

语义相似和多维加权的联合敏感属性隐私保护

徐龙琴¹, 刘双印^{1,2}

(1. 广东海洋大学 信息学院, 广东 湛江 524088; 2. 中国农业大学 信息与电气工程学院, 北京 10083)

(xlqw@126.com; hdsyxq@126.com)

摘要:针对现有 k -匿名方法直接用于多敏感属性数据发布中存在大量隐私泄露的问题,提出一种基于语义相似和多维加权的联合敏感属性隐私保护算法。该算法通过语义相似性反聚类思想和灵活设置多敏感属性值的权值,实现了联合敏感属性值和语义多样性分组的隐私保护,并根据应用需要为数据提供不同的隐私保护力度。实验结果表明,该方法能有效保护数据隐私,增强了数据发布的安全性和实用性。

关键词:隐私保护;联合敏感属性;语义相似度;多维加权; l -diversity

中图分类号: TP309.2 **文献标志码:** A

Privacy protection method for composite sensitive attribute based on semantics similarity and multi-dimensional weighting

XU Long-qin, LIU Shuang-yin

(1. School of Information, Guangdong Ocean University, Zhanjiang Guangdong 524088, China;

2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: In view of a large number of privacy disclosure issues when using k -anonymity method directly for multi-sensitive attribute data publishing, a joint privacy-sensitive properties preserving algorithm based on semantic similarity and multidimensional weighting was proposed. This algorithm realized security protection of the joint-sensitive property value and the semantic diversity of the privacy group with the help of the semantic similarity anti-clustering principle and counter-sensitive property value. According to different application needs, data privacy protection of different extent was provided. The experimental results show that this method can effectively protect data privacy and enhance data security and practicality.

Key words: privacy protection; composite sensitive attribute; semantic similarity; multidimensional weighting; l -diversity

0 引言

随着网络的迅猛发展,大量个人信息被政府机关、商业机构、企事业单位存储发布,虽各发布单位发布数据时均采取了各种隐私数据保护措施,隐匿了个人身份标志和某些隐私数据,但文献[1]研究表明,通过对公开的选民登记表和隐匿了标识属性的医疗信息表进行连接操作,超过87%的美国公民身份被唯一确定。Sweeney等人^[1]率先提出了 k -匿名隐私保护模型,有效解决了连接攻击造成身份泄露的问题,但没考虑多敏感属性的多样化需求,存在敏感属性泄露问题。

针对 k -匿名模型存在的问题,Traian^[2]、Aggarwa^[3]、Machanavajjhala^[4]、Li^[5]、Xiao^[6]、王茜^[7]、杨晓春^[8]等人从数据交换、敏感属性多样化、约束敏感属性值等角度对 k -匿名隐私保护算法进行改进与扩展。以上方案虽然从理论角度对 k -匿名模型中敏感属性泄露问题有所改进,但存在着不同程度的不足。特别是现有研究主要针对单敏感属性数据保护,在多敏感属性隐私保护研究相对较少。例如表1中的PhoneNo、Job、Salary、Disease作为多敏感属性,且Disease属性值为HIV的敏感程度要远高于属性值为flu、bronchitis的敏感程度。因为与高危疾病相比人们相对来说不太在意别人知道自

己患简单易治的疾病,所以具有高敏感度的属性值更需要强力保护。若将单敏感属性隐私保护方法在不考虑属性语义相似及属性值重要性的情况下直接应用于多敏感属性,则不能达到有效防止隐私泄露的目的。

表1 原始数据表T

Rno	Age	Zipcode	Sex	Job	PhoneNo	Salary/元	Disease
r1	20	10015	Male	Craft-repair	775160	8 000	HIV
r2	22	12060	Female	Sales	833110	10 000	HIV
r3	21	21010	Male	Prof-specialty	215686	10 000	Cancer
r4	25	34000	Female	Prof-specialty	234721	8 000	HIV
r5	32	35300	Female	Sales	338410	5 000	Cancer
r6	54	26200	Male	Teacher	284582	18 000	pneumonia
r7	60	38100	Male	Teacher	160187	22 000	Cancer
r8	48	18400	Female	Gov-service	209642	31 000	bronchitis
r9	33	49500	Male	Gov-service	456610	18 000	flu

