

# 基于K均值算法的设备故障分析与应对策略

李思琦<sup>1</sup>, 关博<sup>2\*</sup>

<sup>1</sup>波士顿大学, 文理学院, 美国 波士顿

<sup>2</sup>上海地铁第一运营有限公司, 上海

收稿日期: 2023年11月11日; 录用日期: 2023年12月4日; 发布日期: 2023年12月15日

## 摘要

针对如何通过维护保障管理手段提高设施设备安全运行的问题, 本文以地铁车辆设备为研究对象, 提出使用K均值聚类法对2014~2022年之间的故障数据进行对比分析; 其次, 利用调查问卷的方式, 分别对维保专业内的9位检修组长和18位检修员的日常维护保障行为, 进行数据采集; 最后, 基于应用K均值聚类算法以及调查问卷结果, 提出基于数理统计耦合心理分析的设施设备维保管理应对策略。

## 关键词

地铁设施设备, 数据分析, K均值聚类, 应对策略

# Equipment Fault Analysis and Response Strategies Based on K-Means Algorithm

Siqi Li<sup>1</sup>, Bo Guan<sup>2\*</sup>

<sup>1</sup>College of Arts and Sciences, Boston University, Boston, USA

<sup>2</sup>Shanghai No.1 Metro Operation Co., Ltd., Shanghai

Received: Nov. 11<sup>th</sup>, 2023; accepted: Dec. 4<sup>th</sup>, 2023; published: Dec. 15<sup>th</sup>, 2023

## Abstract

Regarding the issue of how to improve the safe operation of facilities and equipment through maintenance and management measures, this article takes subway vehicle equipment as the research object and proposes to use K-means clustering method to compare and analyze fault data between 2014 and 2022; secondly, using a survey questionnaire, data was collected on the daily maintenance and support behaviors of 9 maintenance team leaders and 18 maintenance personnel in the maintenance profession; finally, based on the K-means clustering algorithm and survey questionnaire results, a facility and equipment maintenance response strategy based on mathematical statistics coupled psychological analysis is proposed.

\*通讯作者。

## Keywords

Subway Facilities and Equipment, Data Analysis, K-Means Clustering, Coping Strategies

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

## 1. 引言

城市轨道交通作为大运量、高效率的运载实体, 支撑着区域经济和社会生产关系的快速发展, 并为了满足日益增加的线路可运载量的需求, 其移动机械整体内部各交叉系统日渐趋于复杂化[1], 必然导致了后期维保压力, 因此如何在现有作业规程约束下, 优化技术人员工作效率, 成为了轨交领域中关注重点[2] [3]。

例如, 上海地铁内人员针对列车运营过程中的信号系统故障问题, 通过内外因素结合测试研究, 利用多维视角对关联因素耦合进行分析, 提出了相应设备的整改解决措施, 有效降低了故障发生率[4]; 广州地铁内部人员针对供电系统中的逆变设备故障进行了系统性的梳理分析, 并利用历史积累数据反向推导出弓网不良关系的底层诱因, 为后续同类故障查找提供了有效依据[5]; 以及成都地铁内部员工针对 5 号线列车调试期间的辅助逆变器 MVB 设备故障问题, 通过其设备运行波形的分析, 找到其故障原因并提出了设备处置的方法[6], 其上述作者均在不同设施设备上取得了不同技术成效。但随着轨道交通的发展不断要求从业人员, 提出创新方法提高维保效率。因此横向观察其他领域, 有学者[7]针对交通影响事故因素, 利用 K 均值聚类法对影响因子进行了研究, 分析主要诱因从而减少重大事故的发生; 并且学者[8]针对既有车辆失稳现象, 提出了利用 K 均值聚类法对历史数据进行挖掘剖析其数据集聚类质心位置, 设计了车辆横向稳定性的判定标准。而值得注意的是, 地铁设施设备的管理本身是属于企业关的范畴, 从管理者的角度观察[9], 需要将注意力逐步从“物的管理”平衡到“人的管理”。因此员工作业时的心理分析是其主体单位不可忽视的内容, 而且良好的员工心理状态, 更有利于企业的管理与发展[10]。

