

Spatiotemporal Association Rules Mining for Key Personnel Based on FP-Growth in Spark Environment

Bo Cheng¹, Weihong Li¹, Haoxin Tong²

¹South China Normal University, Guangzhou Guangdong

²Guangdong Finest Planning Information Technology Co., Ltd., Guangzhou Guangdong

Email: 15625110283@163.com, hongweili9981@163.com

Received: Sep. 7th, 2018; accepted: Sep. 22nd, 2018; published: Sep. 29th, 2018

Abstract

Taking the improvement of public security enforcement efficiency and the urgent need for crime prevention and improvement, focusing on key personnel, car ticketing and hotel accommodation data, we analyzed the key personnel data and other data to get association rules in the framework of Spark distributed computing using the method of spatio-temporal association rules based on FP-growth, and find potential key personnel. This method has been successfully applied to the policing system. It has been proved through practical tests that the method is effective in detecting potential offenders.

Keywords

Spark, FP-Growth, Spatio-Temporal Association Rules, Potential Key Personnel

Spark环境下犯罪人员时空关联规则挖掘

程 博¹, 李卫红¹, 童昊昕²

¹华南师范大学, 广东 广州

²航天精一(广东)信息科技有限公司, 广东 广州

Email: 15625110283@163.com, hongweili9981@163.com

收稿日期: 2018年9月7日; 录用日期: 2018年9月22日; 发布日期: 2018年9月29日

摘 要

针对关于潜在犯罪人员的预测、挖掘效果不佳, 利用犯罪人员、交通出行和住宿消费等数据, 在Spark

文章引用: 程博, 李卫红, 童昊昕. Spark 环境下犯罪人员时空关联规则挖掘[J]. 数据挖掘, 2018, 8(4): 210-219.

DOI: 10.12677/hjdm.2018.84022

分布式运算框架下, 基于FP-growth的时空关联规则方法, 分析犯罪人员数据和其他数据之间的关联规则, 挖掘出潜在的犯罪人员。首次提出将关联规则算法用于普通出行消费数据实现潜在犯罪人员的预测。该方法已成功应用于X市警务系统, 通过实践检验证明该方法在发现潜在犯罪人员方面的有效性。

关键词

Spark, FP-growth, 时空关联规则, 潜在犯罪人员

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目前, 国内外犯罪形势日益严峻, 犯罪类型多样, 犯罪人员数量增加, 危害着人民生命财产安全和社会稳定。公安机关维护公共安全、打击犯罪的任务越来越重, 执法要求越来越高。不断产生的犯罪数据要求分析人员挖掘犯罪数据潜藏的规律、分析犯罪数据之间的隐含关系, 并且预测犯罪的发生、挖掘潜在的犯罪人员, 从而提高公安执法的效率、预防犯罪的发生。

关联规则挖掘(ARM)是数据挖掘领域的一个研究热点, 作为数据挖掘中的经典算法之一, 被广泛运用到犯罪学研究中。关联规则可以从大量数据和数据项之间的关联规则中挖掘犯罪的相关证据, 进一步挖掘不同犯罪之间的犯罪规律、趋势和联系, 向警方提供案件侦查和犯罪预防。在探究过去犯罪成因、确定主要嫌疑人、更好地理解系列犯罪方面, 关联规则挖掘都能起到很好的效果。不同的关联规则挖掘算法在犯罪分析和犯罪预测领域起着重要作用。在国外的关联规则犯罪挖掘研究中, Ng V 等通过引入时间关联规则, 提出一种增量算法, 以解决关联规则中包含时间表达式的时间序列处理问题, 用于发现香港地区犯罪模式[1]。Buczak 研究了模糊关联规则挖掘在社区犯罪模式中的应用, 加速了当地执法工作[2]。Tan 首先分析了 FP-Growth 算法在计算机犯罪取证中的作用, 指出了 FP-Growth 算法能够及时发现最新犯罪和严重犯罪的缺陷, 并对 FP-Growth 算法和测试进行了一些改进。最后指出了未来的研究方向: 多维证据的提取与关联[3]。Joshi 提出 FP-Tree 相似算法, 用于挖掘经常性犯罪集合, 并与 Apriori 算法相比较, 发现该算法更有效[4]。D. Usha 将 Apriori 算法与其他 Apriori、Fp-Growth 算法在真实和合成的犯罪数据集中进行测试, 发现每种算法都有自身的优势, 以便研究人员了解频繁模式挖掘算法在各个领域的应用[5] [6]。Shekhar 在犯罪模式分析(PCA)的基础上探讨了空间频繁模式挖掘(SFPM), 并在空间犯罪数据集上验证了该挖掘方法[7]。Isafiade 重新探讨了犯罪模式挖掘的频繁模式增长模型, 提出了一种基于四分位数(floor-ceil)函数的描述性统计方法, 用于最小支持阈值(MST)选择。修改后的频繁模式增长(RFPG)模型进一步提出了一种模式, 用于识别微妙犯罪模式序列的元组或犯罪活动中反复出现的趋势[8]。Asmai 基于地理和人口因素, 利用关联规则挖掘为犯罪人员制作犯罪映射模型。它检查了特定地点的犯罪发生情况, 可以用来分析相对较高的未来犯罪地点, 可以改进犯罪预防实施[9]。国内利用关联规则算法进行犯罪挖掘的研究也做了大量工作。以模糊集、Rough 集理论为基础, 林和、虞龙江等利用关联规则挖掘对犯罪人员数据库进行定量分析、推断并提取规则, 为犯罪预防提供理论指导[10] [11]。在犯罪画像、犯罪取证分析方面, 关联规则挖掘引起广泛关注, 得到广泛应用[12] [13] [14]。此外, 关联规则挖掘被大量运用在犯罪侦查[15]、犯罪嫌疑人分析[16]、犯罪行为分析[17]、重新犯罪[18] [19]等犯罪研究领域。基于犯罪数据的时间和空间属性, 许多研究提出改进的关联规则挖掘算法, 如时空关联规则[20] [21]、聚类关

