

# 基于关联规则的柴油机装配冷试异常数据分析

赵旭辉<sup>1</sup>, 徐卓<sup>1</sup>, 王辉<sup>2</sup>, 闫伟<sup>1\*</sup>, 李国祥<sup>1</sup>

1. 山东大学核科学与能源动力学院, 山东 济南 250061; 2. 潍柴控股集团有限公司, 山东 潍坊 261001

**摘要:**为分析某柴油机装配冷试试验中异常数据间的关系,提出一种基于关联规则的装配冷试异常数据分析方法,采用 KIH-means 聚类方式,针对该柴油机装配异常构建关联规则样本库;通过支持度-置信度试验,确定最佳的最小支持度和最小置信度阈值组合;采用 Apriori 算法进行关联规则挖掘,揭示异常数据的关联关系,得到针对该柴油机装配异常的挖掘结果。结果表明:该柴油机最小置信度为 0.75,最小支持度为 0.225% 为最佳阈值组合,此时关联规则条数变化最平稳;KIH-means 聚类方式与 Apriori 算法结合,通过对关联规则样本库数据挖掘,确定异常参数的关联规则,提出有效优化措施,从而提高发动机装配质量及装配性能一致性。

**关键词:**柴油机;冷试测试;异常检测;关联规则

**中图分类号:**TK422

**文献标志码:**A

**文章编号:**1673-6397(2025)06-0065-06

**引用格式:**赵旭辉,徐卓,王辉,等.基于关联规则的柴油机装配冷试异常数据分析[J].内燃机与动力装置,2025,42(6):65-70.

ZHAO Xuhui, XU Zhuo, WANG Hui, et al. Association rules based on analysis of anomalies in a diesel engine assembly cold test [J]. Internal Combustion Engine & Powerplant, 2025, 42(6): 65-70.

## 0 引言

柴油机是挖掘机、装载机、推土机等工程机械设备的主要动力源,其性能直接影响整机的工作效率与使用寿命。为确保柴油机性能稳定,通常需要进行包括冷试测试的各种出厂测试。柴油机冷试时,发动机无需点火,由电机倒拖运行,并通过冷试台架上的传感器采集测试参数,该测试方式具有节能环保、污染小、成本低、故障定位准确等优势,在行业中得到了广泛应用。

常规的冷试数据依赖专家经验和先验知识进行分析,存在诊断效率低、误诊率高等问题。随着智能传感器的发展,柴油机出厂测试获得的数据越来越多,亟需使用科学的数据处理分析方法,快速精准地识别异常数据,从而排查出故障问题。研究人员采用多种方法对冷试数据进行了深入研究: Okhionkpanwonyi 等<sup>[1]</sup>针对柴油机冷试中的轴瓦故障检测问题,对振动监测与声发射技术性能进行试验对比,发现声发射技术能更准确指示故障;Guardiola 等<sup>[2]</sup>提出了一种内燃机冷试系统,并借助非侵入式传感技术实现对燃烧系统硬件故障的全方位检测;康娜等<sup>[3]</sup>通过优化链条短滑轨尺寸,解决了柴油机冷试噪声、振动与声振粗糙度(noise vibration harshness, NVH)测试的噪声异常问题;孙俊伟<sup>[4]</sup>构建了针对柴油机冷试的智能质量监控体系,并以冷试振动参数为基础进行了故障诊断;Hernández-nuñez 等<sup>[5]</sup>分析和建立了气缸盖加工偏差与冷试不合格格数据间的相关性,以预测气缸盖失效的原因;Dong 等<sup>[6]</sup>基于发动机冷试中参数的信号特性,发现通过合理的参数估计方法可以有效消除样本误差,获得更可靠的冷试参数

**收稿日期:**2025-09-30

**基金项目:**国家重点研发计划项目(2021YFD2000301);山东省工程机械智能装备创新创业共同体项目(GTT20240101);山东省土方机械智能化施工技术重点实验室项目(PKL2024F13)

**第一作者简介:**赵旭辉(2002—),男,山西长治人,硕士研究生,主要研究方向为发动机冷试技术,E-mail:202434567@mail.sdu.edu.cn。

\* **通信作者简介:**闫伟(1973—),男,济南人,工学博士,教授,博士生导师,主要研究方向为新能源汽车技术、发动机冷试技术等,E-mail:yanwei@sdu.edu.cn。