因此本文针对地铁线路 2014~2022 年之间的车辆故障数据, 进行了基于 K 均值聚类法结合维保工作人员问卷调查的轨道车辆故障分析, 并提出了与之对应的维保策略。

## 2. K 均值聚类法

K 均值聚类由 MacQueen 教授提出, 该算法在进行大数据集处理时伸缩性强、易于实现对球型聚类的发现, 所以在运算方面该算法具有较高的效率, 并在多领域内得到了广泛应用[11] [12]。对于 K 均值的细分是属于分区聚类结构, 以将数据划分为若干类为目标, 并且任意一个数据点只有一个专属类别, 这就意味着 K 均值聚类算法生成的是特定数量、互不相交、非层次的聚类。K 均值算法通过不断进行迭代, 利用最小化平方误差和准则来筛选出目标数据集的最佳类别, 因此可大致分为两步, 第一步是寻找与聚类相关的期望点, 第二步是基于第一步的结果对二次划分进行计算, 因此重复这两个步骤直至数据收敛。

K 均值算法中, 数据集  $X$  各样本点之间采用欧氏距离来计算相异性[13]。设数据集有  $n$  个样本, 它的  $p$  维观测为:

$$X = x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})^T, \quad i = 1, 2, 3, \dots, n \quad (1)$$

任意两个样本点  $x_i, x_j$  ( $i, j=1, 2, 3, \dots, n$ ) 之间的欧式距离表示为  $d_{ij} = d(x_i, x_j)$ 。如果将  $n$  各样本分成  $k$  个聚类, 则选择计算结果中距离最大的两个样本点(序号为  $i_1, i_2$ ) 样本作为初始聚类的中心(聚点):

$$d(x_{i_1}, x_{i_2}) = d_{i_1, i_2} = \max \{d_{ij}\} \tag{2}$$

然后再计算下一个聚点(序号  $i_3$ ), 使得  $i_3$  与  $i_1, i_2$  距离最小值等于其他样本点与  $i_1, i_2$  较小距离中的最大者:

$$\min \{d(x_{i_3}, x_{i_r}), r=1, 2\} = \max \left\{ \min [d(x_j, x_{i_r}), r=1, 2], j \neq i_1, i_2 \right\} \tag{3}$$

不断循环上述过程, 即可得出  $k$  个初始聚点。

### 3. 实验分析

#### 3.1. 车辆例年故障分析

因列车行进过程环境复杂、外界对设备影响突出, 因此检修工作压力较大, 各项工作指标要求较高 [14] [15]。因此为了更好地对设施设备的维护保障进行决策, 特以某线路中车辆专业在 2014~2022 共计 9 年的时间, 相关的故障数据进行分析, 并以每年的故障数据单独列为一个数据集  $X_N, N=1, 2, 3, \dots, 9$ 。

特此以每年故障数据为一簇, 应设  $k=1$ , 对其各数据集的  $x_i^N, x_j^N$  进行  $d_{ij}^N = d(x_i^N, x_j^N)$  分析, 求出各簇的聚点位置, 分析结果如图 1 所示。