关联规则[22]、基于数据立方体的关联规则[23]，其他改进的算法如增量关联规则[24]、改进的 Apriori、FP-growth 算法也在财产犯罪分析等领域取得不错效果[25]。

综上所述，国内外学者都运用关联规则来进行犯罪挖掘的研究，深入探讨了关联规则挖掘在不同犯罪研究领域的效果，并提出增量关联规则、时空关联规则、聚类关联规则等改进的关联规则挖掘算法。上述研究主要着眼于利用犯罪数据进行关联规则挖掘，但是在实际业务中，更多的是与犯罪没有必然联系的业务数据，这些数据也更加容易获取，如何从大量的普通业务数据中挖掘出犯罪信息至关重要。此外，先前研究主要关注犯罪案件分析、犯罪模式挖掘，没有针对潜在犯罪人员的预测。本研究通过对 X 市犯罪人员信息数据、交通出行数据及住宿消费数据等进行综合的关联规则挖掘，找出在不同时间不同区域频繁与犯罪人员一同出行、住宿的普通人员。由于犯罪行为时常是以团伙形式进行，挖掘出的同行人员可以成为潜在的重点关注人员。对潜在犯罪人员进行监控，能够极大减少犯罪案件的发生，提升公共安全。

2. 研究方法

2.1. 关联规则基本概念

关联分析用于挖掘数据间的关联关系，通常使用频繁项集和关联规则来表征此种关联关系。频繁项集表示数据项集合中一起频繁出现的项目子集；关联规则表示 2 个项集间关联度的规则，常以 $X \geq Y$ 表示，其中 X、Y 为 2 个互斥的项集。通常使用支持度(support)和置信度(confidence)来度量关联规则的相关程度。

2.1.1. 支持度和置信度

支持度表示项集 {X, Y} 在总项集里出现的概率。表示为：

$$\text{Support}(X \geq Y) = P(X, Y) / P(I) = P(X \cup Y) / P(I) = \text{num}(XUY) / \text{num}(I) \quad (1)$$

其中，I 表示总事务集。num()表示求事务集里特定项集出现的次数。比如，num(I)表示总事务集的个数，num(XUY)表示含有 {X, Y} 的事务集的个数。

置信度表示在先决条件 X 发生的情况下，由关联规则 “ $X \geq Y$ ” 推出 Y 的概率。即在含有 X 的项集中，含有 Y 的可能性，公式为：

$$\text{Confidence}(X \geq Y) = P(Y | X) = P(X, Y) / P(X) = P(XUY) / P(X) \quad (2)$$

2.1.2. 频繁项集和关联规则

给定最小支持度 min_sup 和最小置信度 min_conf 其范围为 [0, 100%]，则当且仅当满足 $\text{support}(X) \geq \text{min_sup}$ 时的项集 X 为频繁项集，若 $\text{support}(X \geq Y) \geq \text{min_sup}$ ，则称 X 与 Y 是关联规则。若 X 与 Y 同时满足 $\text{confidence}(X \geq Y) \geq \text{min_conf}$ ，则称 X 与 Y 为强关联规则。