为此,提出基于语义相似和多维加权的联合敏感属性隐私保护方法,将多维加权的敏感属性综合语义相似性计算和反聚类思想引入到联合敏感属性多样性划分中,给出了数据集的联合敏感属性多样性分组方法,使每个分组记录同时具有联合敏感属性值多样性和语义多样性,根据数据重要性的差异设置相应的权值,为数据发布提供灵活的隐私保护力

收稿日期:2010-09-20;修回日期:2010-11-24。

基金项目:国家星火计划项目(2007EA780068);广东省自然科学基金资助项目(7010116);广东省粤港关键领域重点突破项目(2010B020315025);广东省科技计划项目(2008B021300002);湛江市科技计划项目(2010C3113011)。

作者简介:徐龙琴(1977-),女,陕西汉中,人,讲师,硕士,CCF会员,主要研究方向:数据库、人工智能、智能信息系统;刘双印(1977-),男,山东菏泽人,副教授,博士研究生,CCF会员,主要研究方向:人工智能、智能计算、智能信息系统。

度,能有效克服现有研究方案的不足,提高隐私数据保护技术的实用性。

1 相关研究及分析

文献[2]通过约束每个等价组中敏感属性至少要有 p 个不同取值,提出了 P -sensitive k -匿名模型,有效阻止了敏感属性泄露问题,但没有对同一等价组中敏感属性值出现频率施加控制,当敏感属性值分布严重不均衡时很难避免概率攻击。文献[3]通过选择性发布每个分组的统计信息来实现隐私保护。文献[4]提出了一种基于 l -diversity 规则的增强型 k -匿名模型,要求每个分组中至少包含 l 个“well-represented”敏感属性值,信息的泄露率控制在 $1/l$ 范围内,但存在一次只能处理一个敏感属性的不足。文献[5]提出了 l -closeness 规则,采用 Earth Mover Distance 来进行分组划分,有效阻止相似攻击,却不能防止身份泄露。文献[6]通过实验验证和理论分析证明了 l -diversity 规则能够为个体提供 stronger 的隐私保护,接着提出了 anatomy 新的隐私规则,将敏感属性与类身份属性分开发布削弱了类身份属性与敏感属性之间联系,避免了对类身份属性泛化处理,提高了数据分析的准确性。但由于直接发布类身份属性值致使不能有效阻止最基本的隐私攻击。文献[7]在考虑属性敏感度约束的基础上改进了 P -sensitive k -匿名模型,提高了隐私信息保护的灵活性。文献[8]通过继承 Classfly 算法的元组概括过滤思想,提出多约束 k -匿名化方法 Classfly+,有效降低了多约束 k -匿名化的信息损失。以上算法在敏感数据发布隐私保护方面取得了一定的成效,但都是针对单敏感属性数据,没有考虑属性取值重要性的差别或把所有属性重要性同等对待,但现实中要保护的敏感属性可能不止一个且属性值敏感程度差别很大。文献[9]给出了基于最大叶子树优先策略的多敏感属性保护算法,改进了 l -diversity k -匿名模型。文献[10]提出了基于有损连接技术的支持多敏感属性的隐私数据发布多维桶分组技术,提高了算法的实用性。因此,数据隐私保护中有必要同时考虑多个敏感属性以及各敏感属性值语义相似性和重要性对发布数据的影响,以对现有的算法进行改进或扩展,进一步提高隐私保护能力。

2 语义相似及多维加权联合敏感属性隐私保护

假设待发布的关系 $T\{IA_1, IA_2, \dots, IA_c, QA_1, QA_2, \dots, QA_f, SA_1, SA_2, \dots, SA_g\}$, 发布关系 $T'\{QA_1', QA_2', \dots, QA_f', SA_1', SA_2', \dots, SA_g'\}$ 。其中身份标识符(Identifier)属性为 $IA_k (1 \leq k \leq c)$, 如 Name、ID(身份证号)、SSNO(社会保险号)等在数据发布时需要被隐匿;准标识符(Quasi-Identifiers, QI)属性为 $QA_i (1 \leq i \leq f)$, 如 Zipcode、Age、Sex 等用做与其他数据源连接可标识身份的属性;敏感属性为 $SA_j (1 \leq j \leq g)$, 如 Disease、Salary、PhoneNo 等需要保护的隐私属性。