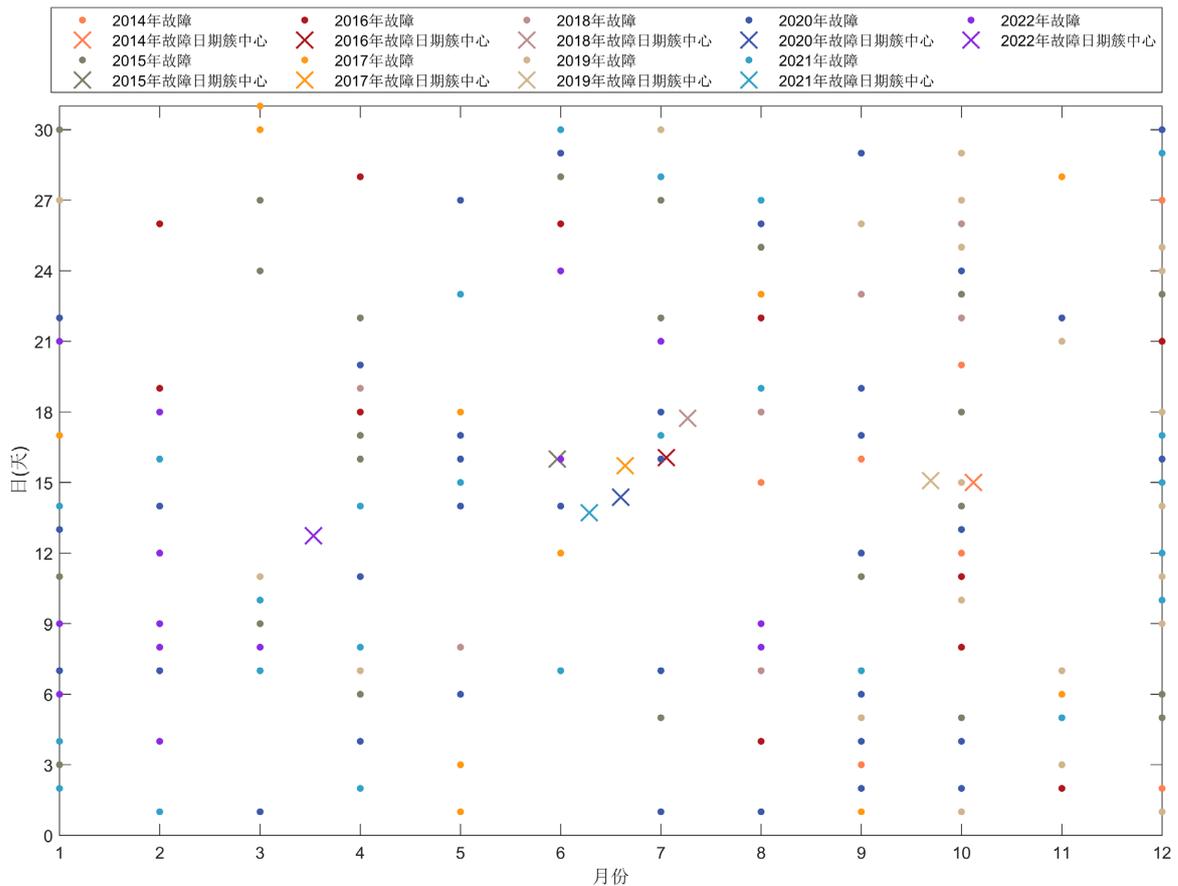


Figure 1. K-means clustering results  
图 1. K 均值聚类结果

从图 1 全局可以看出, 每年故障数据的簇中心主要集中在 6 月份, 并且每年数据的簇中心分布在 6 月后期, 因此可以看出下半年的故障数多于上半年的故障数。并且从工况环境角度看, 从 6 月开始全国的环境逐步转变为不利于设施设备运行的湿热状态, 同时会导致从业人员更易产生疲惫、乏力等不良反应, 及会存在较高的危险系数, 因此使我国将其设为“安全生产月”。企业加大了安全生产各项管理监督措施, 所以从局部具体故障数所反映出的信息, 发现 6 月故障数相对少于 4 月和 10 月的故障数。并且可以看出 3 月和 11 月作为年初和年尾重点安全排查月, 故障数依旧显现出较少的现象。因此可以说明在这些阶段中的安全相关的管理措施是具有成效的, 并且可以进一步说明, 强化设施设备的安全管理, 更能保障地铁的安全运营, 但对于企业管理又是需要考虑多维特别是员工心理抗压因素的复杂过程, 如果持续施加压力反而会发生南辕北辙的后果。

