

doi: 10.7690/bgzdh.2026.06.002

基于 Apriori 算法的装备试验数据挖掘应用

胡春宇, 蒋英杰, 王付忠

(中国人民解放军 96753 部队, 济南 271100)

摘要: 针对因装备作战试验数据来源分散、标准不统一、数据量大导致的难以共享和使用等问题, 基于 Apriori 算法对装备数据挖掘进行研究。介绍关联规则挖掘方法, 并运用到装备试验数据挖掘中, 对装备试验的数据进行分析, 挖掘出试验数据的联系和规律。实践结果表明, 通过数据挖掘, 能够为指挥员分析判断装备信息和评估导弹武器作战效能提供一定决策依据。

关键词: Apriori 算法; 装备试验; 数据挖掘

中图分类号: TP274+.2; E917 **文献标志码:** A

Application of Equipment Test Data Mining Based on Apriori Algorithm

Hu Chunyu, Jiang Yingjie, Wang Fuzhong

(No. 96753 Unit of PLA, Ji'nan 271100, China)

Abstract: Aiming at the problems of the challenges in sharing and using equipment combat test data caused by dispersed data sources, non-uniform standards, and large data volumes, research on equipment data mining based on the Apriori algorithm is conducted. The method of association rule mining is introduced and applied to the mining of equipment test data, enabling the analysis of equipment test data and the discovery of relationships and patterns within the data. The practical results show that data mining can provide commanders with a certain decision-making basis for analyzing equipment information and evaluating the combat effectiveness of missile weapons.

Keywords: Apriori algorithm; equipment test; data mining

0 引言

装备试验是验证和评价装备的作战效能、作战适用性和生存性的试验鉴定活动, 特点是必须要在真实环境中进行^[1], 但由于装备研制生产复杂、采购价格昂贵、试验准备周期较长、试验危险性大、安全保密要求高等特点, 不可能对其进行频繁试验, 有时会将试验任务与实际工作结合起来, 以真实的数据标准检验装备性能, 由此产生的数据作试验更具有说服力^[2]。

在试验中产生的一系列数据是评估装备性能的有力凭证。目前, 装备试验数据来源分散, 大多来源于各试验组织单位, 共享和使用数据困难, 且数据量大、标准不统一, 不利于指挥决策人员分析装备性能^[3]。综上所述, 研究装备作战试验数据挖掘, 对于工作人员和工业部门掌握装备性能、状态, 评估装备性能以及研制生产和改进有着重要意义, 同时通过数据挖掘, 也能为指战员提供战场决策。

笔者通过采集和积累某型装备在作战试验中的数据, 引入 Apriori 算法, 对作战数据进行关联性挖掘, 通过相应的计算和分析, 发现隐藏于数据之间

的联系, 为分析判断装备作战试验效能评估提供数据支撑, 为指挥员提供决策^[4]。

1 关联规则及 Apriori 算法介绍

1.1 关联规则概述及基本概念

数据的关联规则是指通过分析和寻找数据之间隐藏的、有使用价值的关联性, 用一定的规则显示出来, 经过计算机处理和数学计算, 得出重要的结论。为便于理解关联规则, 首先介绍相关概念: 假设 $D=\{T_1, T_2, \dots, T_i, \dots, T_n\}$ 表示一个事务数据库, 其中 $T_i(i=1, 2, \dots, n)$ 表示第 i 个事务。设 $I=\{I_1, I_2, \dots, I_j, \dots, I_m\}$ 表示所有项的集合, 其中 $I_j(j=1, 2, \dots, m)$ 称作一个项目。若 $X \subseteq I$, 则称 X 为一个项集。包含 k 个项的项集称为 k -项集。每个事务具有一个唯一标识符, 记为 TID。某型装备发射数据如表 1 所示。

