

doi: 10.7690/bgzd.2024.09.006

# 一种 Apriori 算法在监狱警察心理健康调查中的应用

吴 萌

(陕西警官职业学院, 西安 710021)

**摘要:** 为从复杂数据中挖掘监狱警察心理健康状况, 提出一种改进的 Apriori 算法进行监狱警察心理健康监测和评估, 并根据狱警心理表现症状进行疏导。针对传统 Apriori 算法计算量大、遍历数据长的问题, 对数据库集心理健康调查结果进行矩阵化处理和和运算, 采用自连接操作方式删除非频繁项集, 降低数据计算量。将改进算法应用于监狱警察心理健康调查, 应用结果表明: 监狱警察普遍存在“内向、抑郁、焦虑、坚强”的心理特征, 后续心理辅导工作中, 需加强户外锻炼, 多接触外界事物, 培养活泼开朗、积极乐观的情绪。

**关键词:** 监狱警察; 心理健康; 改进 Apriori 算法

**中图分类号:** TP311.13 **文献标志码:** A

## Application of Apriori Algorithm in Mental Health Investigation of Prison Police

Wu Meng

(Shaanxi Police Vocational College, Xi'an 710021, China)

**Abstract:** In order to mine the mental health status of prison police from complex data, an improved Apriori algorithm is proposed for monitoring and evaluating the mental health of prison police, and guidance is provided based on the psychological symptoms of prison police. In response to the problems of high computational complexity and long traversal data in the traditional Apriori algorithm, the mental health survey results of the database set are matrix processed and summed, and non-frequent item sets are removed through self connection operation to reduce data computation. The improved algorithm was applied to the mental health survey of prison police, and the application results showed that prison police generally have psychological characteristics of “introversion, depression, anxiety, and resilience”. In subsequent psychological counseling work, it is necessary to strengthen outdoor exercise, have more contact with external things, and cultivate lively, cheerful, positive and optimistic emotions.

**Keywords:** prison police; mental health; improved Apriori algorithm

## 0 引言

狱警作为一类社会特殊警察群体, 长期与罪犯相处, 要时刻保持警惕, 观察罪犯的言行举止, 防止发生罪犯突发事故, 在繁忙的工作中, 业余时间较少, 无法正常与家人团聚。由于工作环境和工作内容特殊性, 面临着来自各方的压力, 经常出现恐慌、焦虑的心理状态, 使得心理健康问题突出<sup>[1-3]</sup>。根据调查发现, 狱警中, 心理不健康比例者达到 38%, 其中男性狱警心理不健康比例达到了近 50%。狱警在心理健康方面已成为一个高危群体<sup>[4]</sup>。目前, 对狱警心理的状态跟踪以及心理干预方法成为研究热点。运用心理健康测试方式对狱警心理健康测试, 如何从庞大测试数据中找到有价值信息, 成为心理健康测试面临的难题<sup>[5-7]</sup>。Apriori 作为一种经典的关联规则挖掘算法, 能从大量数据各数据集的相关关联关系为管理决策提供支持, 但算法的计算较大、数据复杂<sup>[8]</sup>。笔者在基于 Apriori 算法的关联

规则挖掘研究的基础上, 对传统 Apriori 算法进行改进, 降低计算量和遍历数, 通过寻找监狱警察主要因素和心理健康症状间的关系, 为监狱警察心理健康发展提供支持。

## 1 改进的 Apriori 算法

### 1.1 Apriori 算法

Apriori 算法基于逐层搜索的迭代思想<sup>[9]</sup>, 首先寻找到频繁的 1-项集, 根据给定的阈值  $\text{minsup}$  对频繁 1-项集剪枝, 获得初始集表  $L_1$ , 根据先验原理, 若某一项集为频繁集, 则所有子集也是频繁集, 在形成候选的 2-项集时, 使用频繁 1-项集来产生生成候选项  $L_2$ 。产生候选项集 2-后, 根据给定阈值  $\text{minsup}$  对 2-项集剪枝处理, 形成频繁项集 2-, 依次类推, 直到无法寻找到更多频繁项目为止, 每搜索一次, 对数据库进行一次整体扫描, 每一次扫描仅考虑同一宽度  $K$  的所有  $K$ -项集。图 1 为 Apriori

收稿日期: 2024-05-17; 修回日期: 2024-06-20

基金项目: 陕西警官职业学院重点课题(YJKY1901)