### 3.2. 问卷调查

调研了解员工心理状态, 对工作中突出的主要问题以及改进方法进行探讨及交流, 会更加有利于管理措施的制定以及后期的各项工作推进, 为现有的管理方法进行优化提供重要参考依据, 有利于企业经营管理的长治久安[16] [17]。本文通过“职工代表”会谈问询的方法分别调研了 9 位检修组长和 18 位检修员, 以及使用“调查问卷”的方式制定了与其日常工作相关度较高的 20 个典型工作行为, 进行样表数据采集, 结果如图 2 所示。

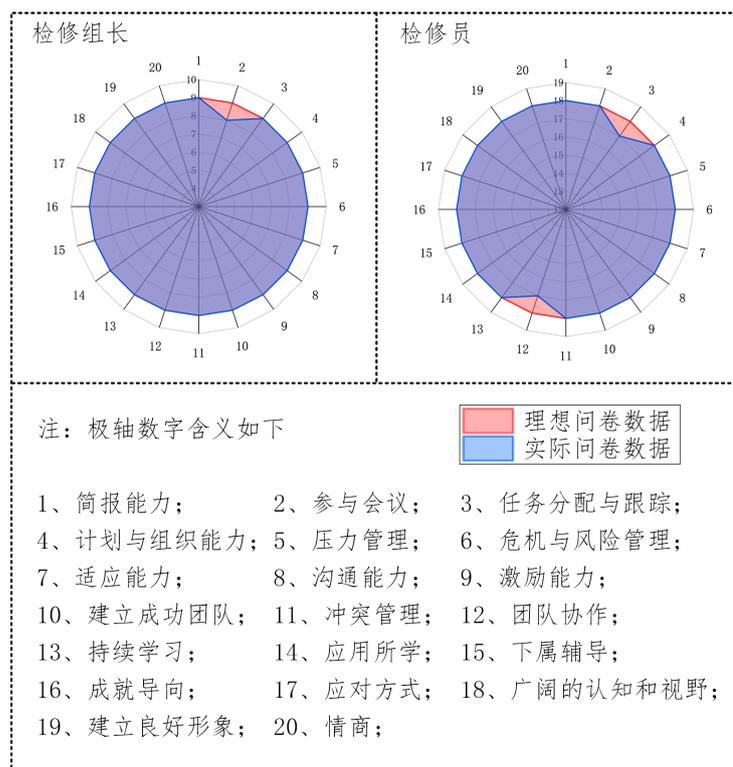


Figure 2. Questionnaire data of line maintenance discipline

图 2. 线路维保专业问卷数据

从图 2 可以看出, 粉色底框为理想问卷数据, 其代表极轴所示的相应人数认为该项工作行为“重要”, 蓝色代表实际问卷数据。例如在检修组长的调查问卷结果中, 可以看到其在“参加会议”选项中, 9 位检修组长的红蓝双色雷达图出现缺口, 也就是代表着此处实际只有 8 位认同“参加会议”这种行

为重要, 有 1 位检修组长不选的原因, 结合企业工作组织结构, 应是班组层级主要负责的是执行步骤, 而具体到生产任务的布置环节主要是相应生产管理部门负责制定, 所以图中调查结果与企业制度吻合。而在图中对检修员所反馈的结果中, 可以看出在“任务分配与跟踪”以及“团队协作”两项指标出现微小残差, 则表示这两处的指标有受调研者认为没有达到重要的程度。但从检修组长的调查中可以看出“任务分配与跟踪”“沟通管理”“冲突管理”等与团队建设相关的内容, 则显示全部班组长认同该选项非常重要。因此可以看出检修组长在基层日常工作中起到了团结和协调的主要作用。因此基于此项问卷调查和理念故障分析结果, 可为后期工作的应对策略提供更为可靠的支撑。

### 3.3. 应对策略

针对地铁实际车辆故障的统计分析, 耦合职员调查问卷的反馈结果, 提出以下 10 条应对策略, 分别是:

1) 细分参与会议的类型, 寻找出各班组长对会议主题的要求, 统一流程规定发言时间、顺序以及格式。并在后期有针对性的去积极引导各班组长的工作, 开设优秀班组长之间的交流会, 加设班组长在“管理”上的培训, 并且平衡好会议与实际工作的总体占比。

