

基于 paillier 的隐私保护下关联规则挖掘方法

邢欢, 张琳

(南京邮电大学计算机学院, 江苏 南京 210003)

摘要: 在基于隐私保护的数据挖掘中, 挖掘的精度和安全性一直是一对矛盾, 针对该矛盾提出了一种在分布式的环境下基于 paillier 同态加密的关联规则挖掘方法, 该方法将计算方和解密方分离保证了数据挖掘的安全性, 同时基于同态加密并不会影响挖掘的精度, 并通过蒙哥马利算法降低了 paillier 算法的开销, 经过实验发现在增加解密过程后算法开销还是在能接受的范畴内。

关键词: 隐私保护; 关联规则; 同态加密

中图分类号: TP311

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2016.00014

Privacy-preserving mining of association rules based on paillier encryption algorithm

XING Huan, ZHANG Lin

(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In privacy preserving association rule mining, the precision and security of mining are always a pair of contradictions. A method of privacy-preserving mining of association rules based on paillier encryption algorithm over distributed databases was proposed. The method separated calculation and decryption so it can solve the problem of accuracy and security from mining of association rules perfectly. The method can reduce the time cost by Montgomery reduction. The experiment shows that the time cost on the basis of adding the process of encryption and decryption is acceptable.

Key words: privacy preserving, association rules, homomorphic encryption

1 引言

目前在大数据的浪潮下促进了一大群学者针对数据挖掘进行研究, 然而在数据的挖掘过程中, 隐私保护的问题日益突出。当年啤酒尿布的经典案例引发了人们对于关联规则挖掘的兴趣。而关联规则挖掘也正是数据挖掘中十分重要的一个领

域, 关联规则挖掘算法如 Apriori 算法、FP 增长算法等经过多年的研究和改进已比较成熟。然而在如今的信息时代, 关联规则挖掘早已不再是单方的数据挖掘, 正是由于多方数据挖掘, 隐私保护也越显重要。关联规则挖掘存在的隐私保护问题主要有如下两种情况: 数据提供方并不是关联规则挖掘方, 所以数据提供方希望在挖掘出正确

收稿日期: 2015-10-19; 修回日期: 2015-12-30。通信作者: 张琳, zhangl@niupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61402241); 江苏省自然科学基金资助项目 (No.BK2012436); 省属高校自然科学研究重大基金资助项目 (No.12KJA520002, No.14KJA520002); 江苏省高校自然科学基金资助项目 (No.13KJB520017)

Foundation Items: The National Natural Science Foundation of China (No.61402241), The Natural Science Foundation of Jiangsu Province (No.BK2012436), Natural Science Key Fund for Colleges and Universities of Jiangsu Province (No.12KJA520002, No.14KJA520002), The Natural Science Fund for Colleges and Universities of Jiangsu Province (No.13KJB520017)

结果的同时并不会泄露提供数据中的隐私内容；另一种情况主要是多个数据提供方想要共同进行数据挖掘，然而多方之间也不想泄露隐私数据信息。针对以上的隐私保护问题，已经有很多学者提出了相关的协议以及算法。文献[1,2]提出了基于差分隐私的算法，差分隐私保护最大的优点是背景知识无关，然而差分隐私保护却不可避免数据失真，改进也只是将失真程度减小到最低。文献[3]利用随机响应技术，在概率的基础上对数据进行挖掘，虽然在数据挖掘的准确性和安全性方面做了很好的平衡，但也无法完美地解决这两方面之间的问题。文献[4]基于 FDM (fast distributed mining of association rules) 算法提出了针对上述第 2 种情况下频繁项集挖掘的协议以及算法。文献[5,6]都提出了基于遗传算法的隐私保护下的关联规则挖掘算法，然而算法在挖掘精度方面并不完美。文献[7]提出了一种基于重构数据分布和添加数据干扰实现隐私保护下关联规则的挖掘算法，但同时由于数据的干扰也导致了数据属性的相关性遭到破坏，数据挖掘精度不够。文献[8]针对多方进行频繁项集挖掘时存在恶意挖掘其他方信息的情况，提出了一种新的算法，但是算法开销稍高。文献[9]根据不同的安全等级设置相应的数据干扰来实现数据挖掘中的隐私保护。文献[10]提出了一种基于 FDM 的水平分布式协议，具有较好的通信效率和安全性。文献[11]针对多方协作数据挖掘隐私泄露问题提出了一种二进制整数规划模型。文献[12]针对两方协作关联规则挖掘提出了基于可交换加密的协议。文献[13]提出了一种基于泛化实现隐私保护的算法，然而使用泛化不可避免地使数据挖掘的精度受到影响。