的控制极限;罗云<sup>[7]</sup>通过调整变量泵参数,使其更好地匹配了发动机油路特性,从而有效降低了冷试测试故障率;杨明川等<sup>[8]</sup>采用可编程逻辑控制器,优化了传统设备,提高了柴油机装配冷试台架的柔性化和智能化水平。

关联规则挖掘是一种用于描述不同事物之间的关联性,从而揭示其中隐含的关系和模式的数据分析方法。在冷试中,可利用其分析冷试数据构建规则数据库,该方法也在汽车行业应用并获得了一系列成果;Ning等<sup>[9]</sup>采用Apriori算法,对汽车装配车间现场数据中表示问题、原因与解决方案的文档集群进行关联挖掘,构建了面向装配故障诊断的知识映射体系;Dou等<sup>[10]</sup>基于某柴油机装配过程数据,构建了和产品质量相关的Simple K-means聚类和Apriori分析模型,有效提高了装配过程质量和效率;王海涛等<sup>[11]</sup>基于知识查询、关联规则和考核评判,构建了发动机维修知识图谱;Zhuang等<sup>[12]</sup>采用Apriori算法挖掘航空航天产品质量异常与失控装配系统之间的强关联规则,解决了质量异常成因复杂、难以追溯的问题。

本文中基于某柴油机装配冷试测试出厂数据,采用KIH-means聚类算法对数据进行初步分析,构建关联规则样本库;通过设置不同最小置信度和最小支持度组合的试验方案,确定最佳阈值组合;采用Apriori算法对样本库进行关联规则处理和分析,得到针对该柴油机装配异常的关联规则挖掘表。该研究对改进发动机装配质量、提高装配性能一致性具有重要的参考价值。

## 1 冷试测试数据库构建

某6缸柴油发动机冷试中采集每个气缸的运行参数,每个气缸采样点数为4 452,采样参数包含最大转矩,进、排气门开启位置,进、排气门开启时间,进、排气门关闭位置,进气门开启前真空度泄漏,进气门关闭后真空度泄漏,最大排气压力,最大进气真空度共11个参数,各参数用 $M_{p,q}$ 表示, $p$ 为采样点数, $p=1,2,3,\dots,4\ 452$ ; $q$ 对应采样参数, $q=1,2,3,\dots,11$ 。采样参数中进、排气门的开启时间、开启及关闭位置均以曲轴转角表示,进、排气门开启时间指进、排气门从开启到关闭过程对应的曲轴转角,测试项目界面如图1所示。

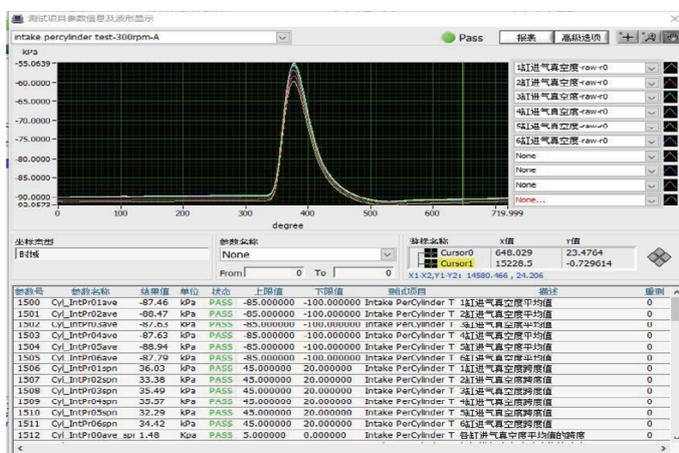
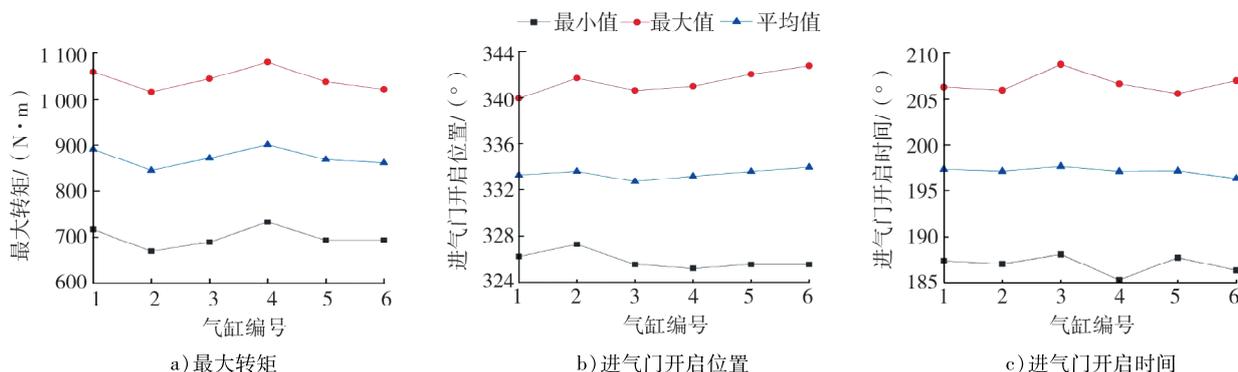