2.1 基本概念

定义 1^{[10]576} 联合敏感属性。关系 T 的全部敏感属性构成一个联合敏感属性,记为 SA 。若关系 T 有 n 条记录 $r_i (1 \leq i \leq n)$, 其中 $|T|$ 为关系 T 的基数,第 j 个敏感属性作为多敏感属性的第 j 维,记做 $SA_j (1 \leq j \leq g)$, 敏感属性 SA_j 值域为 $D(SA_j)$, $|SA_j|$ 为 $D(SA_j)$ 的基数,第 i 条记录的第 j 个敏感属性值表示为 $r_i[SA_j]$ 。

定义 2^{[4]27} 联合敏感属性 l -diversity 规则。对联合敏感属性中任一敏感属性的记录集 R , 设 $\text{Max}(SA_j)$ 为 SA_j 的值在 R 出现最频繁属性,且满足 $\text{Max}(|SA_j|)/|R| \leq 1/l$ 成立, $|R|$ 为 R 中记录的个数,则 R 满足联合敏感属性 l -diversity 规则。

定义 3^{[4]28} 联合敏感属性 l -diversity 分组。关系 T 上的所有分组 $GP\{GP_1, GP_2, \dots, GP_d\}$, 若任一分组 $GP_i (1 \leq i \leq d)$ 都满足联合敏感属性 l -diversity 规则,则称 GP 是 T 上的联合敏感属性 l -diversity 分组。

定理 1 如果 T 上满足联合敏感属性 l -diversity 分组,该表的隐私泄露率不超过 $1/l$, 则发布数据是安全的。

证明 由定义 2 知,满足 l -diversity 规则的联合敏感属性任一分组,则 $\forall SA_j \in D(SA_j)$ 其敏感属性满足 $\text{Max}(|SA_j|)/|G| \leq 1/l$, 即对于该分组中记录的每一维敏感属性,攻击者都无法通过连接操作以大于 $1/l$ 的概率推断得到其真实值。所以发布该分组的数据是安全的,其泄露率不会超过 $1/l$ 。由定义 3 可知,满足联合敏感属性 l -diversity 分组的 T 中, $\forall GP_i \in GP$ 的分组其隐私泄露率 ξ_j 都小于 $1/l$, $T = \bigcup_{i=1}^d GP_i$, 则 $\frac{1}{d} \sum_{i=1}^d \xi_i \leq 1/l$, 按照联合敏感属性 l -diversity 分组原则发布数据,该表的隐私泄露率不超过 $1/l$, 即发布数据是安全的。

证毕。

为了评价发布数据的质量,本文采用文献[10]⁵⁸³ 的附加信息损失度和隐匿率作为算法发布数据质量的评价准则。

定义 4 附加信息损失度。对于 T 上满足联合敏感属性 l -diversity 分组 $GP_i (1 \leq i \leq d)$, $l \leq |GP_i|$, 附加信息损失度(Addition Information Loss, AIL)为:

$$AIL = \sum_{i=1}^n \frac{|GP_i| - l}{nl} \quad (1)$$

定义 5 隐匿率(Suppression Ratios, SR)。 $SR = n_r / |T|$, 用来衡量隐匿的记录数占关系 T 中记录总数的比例,其中 n_r 为隐匿的记录数。