在基于隐私保护的关联规则挖掘中往往数据挖掘的安全性和挖掘精度是一对矛盾。在使用同态加密的情况下，关联规则挖掘的过程既能保证数据的安全性又能保证数据挖掘的精度。唯一的不足是算法开销稍高，然而如今最重要的问题是数据挖掘的精度和安全性、算法和技术的改进以及硬件设备的提升，而算法所需的时间消耗相对没有数据挖掘精度和安全性那么重要。因此，本文提出了一种基于 paillier 同态加密的关联规则挖掘方法。

2 相关定义

2.1 关联规则

事务集为 D ，项集集合为 I ，事务 $t \in T$ 由 I 中元素组成，支持数是项集 A 在事务集 D 中的计数 $|A|$ ，支持度是项集 A 的支持数在事务集中的比例 $\text{sup}(A) = \frac{|A|}{|D|}$ ，最小支持度为 min sup ，若 $\text{sup}(A) \geq \text{min sup}$ ，则 A 是频繁项集。关联规则是形如 $A \Rightarrow B$ 的表达式， A 是规则前件， B 是规则后件。置信度是指项集 A 和项集 B 的支持度与项集 A 的支持度的比值

$$\text{conf}(A \Rightarrow B) = \frac{\text{sup}(A \cup B)}{\text{sup}(A)} \quad (1)$$

关联规则成立的条件是：

1) $A \subseteq I, B \subseteq I$ 且 $A \cap B = \emptyset$ ；

2) $\text{sup}(A \cup B) \geq \text{min sup}$ ；

3) $\text{conf}(A \Rightarrow B) \geq \text{min conf}$ ，其中 min sup 和 min conf 是最小阈值。

2.2 Paillier 同态加密算法

Paillier 是基于合数阶剩余类的概率公钥密码系统，其理论安全性构建在比“RSA 假设”更强的“CCRA 假设”上。Paillier 加密算法具有加法同态和混合乘法同态的特性。在涉及到多方关联规则挖掘时，由于各方都不希望泄露自身的数据，所以可以将项集的支持数等数据特征进行加密。关联规则挖掘的关键在于频繁项集的挖掘，也就是全局支持数的计算。因此关键在于累加各方的项集支持数，而计算过程不应泄露数据信息，所以可以利用 paillier 加密算法的加法同态特性。

Paillier 加密算法的加密过程为

$$c = g^m r^n \bmod n^2 \quad (2)$$

其中， $m \in Z_n$ 是明文， $r \in Z_n^*$ ， $g \in G$ ， $G = \{w | w \in Z_{n^2}^*\}$ ， G 是模 n^2 的乘法群，且满足 $\text{gcd}(L(g^{\lambda} \bmod n^2), n) = 1$ 。解密过程为

$$m = \frac{L(c^{\lambda} \bmod n^2)}{L(g^{\lambda} \bmod n^2)} \bmod n \quad (3)$$

公钥对为 (g, n) ，私钥是 λ ，其中 p, q 是 2 个大素数， $n = pq, \lambda = \text{lcm}(p-1, q-1)$ 。Paillier 加密算法满足加法同态，所以有

$$E_p(x; r_1) E_p(y; r_2) = E_p(x + y; r_1 r_2) \quad (4)$$

3 算法描述

3.1 多方参与的关联规则挖掘方案

问题：存在多个数据提供方，在不泄露自身隐私数据的条件下共同挖掘关联规则。

第 i 个数据提供方记为 P_i ，事务集为 D_i ，其事务集的规模为 $|D_i|$ ，第 i 个数据提供方的局部支持数为 $\minsup_i$ ，全局最小支持度为 $\minsup$ 。