图1 测试项目界面

对采集数据进行统计分析,数据处理结果如图2所示。由图2可知:柴油机6个缸的各参数的最大值、最小值、平均值均不相同;同一气缸相同参数的最大值、最小值区别较大;各缸最大进气门开启前真空度泄漏及最大进气门关闭后真空度泄漏差异均较大。各缸工作不均匀造成发动机性能不一致,需要通过有效管控柴油机装配参数,减小柴油机异常工作状态。



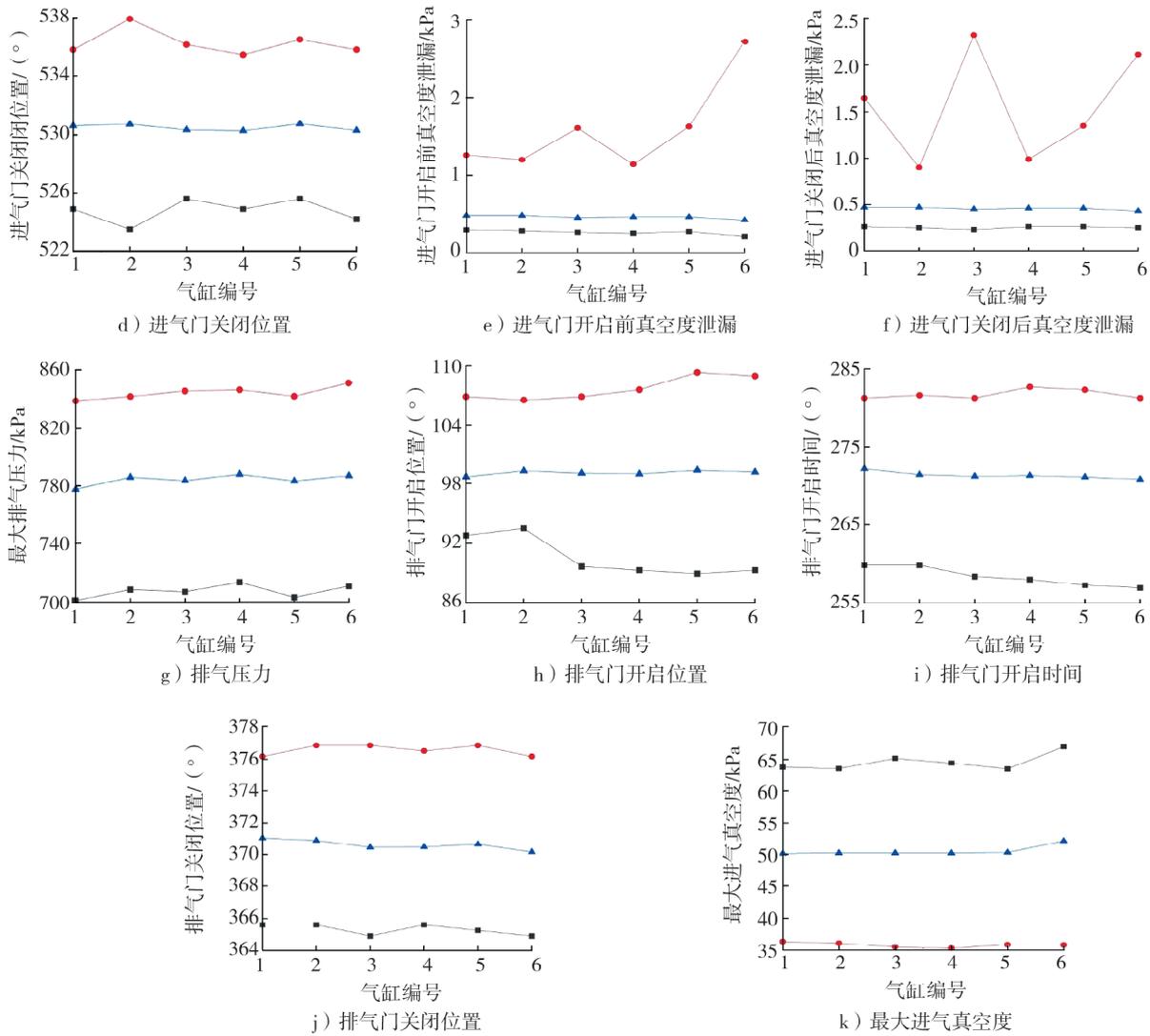


图2 数据处理结果

## 2 关联规则及样本库构建

### 2.1 关联规则

关联规则中的数据一般由布尔值(0和1)组成。通常,0表示无用数据,1表示分析数据<sup>[13]</sup>。数据集  $I = \{I_1, I_2, \dots, I_m\}$  为项集,其中的元素  $I_i (i = 1, 2, \dots, m)$  称为项;  $T = \{T_1, T_2, \dots, T_n\}$  是数据库的集合,其中的元素  $T_j (j = 1, 2, \dots, n)$  称为事务;每个事务是项的集合,即  $I \subseteq T_j$ 。关联规则的语义表达式  $X \Rightarrow Y$ ,其中  $X \not\subseteq I$  且  $Y \subseteq I$ ,  $X$  称为规则前项,  $Y$  称为规则后项,表示“当项集  $X$  出现时,项集  $Y$  也有可能同时出现”。

通常用支持度评价事务出现  $X$  的百分比,即概率。支持度

$$s_x = t_x / t_1, \tag{1}$$

式中: $t_x$  为项集  $X$  出现的次数,也称支持度计数; $t_1$  为所有项集的总次数。最小支持度为项集  $X$  出现次数需满足的最低要求,若低于最小支持度,该项集被淘汰;若不小于最小支持度时,表示该项集为一个频繁项集。

$X \Rightarrow Y$  置信度指  $T$  中事务在包含  $X$  的情况下,同时包含  $Y$  的百分比,即条件概率。 $X \Rightarrow Y$  对应的置信度