2) 针对如何提升团队沟通建设, 建议增设团建活动或体育团队竞技, 利用公司各社团联合各班组, 以班组为单位进行体育或文艺类比赛, 并对其进行宣传及物质奖励, 以此不光提升班组凝聚力, 并且还可以丰富员工们的生活。

3) 增设班组长的“反馈能力”培训。在日常工作中, 对于班组的建设, 上级部门需要班组对目前工作和状况进行有效反馈。因此, 较好地完成对目前工作情况的反馈, 不但是班组建设的重要一环, 而且可以使其在上级部门中更能得到重视, 有利于班组长的对其班组的的管理。并且在一个相较于同车间下平行班组之间, 成绩良好的班组更能提高集体的荣誉感, 能进一步加强班组长的号召能力, 进而可以设置“标兵班集体”流动锦旗等。

4) 要确保管理和维保策略的成功实施, 深刻理解员工和管理者的心理预期和需求至关重要。我们发现, 通过从个体心理学的角度考虑, 对员工和管理者的需求、预期和反应的深入理解能够在推行新的策略时提供坚实的支持。加强沟通并在必要时提供额外培训将确保每个人都能够理解和接受新的策略。同时, 考虑组织心理学的因素——例如, 在特定时期, 如 6 月和 3、11 月, 我们注意到由于某些管理和环境因素的变化, 故障率明显降低。这表明, 我们在制定和实施新政策时, 必须考虑这些影响团体行为和绩效的因素。

5) 通过系统的安全文化培训, 我们可以在组织内部塑造一种广泛接受的安全观念和价值观。当大多数员工认同安全的重要性时, 强大的社会影响力将引导整个团队倾向于严格遵守安全规定, 从而提高整体安全性。定期的安全培训能够刷新员工的知识和技能, 强化安全操作的正确方法, 并提高他们对安全问题的认识和责任感。

6) 促进跨部门协作能够消除信息孤岛, 提高组织的综合效率。加强部门间的互相理解和信任, 结合定期的技能和知识培训, 能进一步提高团队协作效率, 并建立和谐、高效的工作环境。让员工通过跨部门协作拓宽视野和建立更广泛的社交网络, 也将进一步提高他们的工作效率和满意度。

7) 促进有效沟通和建立反馈机制, 有助于增强个体和组织绩效。提高员工的反馈能力不仅可以增强他们的工作理解, 而且也优化与同事和上级之间的沟通。开放和支持性的组织文化能够促进良好的沟通和反馈, 而积极的团队建设活动及奖励机制则可以强化这种文化。这也将支持我们在新政策实施过程中快速识别和解决问题, 确保其成功并不断进行改进。

8) 持续的反馈循环是团队不断优化和进步的关键所在。能够迅速而准确地了解并评估自己的表现和进程, 为团队提供了实施必要改进措施的信息基础。这种透明和开放的沟通机制不仅能增强团队的凝聚

力和改进动力, 也能通过明确、具体的反馈帮助员工了解自己的表现和改进的方向。这种自我发展和实现感将提高员工的工作满意度和投入度。

9) 构建和优化决策支持系统, 确保组织决策的精确性和前瞻性。通过充分利用数据分析和预测工具, 在决策过程中增加数据支持和智能分析, 以增强决策的准确性和前瞻性。运用数据分析工具和技术进行精准的故障预测和预防性维护, 以最小化突发故障的发生。并随技术发展, 不断整合更新的数据分析工具和技术, 适应不断变化的数据环境和分析需求, 确保组织在数据分析和预测方面保持领先和精确。

10) 强化领导层的示范和决策执行, 突出领导力在推行新策略中的决定性角色。认识到领导层对政策和流程的深刻理解及其坚定执行的重要性, 他们的行为和态度将塑造下属的心理预期和行为模式。通过领导的正面行为(如积极参与、展现学习热情及强调团队价值)可以铸造一个积极的工作环境, 激发员工的自我效能感和归属感。最终将倾向于形成一种强大的文化导向和社会影响力, 进一步塑造员工的行为模式和工作心态。