关键在于挖掘全局频繁项集，即如何根据每个数据提供方 P_i 的候选项集 $k - C^m$ 的局部支持数 $\sup_i(k - C^m)$ 在不泄露给其他人的情况下获取全

$$\text{局支持度} = \frac{\sum_{i=1}^n \sup_i(k - C^m)}{\sum_{i=1}^n |D_i|}, \text{ 即只要满足}$$

$$\frac{\sum_{i=1}^n \sup_i(k - C^m)}{\sum_{i=1}^n |D_i|} \geq \minsup \quad (5)$$

则将候选 k 项集 $k - C^m$ 添加到频繁 k 项集 $k - L$ 中。由式(5)得到

$$\sum_{i=1}^n (\sup_i(k - C^m) - \minsup |D_i|) \geq 0 \quad (6)$$

所以只需要考虑 P_i 的 $\sup_i(k - C^m) - \minsup |D_i|$ ，若 P_i 直接公布 $\sup_i(k - C^m) - \minsup |D_i|$ 来求和，则有风险被推测出 $\sup_i(k - C^m)$ 和 $|D_i|$ 。因此，需要将 $\sup_i(k - C^m) - \minsup |D_i|$ 使用 paillier 加密后再公布。每个数据提供方 P_i 都有公钥对 (g, z) ，而只有 P_a 有私钥 λ ， P_b 为数据计算方。 P_i 将 $\sup_i(k - C^m) - \minsup |D_i|$ 加密得到 $E_p(\sup_i(k - C^m) - \minsup |D_i|)$ ，并将其发给 P_b ， P_b 计算 $\prod_{i=1}^n E_p(\sup_i(k - C^m) - \minsup |D_i|)$ ，由于 $\sup_b(k - C^m) - \minsup |D_b|$ 对于 P_a 是未知的，所以将 $\sup_b(k - C^m) - \minsup |D_b|$ 加密后再乘以 $\prod_{i=1}^n E_p(\sup_i(k - C^m) - \minsup |D_i|)$ ，即 $\prod_{i=1}^n E_p(\sup_i(k - C^m) - \minsup |D_i|) E_p(\sup_b(k - C^m) - \minsup |D_b|)$ ，再将此结果发给 P_a ， P_a 通过私钥解密得到

$$\sum_{i=1}^n (\sup_i(k - C^m) - \minsup |D_i|) + \sup_b(k - C^m) - \minsup |D_b| = \text{sum } C \quad (7)$$

此时 $\text{sum } C$ 只有 P_a 知道， $\sup_b(k - C^m) - \minsup |D_b|$ 只有 P_b 知道，需要在各不泄露数据的情况下比较 $\text{sum } C$ 和 $\sup_b(k - C^m) - \minsup |D_b|$ ，若 $\text{sum } C \geq \sup_b(k - C^m) - \minsup |D_b|$ ，即满足式(5)，则 $k - C^m$ 是频繁 k 项集，这里的比较方法就是著名的百万富翁问题，文献[8]已经提出了一种改进的简洁高效的方法，本文使用该方法来处理此处的比较问题。

通过上述方法即可安全挖掘多方频繁项集，在已有频繁项集的基础上再通过 ap-genrules 产生关联规则，为方便描述，将关联规则用 $A \Rightarrow B$ 表示，而判断该关联规则是否是所需要的还需要进行最关键的一步计算： $\text{conf}(A \Rightarrow B) \geq \min \text{conf}$ ， $\min \text{conf}$ 是全局最小置信度阈值。

$$\frac{\sup(A \cup B)}{\sup(A)} \geq \min \text{conf} \Rightarrow \sum_{i=1}^n |A \cup B|_i \geq \min \text{conf} \sum_{i=1}^n |A|_i \Rightarrow \sum_{i=1}^n (|A \cup B|_i - \min \text{conf} |A|_i) \geq 0 \quad (8)$$