2.1.3. 时空关联规则

时空关联规则是在关联规则和空间关联规则的基础上提出的，其理论方法主要是针对于含有时间和空间约束的数据进行分析挖掘，在空间关联规则的基础上添加了时间上的约束。时空关联规则挖掘是要提取出所有符合一定时间约束条件下的强空间关联规则，时空关联规则不仅是时间约束下的频繁项集也是空间约束下的频繁项集，是时间频繁项集与空间频繁项集之间的交集。在时空数据挖掘理论基础之上，时空关联规则挖掘利用常用关联规则算法对时空数据进行关联规则挖掘，寻找数据的时空关联性。近年来，时空关联分析在物流交通、灾害预警、规划建设、土地资源、环境等领域应用广泛[26]-[31]。

2.2. FP-Growth 算法

常见的关联规则算法主要有 Apriori 算法、FP-growth 算法，FP-growth 是一种不产生候选模式而采用

频繁模式增长的方法挖掘频繁模式的算法；此算法只扫描两次，第一次扫描数据是得到频繁项集，第二次扫描是利用支持度过滤掉非频繁项，同时生成 FP-Tree。然后后面的模式挖掘就是在这棵树上进行。此算法与 Apriori 算法最大不同的有两点：不产生候选集，只遍历两次数据，大大提升了效率。

FP-growth 算法为了减少 I/O 操作，提高效率，引入了一些数据结构存储数据，主要包括项头表、FP-Tree 和节点链表。项头表(Header Table)即找出频繁 1 项集，删除低于支持度的项集，并按照出现的次数降序排序，这是第一次扫描数据。然后从数据中删除非频繁 1 项集，并按照项头表的顺序排序，这是第二次也是最后一次扫描数据。FP-Tree (Frequent Pattern Tree)初始时只有一个根节点 Null，将每一条数据里的每一项，按照排序后的顺序插入 FP-Tree，节点的计数为 1，如果有共用的祖先，则共用祖先的节点计数 + 1。FP-Tree 的挖掘过程如下，从长度为 1 的频繁模式开始挖掘。可以分为 3 个步骤：

1) 构造它的条件模式基(CPB, Conditional Pattern Base)，条件模式基(CPB)就是我们要挖掘的 Item 的前缀路径；

2) 然后构造它的条件 FP-Tree (Conditional FP-tree)；

3) 递归的在条件 FP-Tree 上进行挖掘。

伪代码如下：

```

procedure FP-growth(Tree, a)
if Tree 含单个路径 P then{
    for 路径 P 中结点的每个组合 (记作 b)
        产生模式  $b \cup a$ ，其支持度  $support = b$  中结点的最小支持度；
    } else {
        for each  $a_i$  在 Tree 的头部(按照支持度由低到高顺序进行扫描){
            产生一个模式  $b = a_i \cup a$ ，其支持度  $support = a_i.support$ ；
            构造 b 的条件模式基，然后构造 b 的条件 FP-树 Treeb；
            if Treeb 不为空 then
                调用 FP-growth (Treeb, b)；
        }
    }
}

```

FP-growth 函数的输入：*tree* 是指原始的 FP-Tree 或者是某个模式的条件 FP-Tree，*a* 是指模式的后缀 (在第一次调用时 *a* = NULL，在之后的递归调用中 *a* 是模式后缀)；

FP-growth 函数的输出：在递归调用过程中输出所有的模式及其支持度。每一次调用 FP-growth 输出结果的模式中一定包含 FP-growth 函数输入的模式后缀。

2.3. 研究方法流程

如图 1 所示，本研究大致分为以下几个步骤(见图 1)：

1) 将犯罪人员数据划分为训练数据和测试数据，两者的比例为 4:1；对所有数据进行预处理。将犯罪人员训练数据和交通出行、住宿消费数据统一作为算法输入数据上传至 HDFS；