#### 4. 总结

本文使用基于 K 均值算法针对设施设备管理领域中的故障, 进行了关键时间节点及阶段的分析, 利用 K 均值算法得到了近些年中故障数据全局分布特征。从所得出的结果来看, 故障分布具有一定的规律性排布, 并结合往年管理手段, 可以在管理介入密集月份中, 可以看出故障数据量明显下降。

但从个人承受能力上一味增强监管力度, 反而不会得到较好的结果。因此需要从心理学角度对其现象进行分析, 并针对该数据结构特征, 提出了基于社会心理学的构建出的企业内部设施设备故障数据分析方法, 以及与之在全局组织层面的应对策略, 可为后期的企业管理以及生产装备的维护保障管理提供了较为合理的参考。

#### 参考文献

- [1] Li, Y.F. Ma, Y.S. and Feng, S.A. (2011) Application of Embedded System in the Train Information Management System. *Applied Mechanics and Materials*, **44**, 115-119.
- [2] 郭文丽. 基于列控数据完备性的列车故障预警方法研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2017.
- [3] 中国铁路总公司. 列车运行监控装置(LKJ)运用维护规则[M]. 北京: 铁道出版社, 2014.
- [4] 陶文彬. 上海地铁 11 号线列车运行模式丢失故障分析[J]. 交通与运输, 2021, 37(1): 43-46.
- [5] 王志远. 广州地铁二号线整列车辅逆故障分析[J]. 机电工程技术, 2020, 49(5): 210-213.
- [6] 龚承启, 刘瑶. 成都地铁 5 号线列车 MVB 通道故障分析[J]. 电力机车与城轨车辆, 2020, 43(04):91-94.
- [7] 张庆年, 张璠, 杨杰, 等. 基于聚类分析的水上交通事故影响因素研究[J]. 重庆交通大学学报(自然科学版), 2022, 41(6): 1-7.
- [8] 刘宏飞, 徐强, 许洪国, 等. 基于 K 均值聚类分析的车辆横向稳定性判定方法[J]. 湖南大学学报(自然科学版), 2018, 45(8): 48-53.
- [9] 王猛. 有色金属工业企业管理对员工心理健康的影响分析——评《现代冶金企业管理》[J]. 有色金属工程, 2021, 11(6): 150-151.
- [10] 霍静. 国有企业改制期员工心理管理研究[D]: [硕士学位论文]. 南京: 南京航空航天大学, 2006.
- [11] 孙一然. 基于改进 K 均值聚类算法的风电机组齿轮箱故障诊断研究[D]: [硕士学位论文]. 北京: 华北电力大学, 2022. <https://doi.org/10.27139/d.cnki.ghbdu.2021.000590>
- [12] 朱俚治. 一种寻找最优解的新算法[J]. 计算机与数字工程, 2015, 43(11): 1951-1954.
- [13] 何选森, 何帆, 徐丽, 等. K-Means 算法最优聚类数量的确定[J]. 电子科技大学学报, 2022, 51(6): 904-912.
- [14] 王靖铭, 宁静, 赵飞, 等. 基于 Kriging 模型的高速列车齿轮箱箱体裂纹识别[J]. 计算机与数字工程, 2022, 50(4): 892-897.
- [15] 张玮东. 基于 LSTM-BPNN-SVR 的地铁车辆轴箱温度预测方法[J]. 计算机与数字工程, 2022, 50(9): 2079-2082+2108.

- [16] 王芳, 赵小明. 基于多模态社交情感分类的高校学生心理健康分析[J]. 计算机与数字工程, 2022, 50(10): 2166-2170.
- [17] 刘慧娟, 孙红兵. 基于受限审查回归模型的西南地区 95 后大学生学习成绩影响因素实证分析[J]. 计算机与数字工程, 2017, 45(9): 1687-1692.