其中， $|A \cup B|_i$ 和 $|A|_i$ 是 P_i 上项集 $A \cup B$ 和 A 的支持数。此时方法类似于上述全局频繁项集的生成方法，每个数据提供方 P_i 使用公钥对对 $|A \cup B|_i - \min \text{conf} |A|_i$ 加密并将加密结果发给 P_b ， P_b 计算 $(\prod_{i=1}^n E_p(|A \cup B|_i - \min \text{conf} |A|_i)) E_p(|A \cup B|_b - \min \text{conf} |A|_b)$ ，并将结果再发给 P_a ， P_a 通过私钥解密得到

$$\sum_{i=1}^n (|A \cup B|_i - \min \text{conf} |A|_i) + |A \cup B|_b - \min \text{conf} |A|_b = \text{sum Conf} \quad (9)$$

P_a 将 sum Conf 和 P_b 的 $|A \cup B|_b - \min \text{conf} |A|_b$ 使用文献[14]的方法进行比较，若 $\text{sum Conf} \geq |A \cup B|_b - \min \text{conf} |A|_b$ ，则 $A \Rightarrow B$ 是有趣的关联规则。频繁项集挖掘和关联规则挖掘过程中的主要流程相同，如图1所示。

3.2 计算优化

paillier 加密算法中大素数 p 、 q 的生成和私钥 $\lambda = \text{lcm}(p-1, q-1)$ 的生成都很容易。关键在于获取 g ，且满足 $g \in G, G = \{w | w \in Z_{n^2}^*\}$ ， G 是模 n^2 的乘法群， $\text{gcd}(L(g^\lambda \bmod n^2), n) = 1$ 。获取 g 的方法如下。

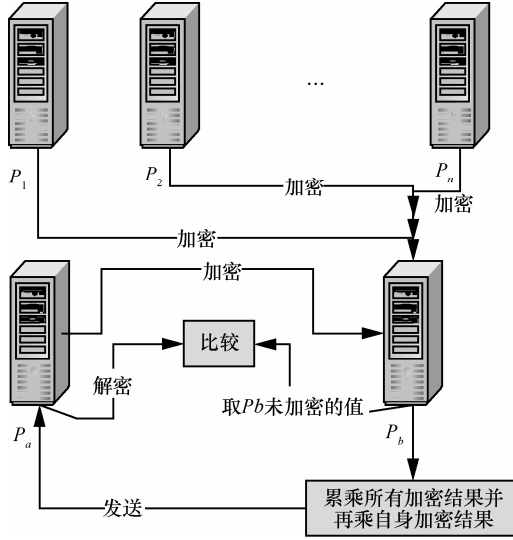


图1 多方规则挖掘时的同态加解密过程

随机生成 $k \in Z_n^*$, $k' \in Z_n$, 且 $k' \lambda \bmod n = k$, $g = k'n + 1$ 。证明过程如下。

$$1) k' \in Z_n, g = k'n + 1$$

$$\Rightarrow 0 < g < n^2, \gcd(g, n) = 1$$

$$\Rightarrow 0 < g < n^2, \gcd(g, n^2) = 1$$

$$\Rightarrow g \in \{w | w \in Z_{n^2}^*\}$$

$$2) g^\lambda \bmod n^2$$

$$= (k'n + 1)^\lambda \bmod n^2$$

$$= (1 + C_\lambda^1 k'n + C_\lambda^2 (k'n)^2 + \dots + C_\lambda^\lambda (k'n)^\lambda) \bmod n^2$$

$$= (1 + \lambda k'n \bmod n^2) \bmod n^2$$

$$\lambda k'n - \lambda k'n \bmod n^2 = Kn^2$$

$$\Rightarrow \lambda k'n \bmod n^2 = \lambda k'n - Kn^2$$

$$= (\lambda k' - Kn)n = ((\lambda k') \bmod n)n$$

$$g^\lambda \bmod n^2 = (1 + \lambda k'n \bmod n^2) \bmod n^2$$

$$= (1 + ((\lambda k') \bmod n)n) \bmod n^2$$

$$= (1 + kn) \bmod n^2$$

$$= 1 + kn$$

$$L(g^\lambda \bmod n^2) = \frac{g^\lambda \bmod n^2 - 1}{n} = k, \text{ 又由于 } k \in$$

