

LBS 连续查询的匿名序列规则挖掘方法研究

陈泽伟¹, 张海涛²

(1. 南京邮电大学 通信与信息工程学院, 江苏 南京 210003;
2. 南京邮电大学 地理与生物信息学院, 江苏 南京 210046)

摘 要:随着 LBS 的深入发展与广泛应用, 隐私保护成为 LBS 深入发展中亟待解决的关键技术问题。时空 K -匿名是 LBS 隐私保护的主要类型, 当前研究尚未涉及匿名集数据的可用性和隐私保护的安全性。针对上述问题, 基于匿名集数据具有时空序列的特性, 提出了一种基于双向不可逆扩展的匿名集序列规则挖掘算法。该算法在扫描序列数据库的过程中, 对相应的项集进行位置标记, 从而保证了对序列数据库一次扫描即能挖掘出用户移动的序列规则。通过对频繁模式进行扩展并发现用户的移动规律、行为模式, 对所提出的算法进行了验证实验及其结果分析。实验结果表明, 所提出算法的挖掘结果会涉及到敏感区域, 如军事领域等, 因此对于实现 LBS 位置隐私保护具有重要的实践意义, 对于丰富隐私保护数据挖掘领域的研究具有一定的理论价值。

关键词:位置服务; 位置隐私保护; 时空 K -匿名; 序列规则

中图分类号: TP301

文献标识码: A

文章编号: 1673-629X(2017)06-0124-06

doi: 10.3969/j.issn.1673-629X.2017.06.026

Investigation on Anonymous Sequential Rules Mining Method with LBS Continuous Query

CHEN Ze-wei¹, ZHANG Hai-tao²

(1. College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

2. College of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

Abstract: With the deep development and wide use of Location-Based Services (LBS), privacy protection has become the key technology to be solved urgently in LBS. The temporal and spatial K -anonymity is the main type of LBS privacy protection, in which the availability of anonymous datasets and the security of privacy protection have not been involved so far. Aimed at this problem and found on a characteristic of spatial-temporal sequences in anonymous dataset, a mining algorithm of anonymity dataset sequence rules has been presented with bidirectional irreversible expansion, which has marked the position of item sets in the process of scanning the sequence database to guarantee mining mobile sequential rules at just one scan of sequence database. Experiments of data mining and result validation have been conducted. Result mined by the algorithm covers sensitive regions like military possessions, which has been proved to have important practical value for realizing LBS privacy protection and certain theoretical value for enriching the study in privacy protection data mining area.

Key words: location based service; location privacy protection; spatial temporal K -anonymity; sequence rules

0 引言

随着 LBS 的深入发展与广泛应用^[1], 因 LBS 引发的隐私泄漏问题日益严重。一些位置隐私泄露事件 (例如恶意的手机软件、手机定位广告等) 引起了公众的广泛关注。隐私保护也成为 LBS 发展过程中亟待

解决的关键问题^[2]。

2003 年, 由 Gruteser 等提出的基于时空 K -匿名的 LBS 隐私保护方法^[3] (简称时空 K -匿名), 以匿名数据的真实可用、方法实现简洁灵活以及更适合 LBS 移动计算环境等特点, 成为近年来研究的主流方向。

收稿日期: 2016-06-06

修回日期: 2016-09-15

网络出版时间: 2017-04-28

基金项目: 国家自然科学基金资助项目 (41201465)

作者简介: 陈泽伟 (1993-), 女, 硕士生, 研究方向为移动大数据技术; 张海涛, 副教授, 研究方向为移动智能地理信息系统。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20170428.1702.020.html>

时空 K -匿名的性能优化主要在快照查询与连续查询两个方面展开。快照查询包括 4 个方面:

- (1)灵活设定隐私保护级别。文献[4]提出了动态感知移动用户时空分布设定 K 值的 Clique-Cloak 方法;文献[5]提出了时空区域 Footprint 的概念,并基于 Footprint 设计了动态设置更人性化的隐私级别的保护方法。
- (2)增强型查询标识保护。时空 K -匿名的匿名集与匿名查询请求为 1:1 关系,而 Clique-Cloak 方法要求匿名集的用户均应提出查询请求,但 Clique-Cloak 方法采用无向图结构生成匿名集会产​​生计算量过大的问题,只适合较小 K 值的匿名保护。
- (3)多模式查询的隐私保护。文献[6]改变时空 K -匿名方法,同时进行查询隐私与位置隐私保护。
- (4)空间网络与分布式传感网的应用,设计了适合道路网络的时空 K -匿名^[7]。

上述时空 K -匿名及优化方法均没有考虑针对匿名集敏感信息模式的隐私攻击问题。这一类攻击确实有存在的可能性:在实际 LBS 的应用中,LBS 服务提供商以及应用第三方,通常会逐渐累积形成具有较大时空跨度的大量匿名集数据。同时,也发现了针对此类数据的关联分析,可能对 LBS 用户产生更具威胁性的隐私推理攻击:发现用户的行为模式,并基于行为模式进行隐私推理攻击。现有方法没有对大量匿名集数据进行分析,没有对匿名集数据的可用性以及隐私保护安全性进行深入研究。

为此,基于对匿名数据特性以及传统的序列规则挖掘方法的分析,提出了一种基于双向不可逆扩展的匿名集序列规则挖掘方法,详细描述了算法步骤。实验首先模拟生成匿名集序列数据,使用基于双向不可逆扩展方法的匿名集序列规则挖掘方法进行数据挖掘,验证算法的有效性。通过结合实际地理环境数据的应用效果,对实验结果的分析,发现传统的时空 K -匿名方法存在隐私保护安全漏洞问题,挖掘结果涉及多处军事敏感区域,因此该算法对于分析隐私攻击推理和用户隐私安全保护研究领域有重要的意义。

1 时空 K -匿名集序列数据

1.1 时空 K -匿名

时空 K -匿名方法的基本思想是:计算当前图幅网格中的用户数,如果大于等于 K ,则匿名成功,生成匿名集;否则,进一步搜寻时空临近的图幅网格。搜寻方法为:依照顺时针方向,依次搜寻空间临近的图幅网格(搜索方向为顺时针,空间最大扩展范围为 8 个邻近的网格)。累加所有图幅网格所包含的用户数,直到总的用户数大于等于 K ,则匿名成功,生成匿名集。否则,进一步进行时间邻近范围的搜寻(最大时间超前/延迟 1 个时段,每个时段的分辨率为 2 个小时),累加前后时段的当前以及空间临近的图幅网格中的用户数,如果大于等于 K ,则匿名成功,生成匿名集。否则匿名失败。

1.2 连续查询生成匿名集序列数据

连续查询^[8]是由同一用户连续两次或多次提出的查询内容相同或高度相关的位置服务查询。直接将快照查询的匿名保护方法应用于连续查询,会引起位置标识与查询标识隐私的泄露^[9-15]。文献[5,10]分别提出了利用初始匿名集作为整个连续查询匿名集的 Memorization 方法与 Plain KAA 方法。但随着匿名集中移动对象的运动,匿名集的时空区域会扩展或收缩,使得位置服务 QoS 下降与位置隐私暴露。

具体的生成序列匿名集数据的方法如下:首先,从匿名用户集 $AUS = \{u_1, u_2, \dots, u_m\}$ 中随机选择用户 u_k ,并将结果保存到数据库中的 AUS 中;然后,每个用户在每个网格均提出一次请求查询,将该网格存储到 CR 中,构成连续查询的网格序列 $CR = \{Cell_1, Cell_2, \dots, Cell_m\}$,服务器上保留该用户提出请求时延 $TD = \{T_1, T_2, \dots, T_m\}$,存储到时间延迟(TD)中;第三,对用户 u_k 参与生成的所有匿名集,按时段先后顺序进行无重复采样,生成相应的匿名集序列,并将结果保存到数据库的序列匿名集表 S 中, $S = \{S_1, S_2, \dots, S_i, \dots, S_n\}$ 。生成匿名集序列的单一序列 S_i 的数据结构如图 1 所示。