2) 设置最小支持度(min_sup)和最小置信度(min_conf)，使用 Spark Mlib 对输入数据进行计算；

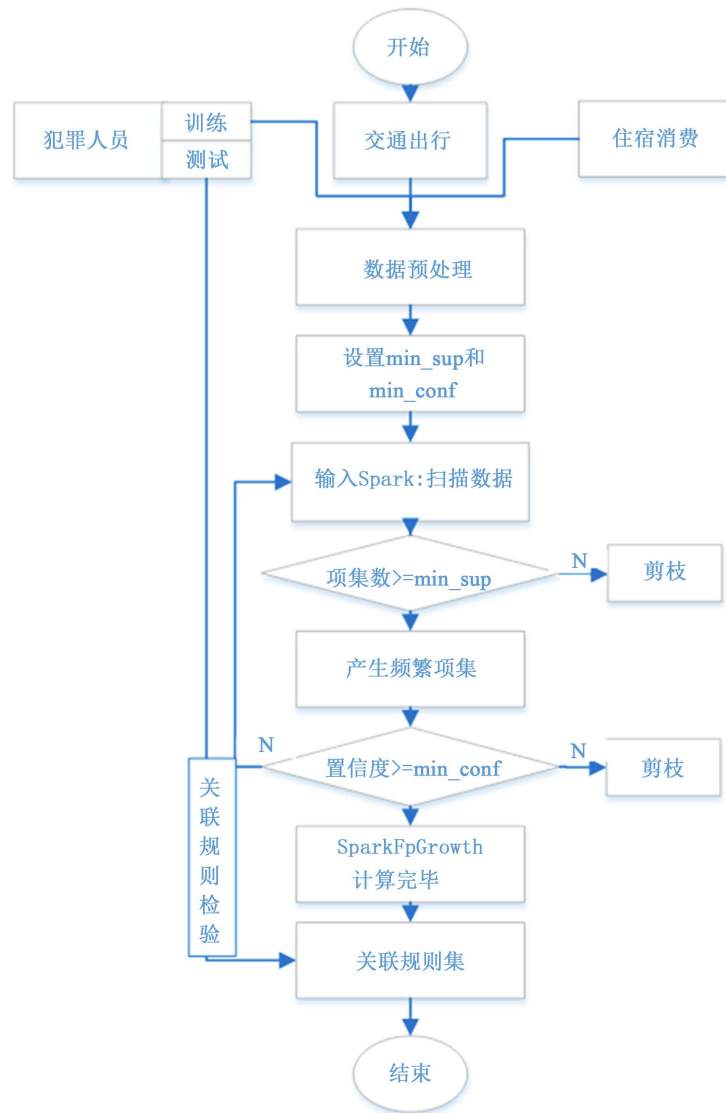


Figure 1. Flow chart of research methods

图 1. 研究方法流程图

- 3) 根据设置的 \min_sup 和 \min_conf 条件对频繁项集进行剪枝，得到关联规则结果集；
- 4) 利用犯罪人员测试数据对关联规则结果集进行检验，检查关联规则的正确性和算法有效性。

3. 实验流程

3.1. 数据预处理

1) 数据提取：从 X 市警务系统数据库中抽取用于关联规则分析的犯罪数据及相关数据，包括犯罪人员数据、交通出行数据、住宿消费数据，并从各数据中提取出分析所需的相应属性字段(见表 1)；

2) 数据清洗：数据表中字段存在空值、错误值等问题，如身份证号码为空或长度不够等，需要对存在问题的字段进行一定处理。本文直接删除存在错误的的数据，因为其所占比例较小，不影响分析；

3) 数据转换、加载：本研究数据中涉及到身份证号码等隐私信息，通过加密算法进行加密处理。由于本研究使用 SparkMllib 中的 FP-growth 算法进行计算，因此将转换后的数据文件上传至 Hdfs 中。

Table 1. Crime and related data information**表 1.** 犯罪及相关数据信息

数据表	属性字段
犯罪人员	身份证号码
交通出行	车次 ID、发车时间、起始站、到达站、乘车人员身份证号码
住宿消费	入住人员身份证号码、旅馆编号、旅馆坐标、入住时间、旅馆名称

3.2. 挖掘过程

本实验在 Spark 分布式运算框架下进行，总共部署五个物理节点，以实现快速、高效的计算。Spark 部署简单、运算高效，是专为大规模数据处理而设计的快速、易用、通用、随处运行的分布式计算引擎，其中间计算结果保存在内存当中，可以极大提高运算效率。本实验中，相对于单机运算，Spark 分布式计算时间效率提高 300%；相对于 Hadoop MapReduce 分布式集群，时间效率提高 150%。本研究利用分布式运算框架 Spark 中的机器学习算法库 SparkMllib 中的 FP-growth 接口来计算各数据的关联规则。