第一作者: 吴 萌(1981—), 女, 陕西人, 硕士。

算法流程。

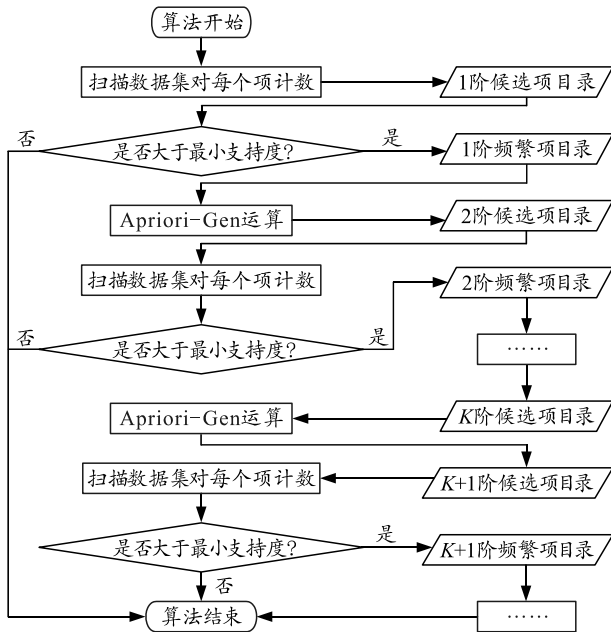


图 1 Ariori 算法流程

在 Ariori 算法中涉及到部分概念：所有项集  $I$ -items，表示所有项目的集合，记录 Transaction，记为  $T$ ；项集 Itemset，同时出现的项集合，表示为  $K$ -itemset ( $K$ -项集)；支持度 Support，置信度 Confidence；候选集 Candidate，根据合并计算生成的项集，为  $C_k$ ；频繁集：Frequent itemset，支持度大于最小支持度集合，表示为  $L_k$ ，提升度 Lift。确定各类参数的计算公式为：

支持度的计算：

$$\text{sup } p(X) = \text{occur}(X) / \text{count}(D) = P(X)。 \quad (1)$$

置信度计算：

$$\text{conf}(X \rightarrow Y) = \frac{\text{sup } p(X \cup Y)}{\text{sup } p(X)} = P(Y|X)。 \quad (2)$$

提升度的计算为：

$$\text{lift}(X \rightarrow Y) = \text{lift}(Y \rightarrow X) = \frac{\text{conf}(X \rightarrow Y)}{\text{sup } p(Y)} = \frac{\text{conf}(Y \rightarrow X)}{\text{sup } p(X)} = \frac{P(X \text{ and } Y)}{P(X)P(Y)}。 \quad (3)$$

### 1.2 算法的改进

Apriori 算法的挖掘过程主要分为 2 部分：1) 在数据库中寻找所有频繁项集  $L$ ；2) 从频繁集  $L$  提取强规则<sup>[10]</sup>。其中，步骤 1) 是评价算法优劣的关键，通常对 Apriori 算法的改进点<sup>[11]</sup>，在步骤 1) 中，首先生成候选项集  $L$ ，在产生的候选项集  $L$  中进行剪枝，寻找频繁集  $L$ 。

从 Apriori 算法可以看出：1) 算法每进行一次

迭代就需要对数据库整体进行一次扫描，若最大频繁项长度为  $N$  时，则需要进行  $N$  次扫描，而数据库存在着海量数据，频繁的数据遍历扫描将会对计算机内存和硬件提出更高的要求，且随着数据记录的增加，系统资源呈现几何级正常，因此到后面数据挖掘的效率将大大降低。2) 在不断迭代计算中，由两两组合形成候选频繁项集，但并没有排除有问题组合元素，造成频繁项集数据庞大，这就需要 CPU 和内存进行数据间的频繁交换，导致算法在深度和广度上适应性较差。

针对算法的缺陷进行改进，首先用矩阵来表示项集数据和运算，即将项集  $L$  的矩阵化表示为：

$$D = |D_1 \ D_2 \ \dots \ D_n| = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & & \vdots \\ d_{m1} & d_{m2} & \dots & d_{mn} \end{bmatrix}。 \quad (4)$$

式中各单元  $d_{ij}$  分别赋值为 0 或 1，将复杂数据变为矩阵中 0 或 1 的处理，对项集中大量数据逻辑运算修改为对矩阵 0 或 1 的运算，则项集  $[J_i, J_j]$  运算表达式为：

$$D_{ij} = D_i \wedge D_j = |d_{i1} \wedge d_{j1} \quad d_{i2} \wedge d_{j2} \quad \dots \quad d_{in} \wedge d_{jn}|。 \quad (5)$$