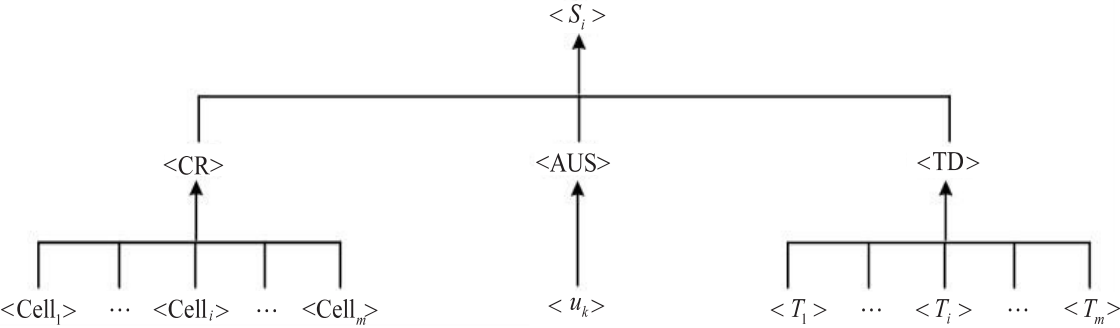


图 1 连续查询生成匿名集数据的单一数据结构

2 序列规则

序列规则挖掘任务是从给定数据库中发现一个属性的集合,在一定时间段上的一些对象都具有这些属性。例如,有一个会员制文具店的销售数据库,其中对象表示顾客,属性表示商品类别或品牌。该数据库记录了在一定时期内被每个顾客买走商品的信息。序列规则挖掘任务就是发现在一定时期内频繁被顾客所购买商品的序列。“凡是买了铅笔刀的顾客中 70% 的人在一个月之内又购买了铅笔”就是一个非常有代表性的序列规则。文具店可以利用这些模式安排促销活动、商品订货周期等。

序列模式只包括“支持度”一个度量指标^[13-15],因此,基于序列模式的事件预测并不能对预测的准确性进行充分估计。关联规则虽然有支持度和置信度两个度量指标的约束,但是不考虑时间的先后顺序。而序列规则克服了序列模式和关联规则各自的缺点,拥有支持度和置信度两个度量指标,并且考虑时间的先后顺序。

现有的序列规则算法不能反映匿名集不确定的特性,因此现有的序列规则挖掘方法不能直接应用于匿名集序列数据。由此,提出基于双向不可逆扩展的匿名集序列规则挖掘算法,解决上述两个问题,以更好地应用于匿名集序列数据,为以后的推理攻击分析打下基础,以实现用户的隐私安全保护。

3 基于双向不可逆扩展的匿名集序列规则挖掘方法

3.1 基本定义

定义 1(匿名集):AS(Anonymous Set)主要包括匿名区域 CR、匿名用户集(UIDS)、查询时间 P ,其中, $AS = \{CR, UIDS, P\}$, $CR = \{Cell_1, Cell_2, \dots, Cell_m\}$, $UIDS = \{U_1, U_2, \dots, U_k\}$ 。

定义 2(匿名集序列):SAS 是由一系列 AS 组成的序列,可以表示为: $SAS = \{AS_1, AS_2, \dots, AS_m\}$,其中 AS_1, AS_2, \dots, AS_m 按时间的先后顺序发生。

定义 3(匿名集序列规则):匿名集序列规则表示为 $A \Rightarrow B$,其中 A, B 代表两个匿名集集合,且 $A \cap B = \emptyset$, $A, B \subseteq I$,且 A 或 B 中的匿名集不分先后顺序,即同时发生。

定义 4(匿名集序列规则的支持度):匿名集序列规则的支持度是描述一个匿名集序列规则的潜在有用性的指标。序列数据库中, $\{i\}$ 和 $\{j\}$ 同时存在于 k 个序列中,对于 k 个序列中的任意序列,假设匿名集 A 中有 n 项,匿名集 B 中有 m 项,且存在 $\{i\} \subseteq A, \{j\} \subseteq B$ 。对于匿名集序列规则“ $\{i\} \Rightarrow \{j\}$ ”,根据匿名集数

