

# 基于大数据聚类分析的动车组质量精准 管控研究

## 一、引言

### 1.1 研究背景

随着我国高速铁路的飞速发展，动车组作为高铁的核心装备，在铁路运输中占据着举足轻重的地位。近年来，中国动车组行业呈现出强劲的增长态势。2023年，我国动车组产量累计值达到1240辆，期末总额比上年累计增长63.2%，截至2022年12月底，全国铁路客车拥有量为7.7万辆，其中动车组4194标准组，共计33554辆，较2020年增长7.06%。动车组以其高速、高效、安全、舒适等特点，不仅极大地满足了人们日益增长的出行需求，还对区域经济发展和城市化进程起到了强大的推动作用，成为国家重要交通基础设施的关键组成部分。

然而，在动车组数量不断增加、运行里程持续增长的同时，其质量问题也日益凸显。动车组零部件质量的优劣直接关乎动车组的运维效率、运营成本以及运行安全。例如，制动系统故障可能导致列车制动失灵，严重危及行车安全；电气系统问题则可能引发列车晚点、停运等情况，给旅客出行带来不便，也给铁路运营部门造成经济损失。近年来国内动车组多次出现质量问题，在追溯质量问题根源的时候，我国目前还是采用人工经验排查的方式，这种方式考验相关人员的素质与经验，并且缺少大数据研究作为技术支持，过于主观。

在当前大数据时代背景下，动车组在运行、维护等过程中产生了海量的数据，这些数据涵盖了动车组的运行状态、零部件性能、维修记录等丰富信息，是解决动车组质量问题的宝贵资源。聚类分析作为一种重要的数据挖掘技术，能够将海量的动车组质量数据按照相似性进行划分，挖掘出数据背后隐藏的模式和规律，帮助我们从纷繁复杂的数据中发现质量问题的潜在因素，实现质量问题的精准定位和有效预测。通过聚类分析，可以将具有相似质量特征的数据归为一类，进而分析每一类数据所对应的质量问题，找出问题的根源所在；还能根据历史数据的聚类结果，预测未来可能出现的质量问题，提前采取措施进行防范，实现从被动维修向主动维护的转变。因此，开展基于聚类分析的动车组质量数据研究具有重要的现实意义和应用价值，有助于提升动车组的质量管控水平，保障铁路运输的安全、高效运行。

## 1.2 研究目的

本研究旨在运用聚类分析技术，深入挖掘动车组质量数据背后的潜在模式和规律，实现对动车组质量问题的精准定位、有效预测以及优化管理，从而提升动车组的质量管控水平，保障铁路运输的安全、高效运行，具体目标如下：

- **实现质量问题精准定位**：通过对动车组运行、维护等过程中产生的海量数据进行聚类分析，将具有相似特征的数据归为一类，从而找出与各类质量问题密切相关的数据特征组合。以此为依据，在出现质量问题时，能够迅速锁定可能的故障源和影响因素，实现质量问题的精准定位。例如，当某类聚类结果频繁出现与制动系统相关的数据特征异常时，即可高度怀疑制动系统存在质量隐患，进而对制动系统的相关部件进行重点检查和维护，避免因质量问题导致的安全事故和运营延误。
- **有效预测质量问题发生**：基于历史质量数据的聚类结果，构建质量问题预测模型。利用该模型对实时监测数据进行分析，预测未来可能出现的质量问题，提前采取相应的防范措施，实现从被动维修向主动维护的转变。比如，根据以往的聚类分析发现，在特定的运行条件下，某些零部件的数据特征变化趋势与后续的故障发生存在关联。通过建立预测模型，当实时数据显示这些零部件出现类似的数据特征变化时，系统能够及时发出预警，提示维修人员提前准备维修方案和备件，降低故障发生的概率和影响程度。
- **优化动车组质量管理**：依据聚类分析所揭示的质量问题规律和影响因素，为动车组的设计改进、生产制造、运维管理等环节提供科学决策依据，优化质量管理流程和策略。在设计阶段，根据聚类分析中发现的常见质量问题，对动车组的结构和零部件进行优化设计，提