表 1 某型装备试验数据

T	装备控制能力	装备突防概率	装备成功率
1	优	差	差
2	良	优	良
3	差	良	差

收稿日期: 2024-12-03; 修回日期: 2025-01-03

第一作者: 胡春宇(1987—), 男, 山西人, 硕士。

表 1 中所有数据表示为 D ，包含 3 个事务即 $T=\{1, 2, 3\}$ ， $I=\{\text{装备控制能力, 装备突防概率, 装备成功率}\}$ ，项集 $X=\{\text{装备控制能力, 装备发射成功率}\}$ 为 2-项集。

若 $X, Y \in I$ ，且 $X \cap Y = \emptyset$ ，则将 $X \Rightarrow Y$ 定义为关联规则。关联规则的相关性强度一般用随对应的支持度和置信度来衡量^[6]。

支持度 (support) 是 D 中同时包含 X, Y 的概率，表示 $X \Rightarrow Y$ 在整个数据中的重要程度^[7]：

$$\text{support}(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{|D|} \times 100\% \quad (1)$$

置信度 (confidence) 是 D 中有 X 的情况下，包含 Y 的概率，表示 $X \Rightarrow Y$ 在整个数据中的可信程度：

$$\text{confidence}(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \times 100\% \quad (2)$$

严格来讲，只有在支持度和置信度都达到特定的数值后，数据间的关联规则才有效。用户在实际使用过程中，一般需要结合实际情况设置最小支持度 (minsup) 和最小置信度 (minconf)^[8]。假如 $\text{support}(X \Rightarrow Y) \geq \text{minsup}$ ，则称 $\{X, Y\}$ 为频繁项目集，反之则称之为非频繁项目集。

1.2 Apriori 算法介绍

Apriori 算法是十种数据关联规则挖掘中的较为经典的算法之一，被广泛应用于大数据挖掘中：首先，找出试验所需的数据库，由用户结合实际情况设置最小支持度和最小置信度，再对数据库进行全面扫描，分别统计每个项出现的次数，生成候选 1 项 C_1 ，删除小于最小支持度和最小置信度的项目，找到频繁 1-项集，记为 L_1 ；然后，将生成的频繁项集数据进行两两比较，生成候选 2 项集 C_2 ，再次扫描 L_1 ，找出频繁项 2 的集合，记为 L_2 ，再使用频繁 2-项集找出 L_3 ，以此类推，直到无法找到更多的频繁 k -项集，找出每个 L_k 需要扫描一次数据库。算法流程如图 1 所示。

1.3 数据挖掘流程

采用 Apriori 算法进行数据挖掘，其过程如图 2 所示。

1) 数据选择。在数据源选择方面，跟试验相关的各种因素作为选择和采集的对象，以便增加数据的有效性和权威性。数据选择主要依托人工收集、装备自行记录和各类传感器敏感元件直接反馈的数