据的特性,将支持度定义为:

$$\text{support}(\{i\} \Rightarrow \{j\}) = \frac{\sup(\{i\} \Rightarrow \{j\})}{|S|} = \frac{\sum_1^k (\frac{1}{n} * \frac{1}{m})}{|S|} \quad (1)$$

其中, $|S|$ 表示序列数据库中所有序列的数目; $\sup(\{i\} \Rightarrow \{j\})$ 表示匿名集序列规则的支持度计数。

定义 5(匿名集序列规则的置信度):匿名集序列规则的置信度是描述一个匿名集序列规则的有效性或“值得信赖性”的确定性度量。对于匿名集序列规则“ $\{i\} \Rightarrow \{j\}$ ”,其置信度定义为:

$$\text{confidence}(\{i\} \Rightarrow \{j\}) = \frac{\sup(\{i\} \Rightarrow \{j\})}{\sup(\{i\})} \quad (2)$$

其中, $\sup(\{i\})$ 表示匿名集序列数据库中包含项 $\{i\}$ 的数目。

3.2 算法描述

基于双向不可逆扩展方法对匿名集数据进行挖掘,其中主算法是双向不可逆扩展算法,同时,主算法中调用了两个子算法。

3.2.1 主算法

在扫描匿名集序列数据库时,将包含 c 项的序列编号(sid)记录为 sids_c , c 项第一次在匿名集序列出现的位置记录为 $\text{firstOccurrences}_c$, c 项最后一次在匿名集序列出现的位置记录为 lastOccurrences_c 。 $\text{sids}_i : j$ 和 $\text{sids}_j : i$, 分别表示匿名集序列规则 $\{i\} \Rightarrow \{j\}$ 和规则 $\{j\} \Rightarrow \{i\}$ 所在的序列编号集合。所以,不需要再次扫描数据库,就可以生成所有大小为 $1 * 1$ 的匿名集序列。

主算法:双向不可逆扩展方法(Bidirectional Irreversible Growth)

输入:序列中 item 的数据库 D 。

输出:经过双向不可逆扩展算法得到的所有规则,以及每条规则对应的支持度和置信度。

子程序:规则生长左扩展(LEFTGROWTH),规则生长右扩展(RIGHTGROWTH)。

参数:匿名集序列数据库 D , 最小支持度阈值 minsup , 最小置信度阈值 minconf 。

(1)扫描匿名集序列数据库 D 一次,计算每个项的支持度计数。生成所有满足条件大小 $1 * 1$ 、 $\text{support}(r) \geq \text{minsup}$ 的规则,并计算各规则的支持度。选择项 i 和 j , 分别记录 i 和 j 的 firstOccurrence 和 lastOccurrence 。

(2)在包含 i 和 j 的 sid 中循环,检查 i 的 firstOccurrence 是否在 j 的 lastOccurrence 之前(由于在扫描数据库时,所有 item 的 firstOccurrence 和 lastOccurrence 均被记录过,所以该步骤运行速度很快,花费时间较

少)。

(3) 如果 firstOccurence_i 在 lastOccurence_j 之前, 则当前的 sid 被添加到 sids_{i:j} 中。如果 firstOccurence_j 在 lastOccurence_i 之前, 则当前的 sid 被添加到 sids_{j:i} 中。

(4) 计算 $\sup(\{i \Rightarrow j\})/|S|$, 得到规则 $\{i \Rightarrow j\}$ 的支持度。

(5) 若 $\sup(\{i \Rightarrow j\})/|S| \geq \text{minsup}$, 则调用子程序 LEFTGROWTH 和 RIGHTGROWTH, 来扩展规则 $\{i \Rightarrow j\}$ 的左半边和右半边。

(6) 计算 $\sup(\{i \Rightarrow j\})/\sup(\{i\})$, 得到规则 $\{i \Rightarrow j\}$ 的置信度。

(7) 若 $\sup(\{i \Rightarrow j\})/\sup(\{i\}) \geq \text{minconf}$, 则输出该规则 $\{i \Rightarrow j\}$ 的支持度和置信度。