关联规则的计算包括三部分，第一部分为利用交通出行数据进行普通的关联规则分析，目的是找出同犯罪人员一起频繁出行的潜在人员；第二部分利用住宿消费数据进行时空关联规则挖掘，目的是找出经常同犯罪人员一起住宿消费的潜在犯罪人员。由于犯罪人员和潜在犯罪人员可能住宿消费的时间地点不同，因此时空关联规则分析考虑在某一时间和空间邻域内的关联规则；第三部分综合利用交通出行数据和住宿消费数据考虑时间和空间属性的关联规则分析，目的是找出经常同犯罪人员一起出行和住宿消费的潜在犯罪人员，由于犯罪人员和潜在犯罪人员存在分开出行、住宿消费的情况，因此关联规则分析考虑某一时间、空间范围内的出行和住宿消费情况。

1) 交通出行数据关联规则挖掘

交通出行数据关联规则挖掘以汽车或者火车为例。将每个车次(包括车次 ID、发车时间、起始站、到达站)当作一个交易(Transaction) ID，该车次上的所有乘客为交易项目。给定支持度和置信度，使用 Spark 对预处理后的汽车数据进行运算，运算的结果规则形式为“车次乘客 A \geq 乘客 B”，表示在给定时间内、符合最小支持度和置信度情况下，乘客 A 和乘客 B 共同乘坐了该车次。对于以上处理结果，并不知道该条规则中是否包含犯罪人员，因此，针对所有关联规则，过滤找出包含犯罪人员的关联规则。对每条包含犯罪人员的关联规则，如“A \geq B”，假设 A 为重点人员，此时可以统计出 A 和 B 共同乘坐的所有车次，那么，经常同犯罪人员 A 一起乘坐该车次的 B 可以作为潜在的重点关注人员。

2) 住宿消费数据时空关联规则挖掘

住宿消费数据时空关联规则挖掘以旅馆住宿或者网吧上网为例。对于每个旅馆或者网吧，以该旅馆或者网吧空间坐标为中心、以当前登记时间为准，给点空间半径 $r = 3 \text{ km}$ ，时间半径 $t = 2 \text{ d}$ ，将该时间和空间范围内所有登记消费的全部人员当作一个交易。给定支持度和置信度，使用 Spark 对预处理后的人员数据进行运算，运算的结果规则形式为“消费场所 ID 人员 A \geq 人员 B”，表示在给定时空范围内、符合最小支持度和置信度情况下，旅客 A 和 B 在该旅馆或者网吧消费。同理，可查找出经常同犯罪人员 A 一起住宿消费的潜在犯罪人员 B。

3) 综合交通出行、住宿消费数据的时空关联规则挖掘

在分别得出汽车和旅馆数据的关联规则后，计算获得汽车和旅馆数据关联规则中相同的关联规则，即得同时乘坐某趟车次并且一同住宿消费的犯罪人员和潜在犯罪人员。

4. 关联规则结果分析

4.1. 交通出行关联规则

计算结果如表 2 所示, 总共发现 433 个潜在犯罪人员。表 2 中 5 个字段分别表示犯罪人员身份证号码(用 A 表示)、潜在犯罪人员身份证号码(用 B 表示)、置信度、犯罪人员和潜在犯罪人员同乘的车次信息以及两者同行的次数。以第二条规则为例, 该规则表示犯罪人员 A 和犯罪人员 B 在研究时间内一起乘坐了 78 次汽车, 其中置信度表示在犯罪人员乘坐的所有车次中, 有 90%是同潜在犯罪人员一起乘车; 此外, 乘坐车次信息表示该车次编号为 653124 xxxxxxxx, 发车时间为 2016-12-16 16:50:00, 从 aks 站开往 awt 站(见表 2)。

4.2. 住宿消费时空关联规则

表 3 为不考虑时间和空间的关联规则分析结果(见表 3), 表 4 为考虑时间和空间的时空关联规则分析结果(见表 4)。从表 3 中可以看出犯罪人员和潜在犯罪人员同住旅馆的信息, 其中时间表示入住的日期; 另外, 同住次数表示两者在研究时间范围内一起外出住宿的次数。总共发现 123 个潜在重点人员。时空关联规则分析中, 设置时间半径 $t = 2 d$, 空间半径 $r = 3 km$ 。通过对比发现, 增加时空条件相当于扩大关联规则搜索范围, 此时, 发现的关联规则数量急剧增加至 207 条, 犯罪人员和潜在犯罪人员同住旅馆的次数显著上升, 说明犯罪人员与潜在犯罪人员不在同一天入住或在不同的旅馆入住的情况比较明显。