据信息等。

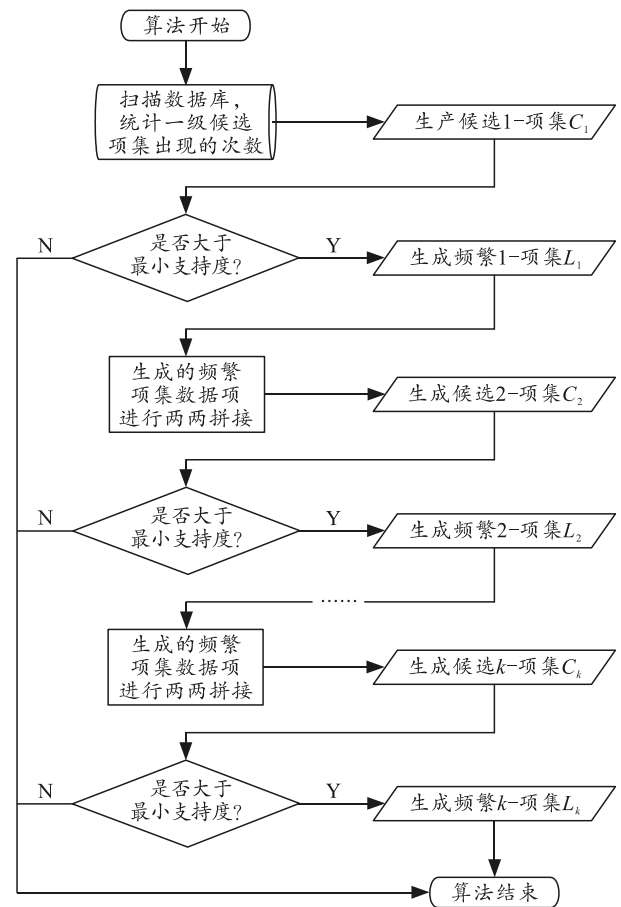


图 1 Apriori 算法搜索频繁项目集流程

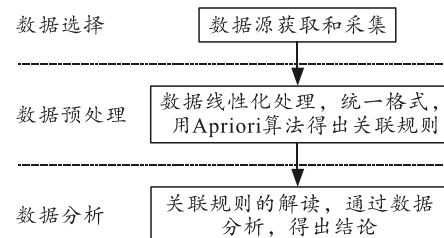


图 2 数据挖掘的流程

2) 数据预处理。由于试验数据结构繁多，标准不统一，有些数据不能直观反映本质现象，通过简单的数学运算和计算机处理，使得数据按照设定的规则进行处理，得到所需的数据格式，应用 Apriori 算法对数据关联性分析，输出最终结果。

3) 数据分析。根据得出的关联规则，结合实际情况，分析判断对装备的相关信息，指导后续工作任务。

2 在装备试验大数据挖掘的应用

2.1 试验数据

根据某装备试验中产生的数据，以及需要分析和了解潜在的关系，笔者选取某型装备试验中采集

的数据，作为数据挖掘的对象。数据内容主要包括领导能力素质、技术人员水平、成员具备的专业能力、装备可用率、装备精度、天气情况、瞬时风速、装备可靠率等，数据统计如表 2 所示。

表 2 原始数据记录

编号	领导能力素质	技术人员水平	成员专业水平度	装备可用率	装备精度	天气	风速/(m/s)	装备可靠率
1	高	高	1	1.00	0	晴	0	高
2	高	高	2	0.90	1	晴	0.1	高
3	低	高	3	0.95	2	阴	0.5	高
4	低	低	1	1.00	5	阴	0	低
5	低	低	2	1.00	5	小雨	1.2	低
6	高	高	3	0.90	4	晴	0.9	高
7	高	低	1	1.00	1	阴	1.0	低
8	低	低	2	0.90	5	小雨	1.5	低
∴	∴	∴	∴	∴	∴	∴	∴	∴

2.2 数据预处理

由于涉及试验数据的指标较多，而且各类数据之间的属性大都不统一^[3]。根据表 2 可知：数据都是不规则的、离散的，为方便后续数据分析和关联性挖掘，对各项数值按照关联规则进行离散性处理，得到处理后的数据如表 3 所示。

表 3 处理后数据

编号	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8
1	K_1	K_3	K_6	K_{11}	K_{12}	K_{16}	K_{19}	K_{26}
2	K_1	K_3	K_6	K_{10}	K_{13}	K_{16}	K_{20}	K_{26}
3	K_2	K_3	K_7	K_{11}	K_{14}	K_{17}	K_{21}	K_{26}
4	K_1	K_4	K_8	K_{11}	K_{15}	K_{17}	K_{19}	K_{27}
5	K_2	K_4	K_5	K_{11}	K_{15}	K_{18}	K_{24}	K_{27}
6	K_1	K_3	K_7	K_{10}	K_{15}	K_{16}	K_{27}	K_{26}
7	K_1	K_4	K_8	K_{11}	K_{13}	K_{17}	K_{23}	K_{27}
8	K_2	K_4	K_5	K_{10}	K_{15}	K_{18}	K_{25}	K_{27}
∴	∴	∴	∴	∴	∴	∴	∴	∴