3.2.2 子算法 1

子算法 1: 左扩展 (LEFTGROWTH)。

输入: 待扩展的匿名集序列规则 $I \Rightarrow J$ (ruleIJ)。

输出: 经过左扩展后的匿名集序列规则。

参数: 待扩展的匿名集序列规则 ruleIJ, 包含项集 I 的序列列表, 包含 $I;J$ 的序列列表 (sids $I;J$), 每个序列中项集 J 最后一次出现的位置结构 (lastOccurences _{J})。

(1) 在 sids $I;J$ 中循环所有序列, 每条匿名集序列中, 从第一个项集开始扫描, 直到项集 J 最后一次出现位置之前的一个项集。找到发生时间早于或等于项集 I 的项集中的所有项, 用 c 表示。

(2) 将项 c 添加到规则左边, 构成规则 $I \cup \{c\} \Rightarrow J$ 。

(3) 计算 $\sup(I \cup \{c\} \Rightarrow J)/|S|$, 得到规则 $I \cup \{c\} \Rightarrow J$ 的支持度。若 $\sup(I \cup \{c\} \Rightarrow J)/|S| \geq \text{minsup}$, 左扩展成功; 若 $\sup(I \cup \{c\} \Rightarrow J)/\sup(I) \geq \text{minconf}$, 那么输出该规则。

(4) 对左扩展得到的规则, 检查该规则的左边项集能否再次进行左扩展。若能, 则调用 LEFTGROWTH 算法, 进行左扩展。调用 LEFTGROWTH 算法, 参数设置为规则 $I \cup \{c\} \Rightarrow J$ 、包含 $I \cup \{c\}$ 的序列列表 (sids _{Ic})、包含 $I;J$ 的序列列表 (sids $I;J$)、每个序列中 J 最后一次出现的位置 (lastOccurences _{J})。若不能, 则进行步骤 (5)。

(5) 计算 $\sup(I \cup \{c\} \Rightarrow J)/\sup(I)$, 得到规则 $I \cup \{c\} \Rightarrow J$ 的置信度。

(6) 若 $\sup(I \cup \{c\} \Rightarrow J)/\sup(I) \geq \text{minconf}$, 那么输出该规则。

3.2.3 子算法 2

RIGHTGROWTH 与 LEFTGROWTH 程序十分相似。但是, RIGHTGROWTH 程序中, 操作步骤多, 有更

多的参数, 因为同时调用了 RIGHTGROWTH 与 LEFTGROWTH。

子算法 2: 右扩展 (RIGHTGROWTH)。

输入: 待扩展的匿名集序列规则 $I \Rightarrow J$ (ruleIJ)。

输出: 经过右扩展后的匿名集序列规则。

参数: 待扩展的匿名集序列规则 ruleIJ, 包含项集 I 的匿名集序列 (sids I), 包含项集 J 的匿名集序列 (sids J), 包含 $I;J$ 的匿名集序列列表 (sids $I;J$), 每个匿名集序列中 I 第一次出现的位置 (firstOccurences _{I}), 每个匿名集序列中 J 最后一次出现的位置 (lastOccurences _{J})。

(1) 在 sids $I;J$ 中循环所有序列, 每条序列中, 从项集 I 第一次出现位置的后一个项集开始扫描, 直到序列的最后一个项集。找到发生时间晚于或等于项集 J 的项集中的所有项 c 。

(2) 将项 c 添加到规则右边, 构成规则 $I \Rightarrow J \cup \{c\}$ 。

(3) 计算 $\sup(I \Rightarrow J \cup \{c\})/|S|$, 得到规则 $I \Rightarrow J \cup \{c\}$ 的支持度。

(4) 若 $\sup(I \Rightarrow J \cup \{c\})/|S| \geq \text{minsup}$, 则调用 LEFTGROWTH 算法和 RIGHTGROWTH 算法, 分别进行左扩展和右扩展。调用 LEFTGROWTH 算法, 参数设置为匿名集序列规则 $I \Rightarrow J \cup \{c\}$ 、包含 I 的匿名集序列 (sids I)、包含 $I \Rightarrow J \cup \{c\}$ 的匿名集序列列表 (sids $I;Jc$)、包含 $I \Rightarrow J \cup \{c\}$ 的匿名集序列最后一次出现的位置结构 (lastOccurences _{Jc})。调用 RIGHTGROWTH 算法, 与 RIGHTGROWTH 主算法参数相同。