为生成  $L_k$ ，对  $L_{k-1}$  执行自连接操作，获得候选项集  $C_k$ ，即  $I_1$  和  $I_2$  作为  $L_{k-1}$  的 2 个可连接项集，要满足条件：

$$\left. \begin{aligned} I_1 &= L_{[1]}, L_{[2]}, \dots, L_{[k-2]}, L_{[k-1]} \\ I_2 &= L_{[1]}, L_{[2]}, \dots, L_{[k-1]}, L_{[k]} \\ I &= I_1 \circ I_2 = L_{[1]}, L_{[2]}, \dots, L_{[k-1]}, L_{[k]} \in L_k \end{aligned} \right\}。 \quad (6)$$

通过自连接避免了对整个数据库集的重复扫描，有效节约了系统资源。

对于出现次数大于最小支持度的项集与事物总数乘积得到频繁项集，则频繁项集  $k$ -中元素一定属于候选项  $C_k$ ，但同时部分非频繁项集元素也在  $C_k$  中，此时  $C_k$  作为  $L_k$  的超集。为减少  $C_k$  中庞大的数据量，对  $C_k$  进行压缩，删除其中的非频繁项。根据 Apriori 性质，当一个子集为非频繁的，则其超集也是非频繁的，定义非频繁项删除规则为：若一个候选  $k-1$ -子集不在  $L_{k-1}$  中，则该  $k$ -项集为非频繁的，从  $C_k$  中剔除。

## 2 监狱警察心理健康测试数据挖掘

### 2.1 数据样本

以某市监狱警察个人基本信息即参与心理健康

测试结果为数据库，选取数据库中“内向(A)、独立(B)、抑郁(C)、焦虑(D)、偏执(E)、坚决(F)”6类心理健康因子为对象，根据性格测试调查、症状自评测试和阿特尔16种个性因素测试，生成100道试题。从监狱系统随机选择9个典型事例进行健康测试，表1为9个典型事例的心理健康问卷统计结果，其中“1”表示“是”“0”表示“否”，给定的阈值  $minsup=3$ 。挖掘各因子对其他因子的影响程度，为心理健康工作队监狱狱警心理疏导提供指导。

表 1 心理健康测试问卷统计结果

| 序号 | A | B | C | D | E | F |
|----|---|---|---|---|---|---|
| 1  | 1 | 1 | 0 | 1 | 0 | 0 |
| 2  | 0 | 1 | 0 | 0 | 1 | 1 |
| 3  | 0 | 0 | 0 | 1 | 1 | 0 |
| 4  | 1 | 1 | 0 | 0 | 1 | 0 |
| 5  | 1 | 1 | 1 | 0 | 0 | 1 |
| 6  | 1 | 0 | 0 | 0 | 1 | 0 |
| 7  | 0 | 1 | 0 | 1 | 0 | 1 |
| 8  | 0 | 1 | 1 | 0 | 1 | 1 |
| 9  | 1 | 1 | 0 | 0 | 1 | 0 |

根据上表建立数据矩阵：

$$D=[D_1, D_2, \dots, D_9]=\begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (7)$$

首先生成频繁项集  $L_1$ ，计算支持度计数： $support\_count(I_1)=5$ ，相应计算得到  $I_2、I_3、I_4、I_5、I_6$  的支持度计数分别为 6、7、4、5、6。由于支持度计数均大于 3；因此，根据计算支持度计数生成频繁项集  $L_1$ ：

$$L_1=\{I_1:5, I_2:6, I_3:7, I_4:4, I_5:5, I_6:6\} \quad (8)$$

利用  $L_1$  生成候选集  $C_2$ ：

$$C_2=\left\{I_1I_2, I_1I_3, I_1I_4, I_1I_5, I_1I_6, I_2I_3, I_2I_4, I_2I_5, I_2I_6, I_3I_4, I_3I_5, I_3I_6, I_4I_5, I_4I_6, I_5I_6\right\} \quad (9)$$

由于  $D_{12}=D_1 \wedge D_2=[0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$ ，确定支持度计算  $support\_count(I_1I_2)=2$ ，计算  $C_2$  其他项支持度计数，并去除支持度小于 3 的项集中，生成频繁项集  $L_2$  为：

$$L_2=\left\{I_1I_3:4, I_1I_4:3, I_1I_6:3, I_2I_3:5, I_2I_4:4, I_3I_4:4, I_3I_5:3, I_3I_6:4, I_4I_6:3\right\} \quad (10)$$