领导能力素质(I_1): K_1 : 高; K_2 : 低。

技术人员能力(I_2): K_3 : 高; K_4 : 低。

成员专业水平(I_3): K_5 : $I_3 < 1$; K_6 : $1 \leq I_3 < 2$;

K_7 : $2 \leq I_3 < 3$; K_8 : $I_3 = 3$ 。

装备可用率 I_4 : K_9 : $I_4 < 0.90$;

K_{10} : $0.90 \leq I_4 < 0.95$; K_{11} : $0.95 \leq I_4 \leq 1$ 。

装备精度 I_5 : K_{12} : $I_5 < 1$; K_{13} : $1 \leq I_5 < 2$;

K_{14} : $2 \leq I_5 < 4$; K_{15} : $4 \leq I_5 \leq 5$ 。

天气 I_6 : K_{16} : 晴; K_{17} : 阴; K_{18} : 小雨。

风速 I_7 : K_{19} : $I_7 < 0.1$; K_{20} : $0.1 \leq I_7 < 0.5$; K_{21} : $0.5 \leq I_7 < 0.9$; K_{27} : $0.9 \leq I_7 < 1.0$; K_{23} : $1.0 \leq I_7 < 1.2$;

K_{24} : $1.2 \leq I_7 < 1.5$; K_{25} : $I_7 \geq 1.5$ 。

装备可靠性 I_8 : K_{26} : 高; K_{27} : 低。

2.3 数据挖掘

采用 Apriori 数据挖掘算法，挖掘其关联性，为使试验效果更加直观，更能反映数据间的隐藏关系，分别将最小支持度和最小置信度设为 20% 和 85%。对所有数据进行扫描和分析，得出所有频繁项集，得出部分关联规则如图 3 和 4，以及表 4 所示。

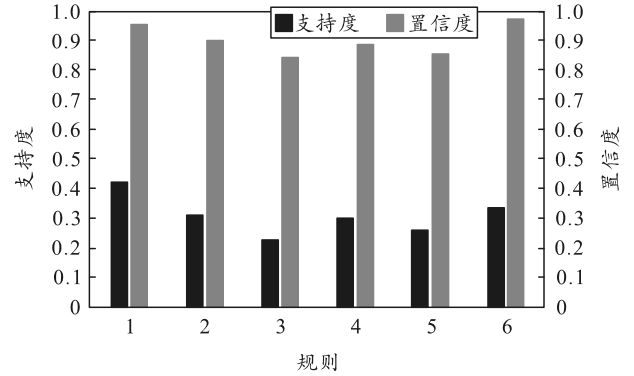


图 3 关联规则

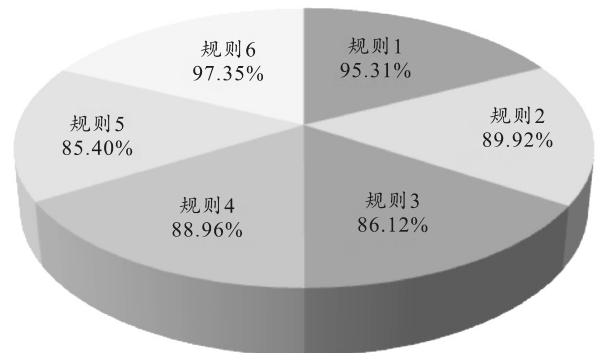


图 4 基于关联规则的置信度

表 4 数据关联结果

编号	关联规则	支持度	置信度
1	$K_1 K_3 K_{11} \Rightarrow K_{12} K_{26}$	42.40	95.31
2	$K_1 K_6 \Rightarrow K_{26}$	31.20	89.92
3	$K_2 K_5 \Rightarrow K_{15}$	22.30	86.12
4	$K_{10} K_{15} \Rightarrow K_{27}$	30.12	88.96
5	$K_3 K_{11} K_{16} \Rightarrow K_{26}$	26.30	85.40
6	$K_5 K_{18} K_{24} \Rightarrow K_{15}$	33.60	97.35
∴	∴	∴	∴