$$c = t_{X \cap Y} / t_x, \tag{2}$$

式中： $t_{X \cap Y}$  为项集  $X$  中出现项集  $Y$  的次数。当  $X \Rightarrow Y$  对应的  $c = 1$  时，表示  $X \Rightarrow Y$  完全成立，即项集  $Y$  必定在项集  $X$  的事务中出现；当  $c < 1$  时，表示  $X \Rightarrow Y$  的关联规则不一定成立，即  $X \Rightarrow Y$  关联规则可靠性相对较低。最小置信度是关联规则可信程度需满足的最低要求，如果同时满足最小支持度和最小置信度，则认为  $X$  和  $Y$  间存在强关联。

## 2.2 建立样本库

柴油机装配冷试出厂数据集样本量较大，无法对每一样本的每一项数据进行处理，且由于关联规则数据集一般由布尔值组成，因此需要寻找一种科学可靠的方法，以构建适合分析用的关联规则样本库。

本文中采用  $K$ -means 聚类方法<sup>[14]</sup> 对柴油机装配冷试出厂参数进行分析。为提高聚类方法的准确性，采用核函数<sup>[15]</sup> 及不同的适应度改进，得到 KIH-means 算法，具体方法为：使用高斯核函数、多项式核函数及 sigmoid 核函数改进  $K$ -means 聚类的每个点到  $K$  个初始聚类中心的欧氏距离，使用相似度比值、模糊熵<sup>[16]</sup>、马氏距离<sup>[17]</sup> 改进适应度函数。

以采样的 4 452 组样本数据为研究对象，由于柴油机为 6 缸，故将所有样本数据分为 6 簇，使用 KIH-means 算法得到聚类簇团，单独计算每个数据点的每一项到聚类中心的距离，将超出聚类簇团最外侧 5% 的数据定义为异常点，其布尔值设为 1，其余正常点的布尔值设为 0，取各测试参数  $M_{p,q}$  对应的布尔值  $B_{p,q}$  构建关联规则样本库，其中第 1 缸部分关联规则样本库如表 1 所示。