$Z_n^* \Rightarrow \gcd(k, n) = 1$, 故有 $\gcd(L(g^\lambda \bmod n^2), n) = 1$ 。

Paillier 加密算法最大的消耗在于 $g^m \bmod n^2$ 这样的模幂运算, 因为一旦 g^m 超过 64 bit, 就需要创建大数运算库, 但根据模运算法则 $g \cdot g \bmod n^2 = (g \bmod n^2 \cdot g \bmod n^2) \bmod n^2$, paillier 算法就需要大量的模运算, 开销是巨大的。使用蒙哥马利算法, 通过不断的降阶, 将模幂运算转化为少数的模乘运算来降低算法开销。算法如下。

步骤1 初始化 $D=1$ 。

步骤2 指数 m 为偶数时, $m = \frac{m}{2}, g = g \cdot g \bmod n^2$ 。

步骤3 指数 m 为奇数时, $m = m - 1, D = Dg \bmod n^2$ 。

步骤4 重复步骤3和步骤4直到 $m = 0$, return D 。

算法流程如图2所示, 加解密过程中的模幂运算全部转化为模乘运算, 实验显示使用蒙哥马利算法能将计算时间降到极低的标准, 使 paillier 同态加密算法的开销达到令人满意的程度。

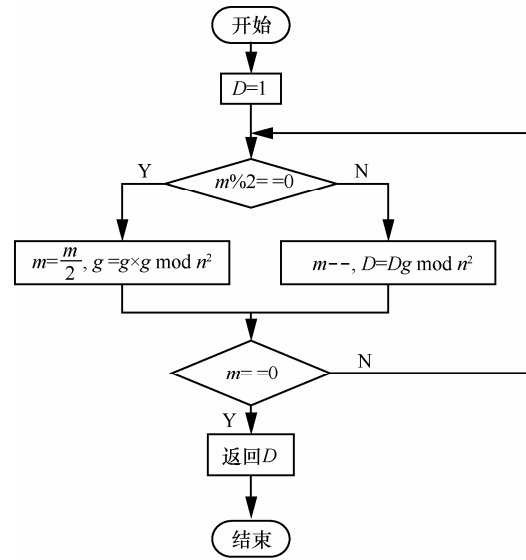


图2 模幂运算转化成模乘运算

4 算法设计

4.1 频繁项集挖掘算法

1) 每个数据提供方 P_i 由 Apriori-gen 算法根据全局 $k-1$ -频繁项集 L_{k-1} 得到本地候选 k -频繁项集的集合 $k-C'_i$, 并由本地支持数 $\min \sup_i$ 剪枝得到 $k-C_i$ 。

2) P_i 公布自己的 $k-C_i$, 获得 $\bigcup_i k-C_i$, P_i 对 $\bigcup_i k-C_i$ 中每个项集计算 $\sup_i(k-C_i) - \min \sup |D_i|$, 使用公钥对 (g, z) 加密, 其中, $z = pq$, p 和 q 是大素数, 对其加密得到 $E_p(\sup_i(k-C_i) - \min \sup |D_i|)$, 并将加密结果发送给 P_b 。

3) P_b 计算 $\prod_1^n E_p(\sup_i(k-C_i) - \min \sup |D_i|)$,

P_b 通过公钥对 (g, z) 加密 $\sup_b(k - C_b^m) - \min \sup |D_b|$ 得到 $E_p(\sup_b(k - C_b^m) - \min \sup |D_b|)$, 再计算 $\prod_1^n E_p(\sup_i(k - C_i^m) - \min \sup |D_i|) E_p(\sup_b(k - C_b^m) - \min \sup |D_b|)$, 并将结果发送给 P_a 。

