:168-175
[57] Dasgupta D, Forrest S. Novelty detection in time series data using ideas from immunology// Proc. of the International Conference on Intelligent Systems. 1999:82-87
[58] Shahabi C, Tian X, Zhao W. TSA-tree: a wavelet-based approach to improve the efficiency of multi-level surprise and trend queries// Proc. of the 12th International Conference on Scientific and Statistical Database Management. 2000:55-68
[59] Keogh E, Lonardi S, Chiu B. Finding surprising patterns in a time series database in linear time and space// Proc. of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, 2002:550-556
[60] Bejerano G, Yona G. Modeling protein families using probabilistic suffix trees// Proc. of the Third Annual International Conference on Computational Molecular Biology. 1999:15-24
[61] Sun Pei, Chawla S, Arunasalam B. Mining for Outliers in Sequential Databases// Proc. of the Sixth SIAM International Conference on Data Mining. Bethesda, 2006: 94-105
[62] 薛安荣,何伟华. 基于时序离群检测的新的分段方法.计算机工程与设计,2007,28(20):4875-4877
[63] 姚卫新. 智能数据分析中异常数据的集成化管理方法研究.上海:复旦大学,2004
[64] 陆声链. 孤立点挖掘及其内涵知识发现的研究与应用. 南宁:广西大学,2005
[65] Gwadera R, Atallah M J, Szpankowski W. Reliable detection of episodes in event sequences. Knowledge and Information Systems, 2005, 7(4):415-437