显然,附加信息损失度和隐匿率越小发布数据的质量越高,理想情况下的隐匿率为 0。

2.2 联合敏感属性值多样性和语义多样性的分组策略

由于数据集 T 中记录有多个敏感属性且敏感属性值类型和属性值重要性不尽不同,在计算相似性时分组划分时应采用不同的计算准则。

联合敏感属性语义多样性和属性值多样性分组策略的关键步骤。1)设置权值参数,根据需要数据提供者可为各敏感属性值和敏感属性设置不同的权值参与相似度的计算。2)根据敏感属性值不同采取不同的计算方案,对于字符型和分类概念型属性值采用基于深度约束 WordNet 语义概念树的语义相似性计算方案;对于数值型属性值分别采用连续数值型和离散数值型的计算方法。3)计算不同记录间多敏感属性综合相似度,由分类概念型语义相似度、字符型语义相似度、基于数值型的属性值相似度加权平均进行求解。4)通过记录间多敏感属性综合相似度反聚类和联合敏感属性 l -diversity 规则进行分组划分,保证每个分组的记录同时具有联合敏感属性值多样性和语义相似性。

2.2.1 基于语义和多维加权联合敏感属性相似度计算

假设关系 T 中有任意两个记录 r_i 和 r_j 及其联合敏感属性,下面具体说明本文记录相似度计算过程。

1) 分类概念型或字符型相似度计算。

对于分类概念型或字符型敏感属性值采用基于深度约束 WordNet 语义概念树的 Wu Palme 语义相似度算法^[12],其算法表达式如下:

$$Sim_1(r_i[SA_j], r_j[SA_j]) = \frac{2 \times \omega_i \times \omega_j \times \text{depth}(p(r_i[SA_j], r_j[SA_j]))}{\omega_i \times \text{depth}(r_i[SA_j]) + \omega_j \times \text{depth}(r_j[SA_j])} \quad (2)$$

其中: $p(r_i[SA_j], r_j[SA_j])$ 代表 $r_i[SA_j]$ 和 $r_j[SA_j]$ 在语义概念树中最近的共同祖先, $\text{depth}(r_i)$ 和 $\text{depth}(r_j)$ 分别表示概念 r_i 和 r_j 在 WordNet 概念树中的深度, ω_i, ω_j 分别为 $r_i[SA_j]$ 和 $r_j[SA_j]$ 的权值。

2) 基于数值类型的属性值相似度计算。

假设给定任意两个记录 r_i, r_j 属性类型相同,属性值分别为 $r_i[SA_j]$ 和 $r_j[SA_j]$,属性权重分别为 ω_i 和 ω_j ,根据属性的数据类型不同,借鉴文献^[12]的 $r_i[SA_j]$ 和 $r_j[SA_j]$ 之间相似度计算模型,其敏感属性值相似度计算表达式如下:

$$Sim_2(r_i[SA_j], r_j[SA_j]) = \frac{L(\omega_i \times r_i[SA_j] \cap \omega_j \times r_j[SA_j])}{L(\omega_i \times r_i[SA_j] \cup \omega_j \times r_j[SA_j])} \quad (3)$$

其中 $L()$ 表示区间长度,该公式适于连续型数字区间属性值。离散型属性值相似度计算表达式为:

$$Sim_3(r_i[SA_j], r_j[SA_j]) = \frac{|\omega_i \times r_i[SA_j] - \omega_j \times r_j[SA_j]|}{\max[SA_j] - \min[SA_j]} \quad (4)$$

3) 记录为 r_i 和 r_j 联合敏感属性综合相似度函数为:

$$Sim_{\text{总}}(r_i, r_j) = \sum_{i=1}^d w_a Sim_k(r_i[SA_j], r_j[SA_j]) \quad (5)$$

其中:根据属性值类型不同在计算综合相似度时, $Sim_k(r_i[SA_j], r_j[SA_j])$ 为 Sim_1, Sim_2, Sim_3 中的一种; $w_a (1 \leq a \leq d)$ 为各敏感属性的权值。

2.2.2 联合敏感属性多样性的分组

根据记录在联合敏感属性的综合相似性和联合敏感属性 l -diversity 分组规则将记录进行反聚类,当不同记录 r_i 和 r_j 联合敏感属性综合相似度值越小二者越不相似,则将记录 r_i 和 r_j 划分在一个分组中,反聚类得到的每个聚簇作为一个 GP 分组。假如一条记录 r 被分配到分组 c 中,则 r 与分组 c 中的代表对象之间的相似性是所有分组中最小的。反聚类必然会使每个 GP 包含记录和独特敏感值个数不同,不理想情况下,反聚类会使某个分组只包含数量极少的记录,使得各 GP 的隐私保护能力差别较大。因此对只含极少记录不满足 l -diversity 限制分组或不能划分到任何分组中的记录做隐匿处理。