4) P_a 根据 $\frac{L(c^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n$ 解密得到 $\sum_1^n (\sup_i(k - C_i^m) - \min \sup |D_i|) + \sup_b(k - C_b^m) - \min \sup |D_b| = \text{sum } C$, 其中, $c = \prod_1^n E_p(\sup_i(k - C_i^m) - \min \sup |D_i|) E_p(\sup_b(k - C_b^m) - \min \sup |D_b|)$ 。

5) 使用文献[8]百万富翁问题中使用的部分标量积保密计算方法比较 P_a 的 $\text{sum } C$ 和 P_b 的 $\sup_b(k - C_b^m) - \min \sup |D_b|$, 如果 $\text{sum } C > \sup_b(k - C_b^m) - \min \sup |D_b|$, 将 $k - C_b^m$ 添加到 $k - L$ 。

6) 频繁 1 项集的获取由 P_i 根据局部最小支持度 $\min \sup_i$ 得到局部候选 1 项集 $1 - C_i$, 再循环第 3) 步至第 5) 步得到频繁 1 项集的集合 $1 - L$ 。

4.2 关联规则挖掘算法

1) 针对每一个频繁 k 项集, 使用 ap-genrules 算法产生候选关联规则 $A \Rightarrow B$, 每个数据提供方 P_i 使用公钥对 (g, z) 对 $|A \cup B|_i - \min \text{conf} |A|_i$ 加密, 并将加密结果 $E_p(|A \cup B|_i - \min \text{conf} |A|_i)$ 发给 P_b 。

2) P_b 计算 $(\prod_1^n E_p(|A \cup B|_i - \min \text{conf} |A|_i)) \cdot E_p(|A \cup B|_b - \min \text{conf} |A|_b)$, 并将结果再发给 P_a 。

3) P_a 通过私钥 λ 解密得到 $\sum_1^n (|A \cup B|_i - \min \text{conf} |A|_i) + |A \cup B|_b - \min \text{conf} |A|_b = \text{sum Conf}$, P_a 将 sum Conf 和 P_b 的 $|A \cup B|_b - \min \text{conf} |A|_b$ 使用文献[8]的方法进行比较, 若 $\text{sum Conf} \geq |A \cup B|_b - \min \text{conf} |A|_b$, 则 $A \Rightarrow B$ 是有趣的关联规则。

5 仿真实验

本文的方法基于同态加密技术, 在关联规则挖掘前并没有更改数据集, 所以并不会影响关联规则挖掘的结果。本文方法的安全性是基于 paillier 加密算法的, paillier 算法的安全性不弱于

“RSA 假设”, 并且 paillier 加密算法是语义安全的, 即不存在任何多项式时间算法能够区分不同值的加密结果。所以, 本文的关联规则挖掘的安全性和挖掘精度问题都是能够保障的。主要通过实验来测试本文算法开销。

实验是由 100 个项组成的项集和 10 000 条事务。将 10 000 条事务分成 5 组, 相当于每个数据方有 2 000 条数据。即仿真 5 个数据方在隐私保护的前提下共同挖掘关联规则。算法最主要的消耗在于加解密过程, 本文对不同密钥长度下 paillier 加密时间与本文改进了 paillier 的加解密计算过程之后的加密时间进行对比。不同密钥位数的算法开销对比如图 3 所示。其中密钥长度为: 16、32、64、128、256、512、1 024、2 048。而当密钥长度大于 1 024 时, 安全性就相当高, 从图 3 可以看出本文方法在密钥长度大于 1 024 时, 加密时间能节省 100 ms 以上, 而在关联规则挖掘过程中, 频繁项集的挖掘都是呈指数增长的, 所以本文方法有较大优势。