Table 2. Traffic Association Rules

表 2. 交通出行关联规则

重点人员	潜在重点人员	置信度	车次信息	同行次数
6529281973xxxxxxx	6529281968 xxxxxxxx	1	{652901xxxxxxx, 2016-12-16 16:50:00, aks 站, awt 站}……{……}	45
6529271992 xxxxxxxx	4128281987 xxxxxxxx	0.9	{653124xxxxxxx, 2016-12-26 16:50:00, aks 站, awt 站}……{……}	78
……	……	……	……	……

注: 共计 433 条关联规则。

Table 3. Accommodation consumption association rules

表 3. 住宿消费关联规则

重点人员	潜在重点人员	置信度	旅馆信息	同住次数
6541261976xxxxxxx	4111231980 xxxxxxxx	0.6	{2016-05-01 jy 宾馆}、{2016-05-03 xh 酒店}……{……}	18
5134291992xxxxxxx	1101111985 xxxxxxxx	0.8	{2016-05-15 fk 酒店}、{2016-05-25 sd 酒店}……{……}	23
……	……	……	……	……

注: 共计 123 条关联规则。

Table 4. Rules for spatial and temporal association of accommodation consumption

表 4. 住宿消费时空关联规则

重点人员	潜在重点人员	旅馆信息
6541261976xxxxxxx	4111231980xxxxxxx	{2016-05-08 sy 宾馆}、{2016-05-09 rj 宾馆}、{2016-05-03 xh 酒店}……{……}
5134291992xxxxxxx	1101111985xxxxxxx	{2016-05-15 fk 酒店}、{2016-05-03 tm 大酒店}、{2016-05-25 sd 酒店}……{……}
……	……	……

注: 共计 207 条关联规则。

4.3. 综合时空关联规则

从表 5 中可知, 综合考虑交通出行和住宿消费数据的时空关联规则数量为 157 条, 相对仅考虑住宿消费数据的时空关联规则数量有所减少, 因为增加了同时乘车这一限制条件。表中同行信息字段表示犯罪人员和潜在犯罪人员同时乘坐该车次或相邻车次并且入住该宾馆或附近旅馆(见表 5)。

4.4. 潜在犯罪人员分析及检验

根据图 1 中各数据关联规则结果可知总共有 920 条关联规则, 即研究发现了 920 个潜在犯罪人员。但是, 通过检查各关联规则发现各数据计算得出的关联规则之间存在大量的重复的潜在犯罪人员, 例如可能在交通出行数据和住宿消费数据或者综合数据中同时发现潜在犯罪人员 A, 此时需要消除重复检测出的潜在犯罪人员。经过上述步骤后, 总计有 648 个潜在犯罪人员被发现(见图 2)。

此外, 针对发现的 648 个潜在犯罪人员, 利用犯罪人员测试数据对其进行检验, 检测出近 20% 的潜在犯罪人员存在于犯罪人员测试数据库中。该检验表明研究方法能够有效发现潜在的犯罪分子在发现上述潜在犯罪人员后, X 市公安部门对其进行了日常出行活动的监控, 经过两个月的监控发现, 部分潜在犯罪人员存在扰乱公共秩序、违反治安条例的行为。由此可见, 本研究发现的潜在犯罪人员在实际生活中可以成为公安机关的重点关注对象, 对其行为的监控能够起到维护社会治安、减少违法犯罪行为的效果。

5. 结束语

本研究从公共安全的实际需求出发, 在 Spark 分布式运算环境下, 实现了犯罪人员快速、高效的挖掘。利用基于 FP-growth 的时空关联规则算法, 对犯罪人员数据、交通出行数据和住宿消费数据进行