对  $L_2$  进行自连接运算，获得候选项集  $C_3$ ：

$$C_3=\left\{I_1I_3I_4, I_1I_3I_6, I_1I_4I_5, I_1I_4I_6, I_2I_3I_4, I_2I_3I_5, I_2I_3I_6, I_2I_4I_5, I_2I_4I_6, I_3I_4I_5, I_3I_4I_6, I_3I_5I_6\right\} \quad (11)$$

考虑到  $\{I_2I_6, I_3I_5, I_4I_5, I_4I_6\}$  为非频繁 2-项集，因此将其从候  $C_3$  中剔除，更新  $C_3$  候选项集，仅需考虑  $\{I_1I_3I_4, I_1I_3I_6, I_1I_4I_6, I_2I_3I_4, I_3I_4I_6\}$  支持度：

$$D_{134}=(D_1 \wedge D_3) \wedge D_4=[1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0]^T \quad (12)$$

确定支持度计数  $support\_count(I_1I_3I_4)=3$ ，同样的计算其他项支持度计数均大于 3，根据算数支持度生成新的频繁集  $L_3$  为：

$$L_3=\{I_1I_3I_4:3, I_1I_3I_6:3, I_1I_4I_6:4, I_2I_3I_4:3, I_3I_4I_6:4\} \quad (13)$$

对  $L_3$  进行自连接运算，生成候选项集  $C_4=\{I_1I_3I_4I_6, I_2I_3I_4I_6\}$ ，其中  $\{I_2I_3I_6\}$  为非频繁的 3-项集，因此从  $C_3$  中剔除  $\{I_2I_3I_4I_6\}$ ，计算获得  $L_4=\{I_1I_3I_4I_6:3\}$ ，算法终止，获得 4 级频繁项集合。

### 2.2 结果分析

根据频繁项计算结果，图 2 为改进 Apriori 算法进行数据挖掘的部分监狱警察心理健康反馈结果。

由  $L_2=\left\{I_1I_3:4, I_1I_4:3, I_1I_6:3, I_2I_3:5, I_2I_4:4, I_3I_4:4, I_3I_5:3, I_3I_6:4, I_4I_6:3\right\}$  的频繁

项计数结果，其中“独立、抑郁”的心理特征较为突出，而对三阶频繁项的计数结果中可得， $L_3=\{I_1I_3I_4:3, I_1I_3I_6:3, I_1I_4I_6:4, I_2I_3I_4:3, I_3I_4I_6:4\}$ ，“抑郁、焦虑、坚强”的心理特征相较于其他 3 类较为突出。出现这种心理特征预工作环境和工作状态密切相关，监狱警察通常都受过专业心理训练，心理素质极强，具有非常强的业务能力和独立、坚强的性格特点，往往能够根据相应的规则来正确处理工作关系和个人关系。但随着工作年限的不断增加，监狱警察处于较为压抑、复杂的工作环境下，长期得不到外界的疏导，在内心容易出现“焦虑”感，这种心理状态一旦碰到棘手的工作问题和家庭关系时，往往会造成心理上造成负担，甚至出现“压抑”的心理亚健康状态。

由  $L_4=\{I_1I_3I_4I_6:3\}$  可以看出，在监狱警察群体中，存在着普遍的“内向、抑郁、焦虑、坚强”的心理特征突出，通过解决其中一方面问题，对其他心理问题也具有一定的改善作用，弥补性格方面的缺陷。如内向的人做事认真、踏实、团队性，在日常工作中需要培养其活泼开朗、积极乐观的情绪，让其积极参加社会活动和人际交换活动，加强户外

锻炼,多接触外界世界,有利于改善“内向、抑郁、焦虑”的性格特点。

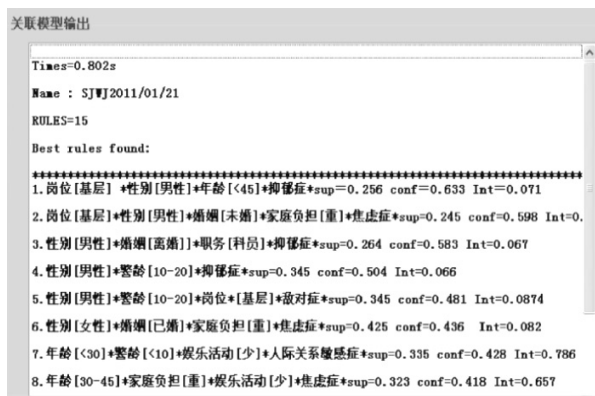


图2 监狱警察心理特征反馈结果

### 3 结论

笔者利用 Apriori 改进算法,通过数据挖掘发现监狱警察存在的心理健康问题,采用针对性措施解决监狱警察心理健康问题。实践结果证明:通过对数据库集中心理健康调查结果进行矩阵化处理和和运算,采用自连接操作方式删除非频繁项集,能大幅降低数据计算量,在保证计算结果精度的同时提升计算效率。通过将改进 Apriori 算法进行监狱警察心理健康调查可以看出:受工作内容和工作环境影响,监狱警察普遍存在“内向、抑郁、焦虑、坚强”的心理特征,需加强户外锻炼,多接触外界事物,培养活泼开朗、积极乐观的情绪。