表 1 第 1 缸部分关联规则样本库

| $p$   | $B_{p,1}$ | $B_{p,2}$ | ... | $B_{p,11}$ |
|-------|-----------|-----------|-----|------------|
| 1     | 0         | 0         | ... | 0          |
| ...   | 0         | 0         | ... | 0          |
| 1 950 | 0         | 1         | ... | 1          |
| ...   | ...       | ...       | ... | ...        |
| 4 451 | 1         | 0         | ... | 0          |
| 4 452 | 0         | 0         | ... | 0          |

## 2.3 Apriori 算法

采用 Apriori 算法<sup>[18]</sup> 分析样本库数据，其算法流程为：1) 扫描全部数据，生成只包含 1 个项的候选项集  $C_1$ ；2) 定义最小支持度、最小置信度，最小支持度和最小置信度由支持度-置信度试验或专家经验确定；3) 根据最小支持度，从当前项数为  $k$  的候选项集  $C_k$  中产生频繁项集  $L_k$ ；4) 若  $L_k$  非空集，将频繁项集进行连接操作，生成候选项集  $C_{k+1}$ ，通过剪枝步多次迭代，直至扫描数据库验证，得到新的频繁项集  $L_k$ ；若  $L_k$  为空集，算法结束。算法流程图如图 3 所示。

Apriori 算法通过先找项数为 1 的频繁项集，逐渐增加项数，直至找不到新的频繁项集为止，结合剪枝步可显著优化迭代效率，为后续关联规则挖掘奠定基础。

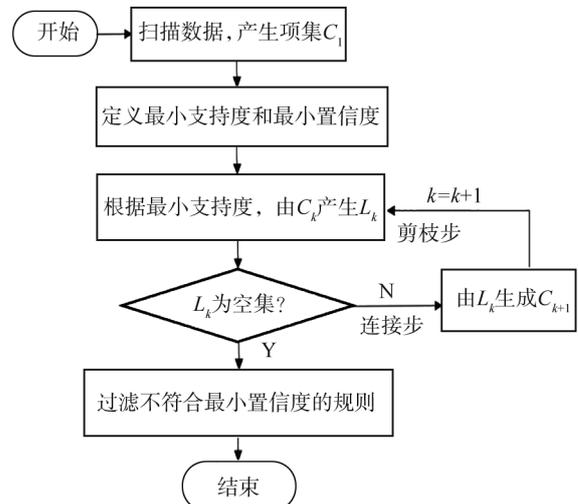


图 3 Apriori 算法流程图

## 3 数据分析

### 3.1 支持度-置信度试验

为了避免出现无意义的冗余规则或遗漏有价值的数据，同时平衡采样数据量和评价质量，本文中通过调整最小支持度和最小置信度，筛选可信度较高的不同项集之间的强关联规则。采用最小支持度对应的  $t_x$  和最小置信度  $c$  计算关联规则条数，选择关联规则条数变化最为平缓时的  $t_x$ 、最小置信度  $c$  组合作为最优阈值，最小置信度和最小支持度对应的选择结果如图 4 所示。

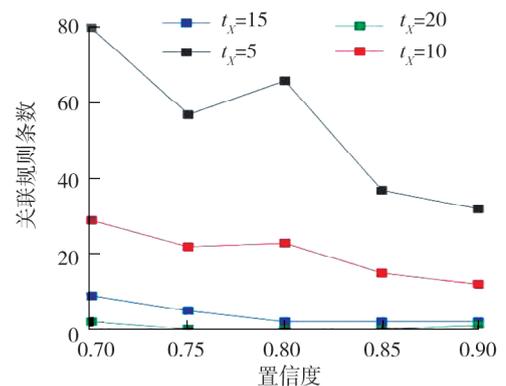


图 4 置信度和支持度阈值选择

由图4可知:当最小支持度对应的 $t_x$ 为10、15、20时,关联规则条数的变化较平稳;当 $t_x$ 为15、20时,基本没有可供参考的关联规则条数,无法提供有效的关联分析;当 $c=0.75$ 时,关联规则条数波动较小;最小置信度为0.75,最小支持度对应的 $t_x$ 为10时,关联规则条数变化最平稳,因此将这一组合确定为最佳阈值,此时对应的关联规则条数为22,由于项集总次数为4452,根据式1)计算,最小支持度为0.225%。