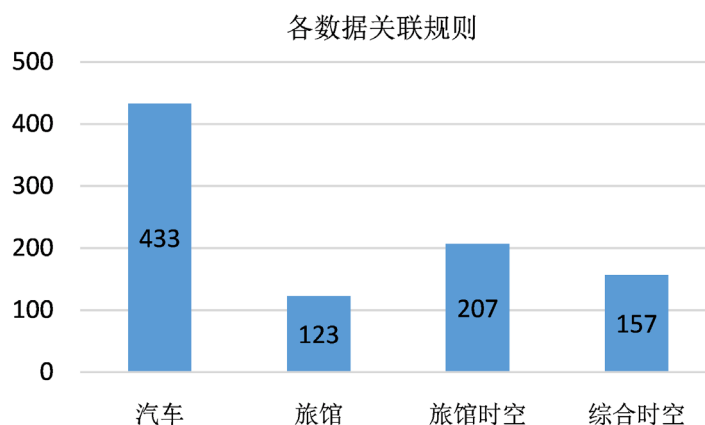


Figure 2. Statistics of the results of the data association rules

图 2. 各数据关联规则结果统计

Table 5. Comprehensive data association rules

表 5. 综合数据关联规则

重点人员	潜在重点人员	同行信息
6523241974xxxxxxxx	1528271980xxxxxxxx	{650202xxxxxxxx, 2017-6-14 10:20:00, ps 站, 叶城站}{2016-05-26 ql 宾馆}……{}
5111291987 xxxxxxxx	2308271969xxxxxxxx	{650202xxxxxxxx, 2017-5-12 10:20:00, ds 站, kt 站}{2016-05-14 yc 宾馆}……{}
……	……	……

注: 共计 157 条关联规则。

时空关联规则挖掘, 研究发现了 648 个潜在犯罪人员。在挖掘效率上, 相对于单机环境和 HadoopMapReduce 分布式环境, Spark 分布式运算环境都极大提高; 此外, 与其他针对犯罪案件的研究不同, 本研究挖掘预测潜在犯罪人员, 更加具有针对性和指向性。本研究方法也可以应用于火车、航班票务数据、网吧上网数据等人员出行数据或其他相关数据, 具有极大的可扩展性。研究已成功应用于 X 市警务系统, 并对潜在犯罪人员进行特殊关注, 发现部分人员存在扰乱社会秩序、违反治安条例的行为, 实践检验了本研究方法的有效性, 对治安防控具有现实意义。

需要注意的是, 研究发现的潜在重点人员中包含大量的儿童, 在实际治安防控中可以减少对这些对象的关注, 因为违法犯罪的可能性较少。此外, 研究中关于犯罪人员和潜在犯罪人员同行次数(即支持度)的设置也是一个难点, 需要根据实际情况不断调整。后续研究将继续利用关联规则算法对相关数据进行挖掘, 满足公共安全的实战需要。