(5) 计算 $\sup(I \Rightarrow J \cup \{c\})/\sup(I)$, 得到规则 $I \Rightarrow J \cup \{c\}$ 的置信度。

(6) 若 $\sup(I \Rightarrow J \cup \{c\})/\sup(I) \geq \text{minconf}$, 那么输出该规则。

4 实验及结果分析

4.1 模拟生成实验数据

根据 2 612 辆出租车上采集的具有时空属性的 GPS 轨迹数据, 以数秒为间隔连续采样得到用户轨迹信息, 包含了每个用户在每个采样时刻的位置序列编号 (VT_ID)、经纬度坐标值 (经度: VT_LONG, 纬度: VT_LAT)、速度 (VT_SPEED)、当前时刻 (VT_DATE) 及状态 (VT_STATE)。

为了便于利用数据的时空特性进行模拟查询的匿名集数据生成, 需要对数据进行预处理, 具体步骤如下:

(1) 将 EXCEL 表格批量导入 SQL 数据库;

(2) 按时段分离整合数据, 存储在 12 个时段信息表中;

(3)空间随机,即对用户相同时段的不同轨迹点,只选取其中一个作为轨迹信息进行存储,保存在 12 个空间随机时段信息表中;

(4)划分网格,将研究区域划分为 250 * 250 个标准正方形空间网格;

(5)用户随机、时段随机,生成匿名集数据,部分匿名集数据示例: $56 * 55 \ 56 * 54 - 1 \ 56 * 50 \ 56 * 49 - 1 \ 58 * 55 \ 58 * 54 - 1 \ 55 * 49 \ 55 * 48 - 1 \ 57 * 51, 57 * 50 - 1 - 2$ 。

表 1 挖掘出的匿名集序列规则

No.	序列规则	支持度	置信度	No.	序列规则	支持度	置信度
1	$105 * 184 => 101 * 153$	112	0.875	10	$104 * 212 => 105 * 214$	104	0.764 7
2	$105 * 185 => 101 * 153$	112	0.875	11	$104 * 213 => 105 * 213$	104	0.764 7
3	$106 * 184 => 101 * 153$	112	0.875	12	$104 * 213 => 104 * 214$	104	0.764 7
4	$106 * 185 => 101 * 153$	112	0.875	13	$103 * 212 => 105 * 213$	104	0.764 7
5	$106 * 186 => 101 * 153$	128	0.888 9	14	$103 * 212 => 104 * 213$	104	0.764 7
6	$106 * 186 => 100 * 153$	112	0.777 8	15	$104 * 213 => 105 * 214$	104	0.764 7
7	$104 * 212 => 105 * 214$	104	0.764 7	16	$103 * 212 => 104 * 214$	104	0.764 7
8	$104 * 212 => 104 * 213$	104	0.764 7	17	$103 * 212 => 105 * 214$	104	0.764 7
9	$104 * 212 => 105 * 213$	104	0.764 7				

17 条匿名集序列规则实际地图表达如图 2 和图 3 所示。

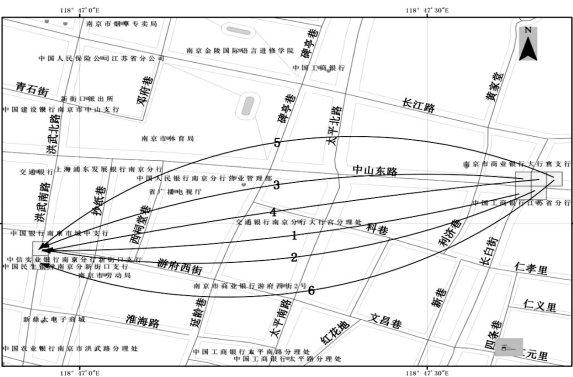


图 2 17 个匿名集序列规则与地理背景数据的叠加显示(1)