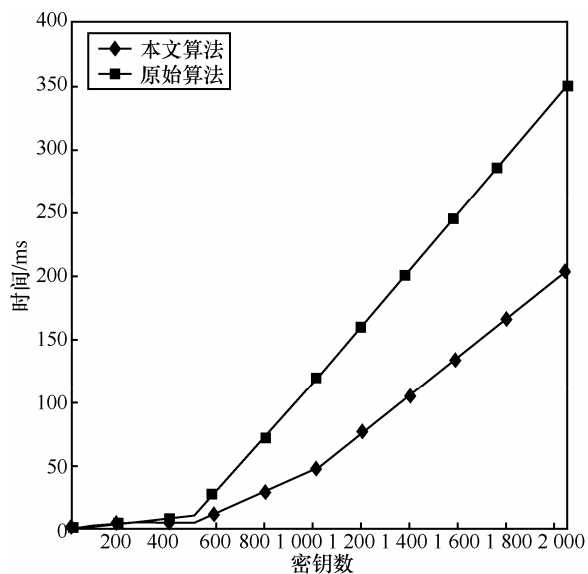


图3 不同密钥数的算法开销对比

在关联规则的挖掘精度、安全性以及算法开销上, 本文基于 paillier 同态加密的算法与文献[5]在分布式环境下基于遗传算法的关联规则挖掘算法以及文献[3]的使用随机响应技术的经典算法进行了对比。在上述介绍的实验环境下对 3 种方法进行对比, 为方便描述, 将文献[5]的算法记为 GADM 算法, 将文献[3]算法记为 RRDM 算法。实验首先将 10 000 条事务在 5 组不同的最小支持

度和最小置信度阈值下使用 Apriori 算法进行关联规则挖掘,这5组值如表1所示。然后分别使用本文算法、GADM算法以及RRDM算法在这5组最小支持度和最小置信度阈值下进行挖掘,其中规则准确率比较结果如图4所示。显然GADM和RRDM规则挖掘的准确率并不是很完美,而由于本文使用了加密解密的形式,并不会影响最后规则的挖掘。3种算法的时间消耗对比如图5所示,其中本文算法分别使用512 bit、1024 bit、2048 bit 密钥进行对比。

表1 测试数据

组数	min sup/%	min conf/%
1	1	50
2	3	60
3	5	60
4	7	50
5	10	60

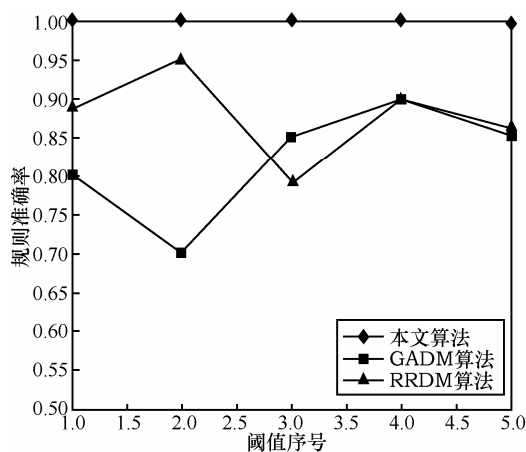


图4 规则准确率对比

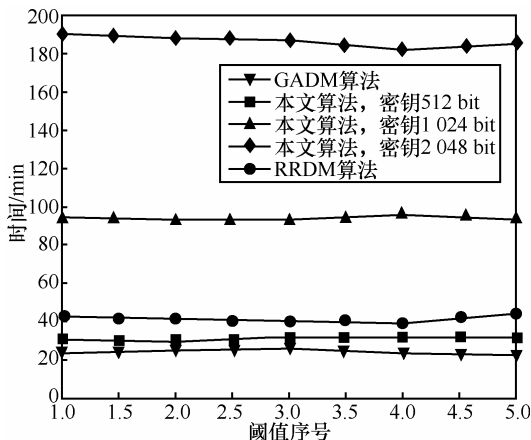


图5 在使用不同秘钥时的时间对比

由此可见,当密钥数为512时,与GADM算法消耗相比是可接受的。其中RRDM是矩阵计算,矩阵规模大算法开销同样巨大,而且基于随机响应的方法存在一定的概率把扰乱后的数据还原,所以安全性并不高。然而本文算法在挖掘的精度和安全性方面却比GADM算法和RRDM算法有较大优势。