2.3 语义相似和多维加权多敏感属性隐私保护算法

关系 T 中的每条记录可以看做是多维敏感属性空间中的一个质点,拥有相同联合敏感属性值的记录是多维敏感属性空间中的相同点, R 为所有记录质点的集合。

输入:记录集 T ,多样性参数 l ;

输出:输出满足联合敏感属性 l -diversity 规则分组的 T^* 。

- 1) 根据关系 T 中联合敏感属性值,设置各敏感属性值及属性的权值,分别统计联合敏感属性值都不同的记录条数,选取记录条数最大的赋值给 ψ 。
- 2) 比较数据集 T 中记录条数 $|T|$ 和 $\text{Max}(|SA_j|)$ 是否大于 l ,

不满足要求则调整 l 的值。若记录数太少则结束程序。

- 3) For $i = 1$ to ψ do
 { 从 R 中随机选择任意点 r 作为分组 GP_i 中心,并将 r 从 R 中移除, R_1 表示所有剩余点的集合};
 For $i = 1$ to ψ do
 { For $j = 1$ to $|R_1|$
 { 从 R_1 中移除一个元素 r_j 与 GP_i 的记录进行综合相似度计算并插入到满足 l -diversity 条件的分组中}
 };
 };
 4) While $R_1 \neq \emptyset$ do //根据相似反聚类处理剩余的记录
 { 从 R_1 中移除任意一个元素 r ;
 For $j = 1$ to ψ do; // c_j, c_q 为存储记录的临时变量
 { IF ($\forall q \in [1, \psi]$ 且 $q \neq j$, $\sum Sim_{\text{总}}(r, c_j) \geq \sum Sim_{\text{总}}(r, c_q)$)
 { 将 r 插入到满足 l -diversity 条件的分组中}
 }
 };
 5) 对只含极少记录不满足 l -diversity 限制的分组或不能划分到任何分组中的记录做隐匿处理。
 6) 根据分组情况对关系 T^* 的准标识符属性进行匿名化处理,输出关系 T^* 。

2.4 算法复杂度分析

2.3 节算法中 1) ~ 2) 对参数 l 进行判断,不合理时提示调整 l 值,直接返回 T 表。设置权值参数和统计联合敏感属性值都不同的记录条数的时间复杂度为 $O(N)$ 。3) 初始化各参数的值,为每个 GP 组随机选择反聚类中心 r ,然后将剩余的记录按照多维加权敏感属性综合相似度大小反聚类到 ψ 个分组中(使每个 GP 中至少有 l 条记录,该过程时间复杂度为 $O((l+1) \cdot \psi)$)。4) 通过计算剩余记录 r' 与所有 GP 记录之间相似性 $Sim_{\text{总}}()$,决定将剩余的记录插入到适当的 GP 中,该过程时间复杂度为 $O(\psi \cdot |R_1|)$ 。5) 对极少的记录作隐匿处理,其复杂度为 $O(N)$ 。由于 $|T| > \psi$, $|T| > |R_1|$,所以整个算法时间复杂度可以表示为 $O(c \cdot |T|)$ 。

3 实验对比分析

3.1 实验数据及实验环境

采用 UCI 机器学习中的真实数据集 Adults (<http://archive.ics.uci.edu/ml/datasets/Adult>) 进行实验测试,对原数据集进行预处理,消除属性值中不完整的记录,从中随机提取 5000 条记录,选取 adults 中 Age, Sex, Education, Marital Status 和 race 作为 QI 属性,将 Occupation, Salary Class 和 Work Class 作为敏感属性。实验环境为 Microsoft Windows 2003 Server, MSSQL Server 2005, Intel 双核 CPU 1.66 GHz, 1 GB 内存, 120 GB 硬盘。采用 Visual Studio 2008 C#. NET 实现了本文提出的算法并进行大量实验测试,实验结果如下。