### 3.2 关联规则分析

采用Apriori算法对构建的样本库进行分析,最小支持度对应的 $t_x$ 为10,最小置信度为0.75,结合图3的流程进行计算,获得的关联规则挖掘结果如表2所示。

表2 关联规则挖掘结果

| 气缸编号 | 规则序号 | 规则前项 X                           | 规则后项 Y    | 置信度  |
|------|------|----------------------------------|-----------|------|
| 1    | 1    | 排气门开启位置偏大,排气门关闭位置偏小              | 排气门开启时间偏小 | 1.00 |
| 2    | 1    | 进气门开启位置偏大,进气真空度最大值偏小             | 进气门开启时间偏小 | 0.82 |
|      | 2    | 排气门开启位置偏大,排气门关闭位置偏小              | 排气门开启时间偏小 | 1.00 |
| 3    | 1    | 进气门开启前泄漏值偏小,排气门开启时间偏小,进气真空度最大值偏小 | 排气门开启位置偏大 | 1.00 |
|      | 2    | 进气门开启前泄漏值偏小,排气门开启位置偏大,进气真空度最大值偏小 | 排气门开启时间偏小 | 1.00 |
| 4    | 1    | 进气门开启位置偏小,进气门开启时间偏小,进气真空度最大值偏小   | 进气门开启位置偏大 | 0.82 |
|      | 2    | 排气门开启位置偏大,排气门关闭位置偏小,进气真空度最大值偏小   | 排气门开启时间偏小 | 1.00 |
| 5    | 1    | 排气门开启位置偏大,排气门关闭位置偏小              | 排气门开启时间偏小 | 1.00 |
|      | 2    | 进气门开启位置偏小,排气门开启时间偏小              | 排气门关闭位置偏小 | 0.91 |
| 6    | 1    | 进气门开启时间偏小,排气门关闭位置偏大,排气门开启时间偏大    | 进气门开启位置偏大 | 1.00 |

以第1缸和第2缸为例,结合 $X \Rightarrow Y$ 进行分析。对于第1缸,排气门开启时间偏小可能是由于配气机构装配误差,或者气门座圈与气门杆存在缺陷,因此需要排查配气机构以及气门座圈密封性。对于第2缸,规则1指出进气门开启时间偏小,可能存在进气量不足,需要排查气门间隙;规则2指出排气门启闭角度错位(延迟开启、提前关闭)导致开启时间偏小,可能存在气门升程不足的问题,需要排查凸轮升程曲线;综合分析,第2缸需重点排查凸轮升程曲线及气门间隙。

## 4 结论

- 1)使用KIH-means算法对某柴油机装配冷试测试出厂数据进行聚类分析,将数据转化为布尔数据库,构建可供挖掘的关联规则样本库。
- 2)通过支持度-置信度试验,确定最小支持度为0.225%、最小置信度为0.75为最优阈值组合。
- 3)采用Apriori算法对样本库进行关联规则处理和分析,得到了针对该柴油机装配异常的关联规则挖掘结果,根据这些规则可采取针对性的改进措施,以提高发动机装配的质量和性能一致性。

### 参考文献:

- [1] OKHIONKPAMWONYI S, LI G X, XU Y D, et al. Comparative study of vibration and acoustic emission strategies applied on monitoring of diesel engine fault [C]//Proceedings of IncoME-V & CEPE Net-2020. Cham, Switzerland: Springer Nature, 2021: 310-319.
- [2] GUARDIOLA C, BARON P, SALA S, et al. Engine combustion hardware diagnostics in an end-of-line cold test stand [C]//Proceedings of WCX SAE World Congress Experience. Detroit, USA;SAE International, 2022.
- [3] 康娜,姜皓宇,蔡志鹏,等. 发动机冷试噪声与振动问题分析与解决[J]. 北京汽车,2025(1):39-41.
- [4] 孙俊伟. 基于大数据分析的柴油机冷试测试分析研究[D]. 济南:山东大学,2021.