参考文献

- [1] Ng, V., Chan, S., Lau, D., *et al.* (2007) Incremental Mining for Temporal Association Rules for Crime Pattern Discoveries. *Proceedings of the 18th Australasian Database Conference (ADC 2007)*, Ballarat, Victoria, 29 January-2 February 2007, 123-132.
- [2] Buczak, A.L. and Gifford, C.M. (2010) Fuzzy Association Rule Mining for Community Crime Pattern Discovery. *ACM SIGKDD Workshop on Intelligence and Security Informatics*, Washington, DC, 25-28 July 2010, 1-10. <https://doi.org/10.1145/1938606.1938608>
- [3] Tan, Y., Qi, Z. and Wang, J. (2012) Applications of Association Rules in Computer Crime Forensics. *Applied Mechanics & Materials*, **157-158**, 1281-1286. <https://doi.org/10.4028/www.scientific.net/AMM.157-158.1281>
- [4] Joshi, A. and Suresh, M.B. (2014) Compact Structure of Felonious Crime Sets Using FP-Tree Comparable Algorithms Analysis. *International Journal of Computer Science and Information Technologies*, **5**, 2694-2699.
- [5] Usha, D. and Rameshkumar, K. (2013) Frequent Pattern Mining Algorithm for Crime Dataset: An Analysis. *International Journal of Engineering Sciences & Research Technology*, **2**, 3379-3384.
- [6] Usha, D. and Rameshkumar, K. (2014) A Complete Survey on Application of Frequent Pattern Mining and Association Rule Mining on Crime Pattern Mining. *International Journal of Advances in Computer Science and Technology*, **3**, 264-275.
- [7] Shekhar, S., Mohan, P., Oliver, D., *et al.* (2014) Crime Pattern Analysis: A Spatial Frequent Pattern Mining Approach Crime Pattern Analysis: A Spatial Frequent Pattern Mining Approach. *Crime Pattern Analysis a Spatial Frequent Pattern Mining Approach*.
- [8] Isafiade, O., Bagula, A. and Berman, S. (2015) A Revised Frequent Pattern Model for Crime Situation Recognition Based on Floor-Ceil Quartile Function. *Procedia Computer Science*, **55**, 251-260. <https://doi.org/10.1016/j.procs.2015.07.042>
- [9] Asmai, S.A., Roslin, N.I.A., Abdullah, R.W., *et al.* (2014) Predictive Crime Mapping Model Using Association Rule Mining for Crime Analysis. *Science International*, **26**, 1703-1706.
- [10] 林和, 莫照, 虞龙江, 等. 基于 Rough 集数据挖掘在犯罪人口数据库中的应用[J]. 计算机科学技术, 2007.
- [11] 虞龙江. 模糊与粗集理论在犯罪人口数据库中的应用[D]: [硕士学位论文]. 兰州: 兰州大学, 2003.
- [12] 魏士靖, 张基温, 耿汝年. 基于犯罪画像的计算机取证分析方法研究[J]. 微计算机信息, 2006, 22(4): 237-239.
- [13] 魏士靖. 关联规则在画像分析取证中的应用[J]. 中国科技论文在线, 2005.
- [14] 宿娜. 基于关联规则的实时网络取证模型研究[D]: [硕士学位论文]. 青岛: 山东科技大学, 2013.
- [15] 盛巍. 侦查信息关联规则原理[J]. 科技信息, 2012(36): 300-300.
- [16] 于娜. 基于关联规则算法的嫌疑程度关系发现方法研究[D]: [硕士学位论文]. 大连: 大连工业大学, 2015.
- [17] 徐伟, 张军. 关联规则在犯罪行为分析中的应用研究[C]//中国科协年会. 第十届中国科协年会论文集: 2008 年卷.
- [18] 汤毅平. 基于 Apriori 算法的重新犯罪关联规则挖掘[J]. 指挥信息系统与技术, 2016(3): 91-95.
- [19] 冯卓慧, 冯前进. 基于关联规则的再犯罪特征分析[J]. 浙江理工大学学报(社会科学版), 2017, 38(1): 57-60.
- [20] 闫密巧, 过仲阳, 任浙豪. 基于聚类关联规则的公交扒窃犯罪时空分析[J]. 华东师范大学学报: 自然科学版,

2017(3): 145-152.

- [21] 孙越恒, 王文俊, 迟晓彤, 等. 基于多维时间序列模型的社会安全事件关联关系挖掘与预测[J]. 天津大学学报(社会科学版), 2016, 18(2): 97-102.
- [22] 王慧, 郑涛, 张建岭. 基于聚类的关联规则算法在刑事犯罪行为分析中的应用[J]. 中国人民公安大学学报(自然科学版), 2010, 16(3): 64-67.
- [23] 王海波, 张永田, 吴升. 基于数据立方体的多最小支持度关联规则在犯罪分析中的应用[J]. 测绘科学技术学报, 2016, 33(4): 405-409.
- [24] 杜威, 邹先霞. 增量关联规则挖掘算法在犯罪行为中的应用研究[J]. 中国人民公安大学学报(自然科学版), 2011, 17(2): 56-58.
- [25] 许阳泉. 关联规则在财产犯罪分析中的应用[D]: [硕士学位论文]. 福州: 福州大学, 2014.
- [26] 张予. 数据挖掘技术在高危人员犯罪信息挖掘的应用研究[D]: [硕士学位论文]. 南昌: 南昌大学, 2009.
- [27] 夏英, 张俊, 王国胤. 时空关联规则挖掘算法及其在 ITS 中的应用[J]. 计算机科学, 2011, 38(9): 173-176.
- [28] 李晶晶. 时空数据挖掘在环境保护中的应用研究[D]: [硕士学位论文]. 长沙市: 中南大学, 2008.
- [29] Xue, C.J., Dong, Q. and Ma, W.X. (2014) Object-Oriented Spatial-Temporal Association Rules Mining on Ocean Remote Sensing Imagery. *35th International Symposium on Remote Sensing of Environment*, Beijing, 22-26.
- [30] Mennis, J. and Liu, J.W. (2005) Mining Association Rules in Spatio-Temporal Data: An Analysis of Urban Socioeconomic and Land Cover Change. *Transactions in GIS*, 9, 5-17. <https://doi.org/10.1111/j.1467-9671.2005.00202.x>
- [31] 叶文菁, 吴升. 基于加权时空关联规则的公交扒窃犯罪模式识别[J]. 地球信息科学学报, 2014, 16(4): 537-544.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2163-145X, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: hjdm@hanspub.org