\*\*\*\*\*

(上接第 15 页)

身管的冷却可以根据冷却剂的不同分为气冷、液冷和化学试剂冷却 3 种,而按照冷却作用方式的性质可分为连续冷却和间断冷却。间断冷却又可以分为单发间冷却和连发间冷却。通过采用一系列结构性措施,如增大身管的重量、增加身管的冷却面积、采用可速换的结构以及射后向膛吹气等方式,可以降低身管发热量并提高空气冷却效率。

在身管设计时,提高身管寿命<sup>[6]</sup>,主要采取技术措施如下:

- 1) 采用高能、低温、低烧蚀的发射药。
- 2) 采用大药室、低膛压发射。
- 3) 装药中加入缓蚀剂。
- 4) 加强身管散热(冷却等)。
- 5) 加强内膛表面处理(镀铬等耐蚀、耐磨覆层)。
- 6) 改进弹丸弹带结构和材料。
- 7) 改进身管材料(高强度)。
- 8) 改进身管制造工艺(自紧,活动衬管等)。

### 参考文献:

- [1] 胡珊. Apriori 算法改进及其在大学生心理健康测试中的应用[J]. 自动化与仪器仪表, 2016(6): 222-224.
- [2] 钱程. Apriori 算法在大学生情感素质研究中的改进与应用[D]. 上海: 上海师范大学, 2019: 23-24.
- [3] 江家龙, 秦亮曦. C4.5 算法在大学生心理健康分析中的应用[J]. 现代计算机(专业版), 2016(21): 15-19.
- [4] 陈婉. 数据挖掘在高职学生心理健康管理系统中的应用研究[D]. 郑州: 郑州大学, 2017: 18-21.
- [5] 孙伟平. 决策树技术在大学生心理健康测评中的应用研究[D]. 郑州: 郑州大学, 2020: 31-33.
- [6] 杨昱梅, 李婧. 聚类分析算法在大学生心理健康分析中的应用研究[J]. 中国教育学报, 2015(S1): 27-29.
- [7] 郭友倩. Apriori 算法在心理健康测试分析中的应用研究[J]. 现代经济信息, 2015(12): 380-381.
- [8] 吴小刚, 周萍, 彭文惠. 决策树算法在大学生心理健康评测中的应用[J]. 计算机应用与软件, 2018, 28(10): 240-244.
- [9] 熊祖涛. 基于集成学习算法的高职生心理健康预测方法[J]. 软件导刊, 2016, 15(6): 162-164.
- [10] 晏杰. 决策树算法的研究及其在大学生心理健康数据处理中的应用[J]. 江汉大学学报(自然科学版), 2015, 43(4): 371-375.
- [11] 何晓薇. 基于 MD5 算法的大学生心理健康跟踪系统设计及实现[J]. 计算机光盘软件与应用, 2014, 17(13): 268-270.

### 5 结束语

笔者分析身管射击中的能量转换,对射击过程中身管吸收的热量提供一种简便算法。以某科研产品身管数据进行了验证,分析身管发热的影响,介绍身管烧蚀寿命以及烧蚀寿命终止标准、冷却措施以及提高身管寿命措施,可供大家参考使用。

### 参考文献:

- [1] 郭张霞. 炮身设计[M]. 北京: 国防工业出版社, 2010.
- [2] 高跃飞. 弹道学[M]. 北京: 国防工业出版社, 2010.
- [3] 耿晓虎, 韩崇伟, 马捷, 等. 基于电压反馈的火炮数据处理模块热插拔方法[J]. 兵工自动化, 2023, 42(10): 21-24.
- [4] 杨国来. 火炮发射动力学概论[M]. 北京: 国防工业出版社, 2018.
- [5] 谈乐斌. 火炮概论[M]. 北京: 北京理工大学出版社, 2014.
- [6] 张相炎. 火炮设计理论[M]. 北京: 北京理工大学出版社, 2014.