3.2 实验结果及分析

实验主要从两个条件下对算法性能指标进行分析:1) 敏感属性个数 d 、敏感属性值及敏感属性权值都固定不变,附加信息损失度随 l 值和记录条数 $|T|$ 不同时的变化规律;2) 多样性参数 l 、敏感属性值及敏感属性权值都固定不变,隐匿率随 d 值和记录条数 $|T|$ 不同时的变化规律。

在随机产生的权值参数和敏感属性个数 $d = 3$ 都固定不变的条件下,多样性参数 l 分别取 2, 3 和 4 时,随着 $|T|$ 由 1000 增加到 5000 时,其实验结果如图 1 所示。

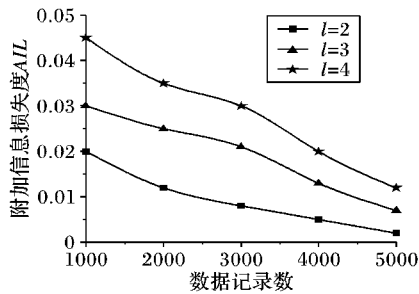


图1 附加信息损失度随 l 值和 $|T|$ 不同的变化情况

由图1知,随着 l 值和 $|T|$ 变大时,附加信息损失度变小,隐私保护程度提高。但在 d 和 $|T|$ 相同的条件下, l 的值越大,附加信息损失度也越大,这是因为数据集中各敏感属性不同取值的个数影响,使得整体的分组变差。当 l 和 d 相同时,随着 $|T|$ 的增加,各个敏感属性取值的多样化程度越来越好,使得分组的效果逐渐变好,附加信息损失度也在变小。相同条件下隐匿率变化也与附加信息损失度变化规律类似。

在随机产生的权值参数和多样性参数 $l=3$ 都固定不变,敏感属性个数 d 分别取2和3时,随着 $|T|$ 由1000增加到5000时,其实验结果如图2所示。

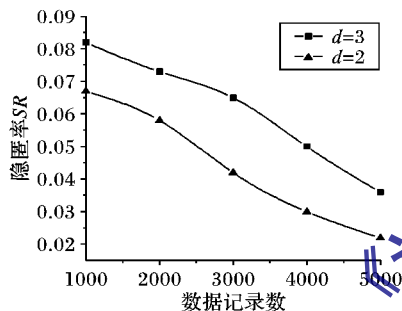


图2 隐匿率随 d 值和 $|T|$ 不同的变化情况

由图2可知,随着 d 由2变为3, $|T|$ 增大时,隐匿率变小,发布数据可用性好。但在 l 和 $|T|$ 相同的条件下,敏感属性的个数越多,隐匿率越大,这是由于复合敏感属性的维数越高,复合敏感属性每一维上都满足 l -diversity分组越困难造成的。相同条件下附加信息损失度变化也与隐匿率变化规律类似。

4 结语

隐私数据发布保护一直是安全领域研究的热点。现有的 k -匿名方法仅考虑单敏感属性数据保护,没有注意多个敏感属性、属性语义相似及属性值重要性对发布数据的影响,不适用于多敏感属性数据发布保护,若直接使用则不能达到隐私

保护的目。为此,本文提出了基于语义相似和多维加权的联合敏感属性隐私保护方法,通过多维加权的联合敏感属性语义综合相似度进行反聚类,实现了联合敏感属性值和语义多样性划分,根据数据提供者应用的需要可灵活设置多敏感属性值的权值,为数据提供不同的隐私保护力度。实验结果表明,该方法可以有效解决相似攻击等多敏感属性隐私泄露问题,增强了数据发布的实用性。隐私数据发布保护问题中数据的准标识属性与多敏感属性存在的函数依赖关系对发布数据结果的影响,是下一步研究的主要工作。

参考文献:

- [1] SWEENEY L. k -Anonymity: A model for protecting privacy [J]. International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems, 2002, 10(5): 557-570.
- [2] TRAIAN T M, BINDU V. Privacy protection: p -sensitive k -anonymity property [C]// Proceedings of the 22nd International Conference on Data Engineering. New York: ACM, 2006: 94-106.
- [3] AGGARWAL C C, YU P S. A condensation approach to privacy preserving data mining [C]// EDBT'04: the 9th International Conference on Extending Database Technology, LNCS 2992. Berlin: Springer-Verlag, 2004: 183-199.
- [4] MACHANAVAJJHALA A, GEHRKE J, KEFER D. l -diversity: Privacy beyond k -anonymity [C]// ICDI 2006: Proceedings of the 22nd International Conference on Data Engineering. Atlanta, Georgia: ACM, 2006: 24-35.
- [5] LI JINCHUI, LI TIANCHENG, VENKATASUBRAMANIAN S. l -Closeness: Privacy beyond k -anonymity and l -diversity [C]// ICDE'07, the 23rd International Conference on Data Engineering. Istanbul, Turkey: IEEE Computer Society, 2007: 106-115.
- [6] XIAO XIAOKUI, TAO YUFEI. Personalized privacy preservation [C]// SIGMOD'06: the 25th ACM SIGMOD International Conference on ACM Management of Data. Chicago, Illinois, USA: ACM, 2006: 229-240.
- [7] 王茜,曾子平. (p, a) -sensitive k -匿名隐私保护模型[J]. 计算机应用研究, 2009, 26(6): 2177-2179.
- [8] 杨晓春,刘向宇,王斌,等. 支持多约束的 K -匿名化方法[J]. 软件学报, 2006, 17(5): 1222-1231.
- [9] 祁瑞丽,王可,郭学涛,等. 基于最大叶子子树优先策略的多敏感属性保护方法[J]. 燕山大学学报, 2009, 33(5): 432-437.
- [10] 杨晓春,王雅哲,王斌,等. 数据发布中面向多敏感属性的隐私保护方法[J]. 计算机学报, 2008, 32(4): 574-587.
- [11] BUDANITSKY A, GRAEME H. Evaluating WordNet based measures of semantic distance [J]. Computational Linguistics, 2006, 32(1): 13-47.
- [12] 刘平峰. 基于知识网络的电子商务智能推荐理论方法研究[D]. 武汉: 武汉理工大学, 2006.

(上接第924页)

- [3] 史龙,王福豹,段渭军,等. 无线传感器网络 Range-Free 自身定位机制与算法[J]. 计算机工程与应用, 2004, 40(23): 127-130
- [4] NICULESCU D, NATH B. DV based positioning in Ad Hoc networks[J]. Journal of Telecommunication Systems, 2003, 22(1/4): 267-280.
- [5] 孙美玲. 基于遗传算法的无线传感器网络节点自身定位算法研究[D]. 北京: 中国石油大学, 2009: 16-19.
- [6] 林金朝,刘海波,李国军,等. 无线传感器网络中 DV-Hop 节点定位改进算法研究[J]. 计算机应用研究, 2009, 26(4): 1272-1275.
- [7] 赵清华,刘少飞,张朝霞,等. 一种无需测距节点定位算法的分析和改进[J]. 传感技术学报, 2010, 23(1): 122-127.

- [8] EUSUFF M M, LANSEY K E. Optimization of water distribution network design using the shuffled frog leaping algorithm [J]. Water Resources Planning and Management, 2003, 129(3): 210-225.
- [9] MOHAMMAD B A, MAROOSI F A. Application of shuffled frog-leaping algorithm on clustering[J]. The International Journal of Advanced Manufacturing Technology, 2009, 45(1/2): 199-209.
- [10] 杨祖元,徐姣,罗兵,等. 基于 SFLA-FCM 聚类的城市交通状态判别研究[J]. 计算机应用研究, 2010, 27(5): 1743-1745.
- [11] 郑仕健,楼才义,杨小牛. 基于改进混合蛙跳算法的认知无线电协作频谱感知[J]. 物理学报, 2010, 59(5): 3611-3616.
- [12] 嵇玮玮,刘中. DV-Hop 定位算法在随机传感器网络中的应用研究[J]. 电子与信息学报, 2008, 30(4): 970-974.