- [5] HERNÁNDEZ-NUÑEZ M, BONILLA-BLANCAS A E. Inferential statistics models to relate the rejections of an engine cold testing and the machining defects in camshaft assembly bores[J]. *Ingeniería y Competitividad*, 2022, 24(2): 1–13.
- [6] DONG F, ZHANG H, DU Z G, et al. Comparative analysis of estimation methods of control limits for the engine cold test parameters[C]//*Proceedings of China SAE Congress 2021: Selected Papers*. Singapore: Springer Nature, 2022: 892–901.
- [7] 罗云. 基于冷试的发动机建压故障分析与优化[J]. *南方农机*, 2024, 55(增刊1): 88–91.
- [8] 杨明川, 陈鑫, 李志遥, 等. 数字化控制提升发动机冷试台架柔性[J]. *现代制造技术与装备*, 2023, 59(8): 144–147.
- [9] NING W H, LIU J H, XIONG H. Knowledge discovery using an enhanced latent Dirichlet allocation-based clustering method for solving on-site assembly problems[J]. *Robotics and Computer-Integrated Manufacturing*, 2022, 73: 102246.
- [10] DOU X, XU J L, WU J J, et al. Research on the analysis method of engine assembly quality based on data driven[J]. *Journal of Physics: Conference Series*, 2021, 1948(1): 012171.
- [11] 王海涛, 翟仕奇, 胡嘉旭, 等. 基于关联规则的发动机维修知识图谱构建[J]. *装备制造技术*, 2023(11): 5–9.
- [12] ZHUANG C B, LIU Z W, LIU J H, et al. Digital twin-based quality management method for the assembly process of aerospace products with the grey-Markov model and apriori algorithm[J]. *Chinese Journal of Mechanical Engineering*, 2022, 35(1): 105.
- [13] 胡友鑫, 林茂彦, 罗翦秋, 等. 复杂网络高阶结构的关联规则挖掘及其应用[J]. *电子科技大学学报*, 2025, 54(1): 152–160.
- [14] 祝昶博. 基于改进的K-均值聚类算法的汽车用户画像研究与应用[D]. 长春: 吉林大学, 2022.
- [15] 谢爽. 基于核函数的抑制式可能性聚类算法研究[D]. 西安: 西安邮电大学, 2024.
- [16] LIANG Y, CHEN Y J, HUANG Q, et al. An effective optimization method for fuzzy K-means with entropy regularization [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(7): 2846–2861.
- [17] TATTERSALL R J, BAGGALEY A W, BILLAM T P. Non-equilibrium dynamics of vortices in two-dimensional quantum gases: determining the dynamical scaling region using the mahalanobis distance[J]. *Journal of Low Temperature Physics*, 2024, 215(5): 416–429.
- [18] WANG C M, FU A Q, LI W D, et al. Intelligent identification of hidden dangers in hydrogen pipeline transmission station using GWO-optimized apriori algorithm[J]. *Energies*, 2024, 17(18): 4539.

## Association rules based on analysis of anomalies in a diesel engine assembly cold test

ZHAO Xuhui<sup>1</sup>, XU Zhuo<sup>1</sup>, WANG Hui<sup>2</sup>, YAN Wei<sup>1\*</sup>, LI Guoxiang<sup>1</sup>

1. School of Nuclear Science and Energy Power Engineering, Shandong University, Jinan 250061, China;

2. Weichai Holding Group Co., Ltd., Weifang 261001, China

**Abstract:** To analyze the correlations among abnormal data in the cold test of a certain diesel engine assembly, an abnormal data analysis method for assembly cold test based on association rules is proposed. This method adopts the KIH-means clustering method to construct an association rule sample library for the diesel engine assembly data. Through the support-confidence test, the optimal combination of minimum support and minimum confidence thresholds is determined. The Apriori algorithm is used for association rule mining to reveal the correlation among abnormal data, and the mining results for the diesel engine assembly anomalies are obtained. The results show that the optimal threshold combination for the diesel engine is a minimum confidence of 0.75 and a minimum support of 0.225%, at which the change in the number of association rules is the most stables. The combination of KIH-means clustering method and Apriori algorithm can determine the association rules of abnormal parameters through data mining of the association rule sample library. On this basis, targeted optimization measures can be effectively proposed, thereby improving the engine assembly quality and the consistency of assembly performance.

**Keywords:** diesel engine; cold test; anomaly detection; association rule

(责任编辑:胡晓燕)