表 4 中的关联规则：

规则 1：领导者能力素质高，且技术人员水平高，装备性能好，装备精度高的情况下，装备可靠率达到 95.31%，表面装备质量性能完好。

规则 2：领导者能力素质高，成员专业水平度高，则装备可靠率有 89.92%。

规则 3：领导者能力素质低，且班组成员水平度低下，则有 86.12% 的概率在操作中精度偏差较大。

规则 4：装备性能差且精度偏差较大的情况下，

装备可靠率较低。

规则 5：技术人员能力水平高且装备精度高，天气状况好的情况下，装备可靠率较高。

规则 6：成员能力水平不过硬，天气条件较差的情况下，装备可靠率较低。

2.4 数据分析与解释

分析规则 1、2 和 5，表明领导者能力素质高，成员专业素质过硬，各类技术专家、装备精度、天气条件等外在因素较好的情况下，装备可靠率较高。说明一个成熟的领导者，对试验需要掌握的必备技能成竹于胸，注重成员的专业水平、装备质量管理、技术人员的能力素质、天气条件等各种影响和制约装备可靠率的因素，并能根据结果做出改进补偿措施。

分析规则 3 和 4，表明成员专业素质较低和装备性能相对较差，导致装备成品精度较低，最终装备可靠率低下。说明不注重培养成员的专业，装备管理能力差，各项准备工作不拖底，出现一系列紧张心理和不懂不会现象，操作上出现种种故障，致使装备可靠率不高。

综合分析结果，装备可靠率与人员专业能力素质、装备性能精度、天气条件等元素息息相关，不管是在日常培训还是实际作业，一方面要抓好人才的培训力度和传帮带，另一方面抓好与装备息息相关的技术条件。

2.5 算法评价

通过 Apriori 算法在实际案例中的应用，并检验效果，能够直观地反映试验数据之间的关联性，结

论通俗易懂，具有很强的指导意义。在实际作战试验运用中，由于数据源的获取途径有限，加之数据采集中存在一定的漏洞，影响结果的分析，但随着数据量的增加，在大数据的计算功能下，该算法的误差将会越来越小。

3 结束语

笔者从装备试验的数据出发，通过数据挖掘计算，得出影响装备可靠率的部分因素，结果符合客观情况。实践结果表明：通过分析试验数据之间的关系，能够在制定项目研究方案、设计路线、产品规划等方面提供参考依据。

参考文献：

- [1] 薛晓君. 作战保障基于大数据的作战指挥决策刍议[J]. 国防信息学院学报, 2015, 158(3): 1-2.
- [2] 李建平, 曹炳华, 任逸然. 大数据在未来作战中的应用研究[J]. 网信军民融合, 2018(12): 22-24.
- [3] 贾晨星, 李立纲, 任重. 强军新时代军队作战数据建设[J]. 国防科技, 2018, 39(6): 48-52.
- [4] 左钦文, 张杰民, 刘晓宏, 等. 基于大数据及机器学习的智能作战评估方法[J]. 兵器装备工程学报, 2020, 41(2): 107-110.
- [5] 孙媛妮, 吕学义, 王申坪. 装备作战试验指标体系构建问题研究[J]. 价值工程, 2019, 38(20): 243-244.
- [6] 郭宝宝, 刘森琪, 杨章勇, 等. 作战实验中大数据技术运用探讨[J]. 科技创新与应用, 2019(25): 153-154.
- [7] 朱子薇, 朱元武, 范玲瑜, 等. 作战仿真数据在效能评估中的运用[J]. 火力与指挥控制, 2019, 44(5): 161-164.
- [8] 范绍华, 谭凯家. 基于“云作战”样式的武器装备发展探析[J]. 国防, 2019(3): 17-20, 24.