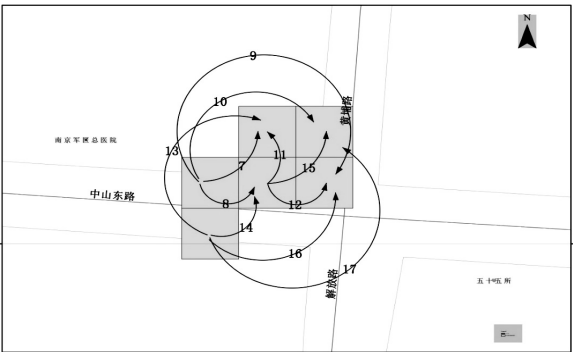


图 3 17 个匿名集序列规则与地理背景数据的叠加显示(2)

4.2 双向不可逆扩展的序列规则挖掘

4.2.1 挖掘结果

根据双向不可逆扩展的匿名集序列规则挖掘算法,对 SPMF 开源框架中的 RULEGROWTH 进行改造后,对 4.1 节中模拟生成的匿名集序列数据进行数据挖掘,并设置最小支持度阈值和最小置信度阈值分别为 100 和 0.7。经过挖掘算法得到的 17 条匿名集序列规则及其支持度、置信度,如表 1 所示。

4.2.2 隐私安全分析

根据表 1,并结合图 2 和图 3,可以发现 LBS 匿名查询的运动规律。具体分析如下:

(1)LBS 匿名查询分布在两个相互独立的区域内。一个由汉府街与长白街交接处向洪武南路运动,另一个在中山东路、解放路和黄埔路交界处运动。

(2)匿名集序列规则 1 ~ 7,由汉府街与长白街交界处 $106 * 184$ 、 $106 * 185$ 、 $106 * 186$ 、 $105 * 184$ 、 $105 * 185$ 五个网格,向洪武南路的 $101 * 153$ 和 $100 * 153$ 网格运动。运动跨度较大,规律比较明显。

(3)从图 2 中可以看出,匿名集序列规则 1 ~ 7 涉及的网格,主要分布在新街口商业繁华区(东起汉府街、长白街,西至洪武南路;北起中山东路,南至淮海路),该区域是南京市交通密集、人口密度最大的区域之一。因此上述规律可为该区域在交通高峰时的交通疏导提供一定的参考。

(4)从图 3 中可以看出,基于匿名集序列规则的预测特性,也给用户的位置隐私带来更大风险:攻击者可对进入或离开敏感时空区域的用户进行时空推理分析,以实现更具威胁性的用户隐私攻击。

(5)图 2 和图 3 涉及的新街口商业圈位于南京市的中心区,是中国著名的商业中心,拥有近百年历史,近百家世界五百强分支机构入驻。其中,涉及的隐私敏感区繁多,攻击者可对进入或离开该区域的用户进行更具威胁性的推理攻击。

5 结束语

现有的隐私保护方法,没有对大量匿名集数据进行分析,没有对匿名集数据的可用性以及隐私保护安全性进行深入研究。因此,通过对匿名数据特性以及传统序列规则挖掘方法的分析,提出了一种基于双向不可逆扩展的匿名集序列规则挖掘方法,对挖掘出的序列规则涉及位置隐私的部分结合地图进行了综合分析。验证实验结果表明,所提算法可行有效,对于隐私保护的未來方向具有重要意义。下一步将对算法进行改进以更好适应匿名集数据,并重点研究基于匿名集数据的隐私推理攻击以及相应的隐私保护算法。

参考文献:

[1] 赵文斌,张登荣. 移动计算环境中的地理信息系统[J]. 地理与地理信息科学,2003,19(2):19-23.

[2] 彭志宇,李善平. 移动环境下 LBS 位置隐私保护[J]. 电子与信息学报,2011,33(5):1211-1216.

[3] 张海涛,高莎莎,徐 亮. 空时 K-匿名数据的关联规则挖掘研究[J]. 地理与地理信息科学,2012,28(6):13-16.

[4] Gedik B,Liu L. Location privacy in mobile systems:a personalized anonymization model[C]//Proceedings of ICDCS. [s. l.]:[s. n.],2005:620-629.