6 结束语

目前针对数据挖掘的隐私保护已经有了很多研究,但是数据挖掘的精度和数据挖掘过程中的安全性一直是一对矛盾。本文提出了一种基于paillier同态加密的能够实现隐私保护关联规则挖掘方法,能很好地解决挖掘精度和安全性问题,虽然增加了加解密过程,但是通过蒙哥马利算法提高了加解密性能,最终实验显示算法消耗还是可接受的。并且本文使用的paillier加密算法的安全性是建立在比“RSA假设”更强的“CCRA假设”上,所以相比于其他多方安全计算中涉及到的加密算法更安全。但是本文方法的算法开销稍大,这是以后需要继续深入研究的内容。

参考文献:

- [1] 丁丽萍, 卢国庆. 面向频繁模式挖掘的差分隐私保护综述[J]. 通信学报, 2014, 35(10): 200-209.
DING L P, LU G Q. Survey of differential privacy in frequent pattern mining[J]. Journal on Communication, 2014, 35(10): 200-209.
- [2] 张啸剑, 孟小峰. 面向数据发布和分析的差分隐私保护[J]. 计算机学报, 2014, 37(4): 927-949.
ZHANG X F, MENG X F. Differential privacy in data publication and analysis[J]. Chinese Journal of Computers, 2014, 37(4): 927-949.
- [3] SUN C J, FU Y, ZHOU J L, et al. Personalized privacy-preserving frequent itemset mining using randomized response[J]. The Scientific World Journal, 2014(3): 686151.
- [4] TASSA T. Secure mining of association rules in horizontally distributed databases[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(4): 970-983.
- [5] KESHAVAMURTHY B H, KHAN A M, TOSHNIWAL D. Privacy preserving association rule mining over distributed databases using genetic algorithm[J]. Neural Comput & Applic, 2013, 22(1): 351-364.
- [6] LIN C W, ZHANG B B, YANG K T, et al. Efficiently hiding sensitive itemsets with transaction deletion based on genetic algorithms[J]. The Scientific World Journal, 2014: 398269.

- [7] ANEKRITMONGKOL S, KASEMSAN K. SQL model in language encapsulation and compression technique for association rules mining[J]. International Journal of Intellectual Property Management, 2013, 4(1): 65-75.
- [8] SEKHAVAT Y A, FATHIAN M. Mining frequent itemsets in the presence of malicious participants[J]. The Institution of Engineering and Technology, 2010, 4(2): 80-92.
- [9] LI Y P, CHEN M H, LI Q W, et al. Enabling multilevel trust in privacy preserving data mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(9): 1598-1612.
- [10] TASSA T. Secure mining of association rules in horizontally distributed databases[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(4): 970-983.
- [11] BHANUMATHI S, SAKTHIVEL A. A new model for privacy preserving multiparty collaborative data mining[C]//International Conference on Circuits, Power and Computing Technologies. c2013: 845-850.
- [12] ZHANG F, RONG C M, ZHAO G, et al. Privacy-preserving two-party distributed association rules mining on horizontally partitioned data[C]//International Conference on Cloud Computing and Big Data. c2013: 633-640.
- [13] HAJIAN S, DOMINGO-FERRER J, FARRÀS O. Generalization-based privacy preservation and discrimination prevention in data publishing and mining[J]. Data Mining & Knowledge Discovery, 2014, 28: 1158-1188.
- [14] 李顺东, 王道顺. 基于同态加密的高效多方保密计算[J]. 电子学报, 2013, 41(4): 798-803.
- LI S D, WANG D S. Efficient secure multiparty computation based on homomorphic encryption[J]. Acta Electronica Sinica, 2013, 41(4): 798-803.

作者简介:



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、数据挖掘、隐私保护等。



张琳 (1980-), 女, 江苏丰县人, 博士后, 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、隐私保护等。