[5] Xu T,Cai Y. Feeling-based location privacy protection for location-based services[C]//ACM conference on computer and communications security. [s. l.]:ACM,2009:348-357.

[6] Mokbel M F,Chow C,Aref W G. The new casper:query processing for location services without compromising privacy

[C]//Proceedings of VLDB. [s. l.]:[s. n.],2006:763-774.

[7] Ku Wei-Shinn,Zimmermann R,Peng Wen-Chih,et al. Privacy protected query processing on spatial networks[C]//Proceedings of ICDE workshops. [s. l.]:[s. n.],2007:215-220.

[8] 林 欣,李善平,杨朝晖. LBS 中连续查询攻击算法及匿名性度量[J]. 软件学报,2009,20(4):1058-1068.

[9] Xu T,Cai Y. Location anonymity in continuous location-based services[C]//Proceedings of the 15th annual ACM international symposium on advances in geographic information systems. [s. l.]:ACM,2007:39.

[10] Chow C Y,Mokbel M F. Enabling private continuous queries for revealed user locations[C]//Proceedings of SSTO. [s. l.]:[s. n.],2007:258-275.

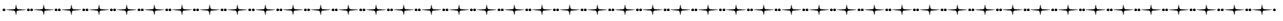
[11] Chen J,Cheng R,Mokbel M,et al. Scalable processing of snapshot and continuous nearest-neighbor queries over one-dimensional uncertain data[J]. The VLDB Journal,2009,18(5):1219-1240.

[12] Wang Yiming,Wang Lingyu,Fung B C M. Preserving privacy for location-based services with continuous queries[C]//Proceedings of ICC. [s. l.]:[s. n.],2009:1-5.

[13] 王 虎,丁世飞. 序列模式挖掘研究与发展[J]. 计算机科学,2009,36(12):14-17.

[14] 汪林林,范 军. 基于 Prefixspan 的序列模式挖掘改进算法[J]. 计算机工程,2009,35(23):56-58.

[15] 常 鹏,陈 耿,朱玉全. 一种分布式序列模式挖掘算法[J]. 计算机应用,2008,28(11):2964-2966.



(上接第 123 页)

ton,USA;ACM,2010:493-502.

[12] Blum A,Dwork C,McSherry F,et al. Practical privacy: the SuLQ framework[C]//Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems. Baltimore,USA;ACM,2005:128-138.

[13] Chaudhuri K,Monteleoni C. Privacy-preserving logistic regression[C]//Proceedings of the 22nd annual conference on neural information processing systems. Vancouver,Canada:[s. n.],2008:289-296.

[14] Shen E,Yu T. Mining frequent graph patterns with differential privacy[C]//Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. Chicago,USA;ACM,2013:545-553.

[15] McSherry F. Privacy integrated queries:an extensible platform for privacy-preserving data analysis[C]//Proceedings of the ACM SIGMOD international conference on management of data. Providence,Rhode Island,USA;ACM,2009:19-30.

[16] Mohan P,Thakurta A,Shi E,et al. GUPT:privacy preserving data analysis made easy[C]//Proceedings of the ACM SIG-

MOD international conference on management of data. Scottsdale,USA;ACM,2012:349-360.

[17] Roy I,Setty S T V,Kilzer A,et al. Airavat:security and privacy for MapReduce[C]//Proceedings of the 7th USENIX symposium on networked systems design and implementation. San Jose,USA;USENIX,2010:297-312.

[18] Peng S,Yang Y,Zhang Z,et al. Query optimization for differentially private data management systems[C]//Proceedings of the 29th IEEE international conference on data engineering. Brisbane,Australia;IEEE,2013:1093-1104.

[19] ISO/IEC JTC1/SC27 N10360,Information technology-security techniques-privacy framework[S]. Geneva:ISO/IEC,2011.

[20] ISO/IEC JTC1/SC27 N14300,Information technology-security techniques-privacy capability assessment model[S]. Geneva:ISO/IEC,2015.

[21] 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.

[22] Nisan N,Tim R,Eva T,et al. Algorithmic game theory[M]. Cambridge:Cambridge University Press,2007:216-219.