高其可靠性和稳定性；在生产制造环节，加强对关键质量特征的监控和管理，确保产品质量符合标准；在运维管理方面，制定更加合理的维修计划和维护策略，提高运维效率，降低运营成本。通过这些措施，实现对动车组全生命周期的质量优化管理，提升动车组的整体质量水平和市场竞争力。

### **1.3 研究意义**

本研究聚焦于动车组质量数据的聚类分析，无论是在理论层面还是实践领域，都具有不可忽视的重要意义。

在理论层面，本研究丰富和拓展了数据挖掘技术在交通运输领域的应用。通过对动车组质量数据这一复杂且具有行业特色的数据集合进行深入分析，为聚类分析算法在实际场景中的应用提供了新的案例和研究视角。动车组质量数据涵盖了运行状态、零部件性能、维修记录等多方面信息，其数据类型多样、维度高且数据量大的特点，对传统聚类分析算法提出了挑战。在研究过程中，针对这些问题对算法进行改进和优化，推动了聚类分析算法的发展，使其能够更好地适应复杂数据环境，为其他领域处理类似数据提供了可借鉴的方法和思路。同时，本研究也促进了大数据技术与交通运输工程学科的交叉融合，为相关学科的理论发展注入了新的活力，有助于形成更加完善的交通运输大数据分析理论体系。

在实践领域，本研究对动车组的运维管理具有直接且重要的应用价值。通过聚类分析实现对动车组质量问题的精准定位，能够帮助运维人员在众多可能的因素中迅速锁定问题根源，避免盲目排查，提高故障诊断效率。例如，当制动系统出现异常时，通过聚类分析能够快速确定与制动系统相关的数据特征组合，进而对制动系统的关键部件进行针对性检查和维修，减少故障排查时间，提高动车组的可用性。在质量问题预测方面，基于聚类分析建立的预测模型可以提前预警潜在的质量问题，使运维部门能够提前制定维修计划，准备维修备件，合理安排维修资源，实现从被动维修向主动维护的转变。这不仅可以降低因突发故障导致的运营延误和安全风险，还能通过优化维修计划降低运维成本，提高动车组的运营效率。从更宏观的角度来看，提升动车组的质量管控水平有助于保障铁路运输的安全、高效运行，增强人们对铁路出行的信心，促进铁路运输行业的可持续发展，对国家经济和社会发展产生积极的推动作用。

## 1.4 国内外研究现状

### 1.4.1 国外研究情况

国外在动车组质量数据管理和聚类分析应用方面起步较早，积累了丰富的经验和先进的技术。以德国、日本等高铁技术发达的国家为例，德国的西门子公司在动车组质量管控中，利用先进的数据采集系统收集列车运行过程中的各类数据，包括速度、加速度、零部件温度、振动等信息。通过对这些数据的聚类分析，实现了对动车组关键部件的状态监测和故障预测。例如，他们运用 **K-Means**

聚类算法对电机运行数据进行分析，将电机的运行状态划分为正常、轻微异常和严重异常三类。当监测到电机数据落入轻微异常聚类时，系统会及时发出预警，提示运维人员进行检查和维修，避免故障进一步发展；当数据落入严重异常聚类时，则立即采取相应的应急措施，保障列车运行安全。这种基于聚类分析的质量管控方式，大大提高了西门子动车组的可靠性和运维效率。

日本的 **JR 东日本** 公司同样重视动车组质量数据的挖掘与应用。他们建立了庞大的动车组质量数据库，涵盖了从车辆制造到运营维护的全生命周期数据。在聚类分析技术的应用上，采用层次聚类算法对车辆的维修数据进行分析，找出不同故障类型之间的关联关系，从而优化维修策略。通过分析发现，某些电气部件的故障与环境温度、湿度等因素存在聚类关系，基于此，**JR 东日本** 公司制定了针对性的维护计划，在特定的环境条件下提前对相关电气部件进行检查和保养，有效降低了故障发生率。此外，**JR 东日本** 公司还利用聚类分析结果指导动车组的设计改进，根据不同线路的运行特点和故障聚类情况，对车辆的悬挂系统、制动系统等进行优化设计，提高了动车组在不同工况下的适应性和可靠性。

在技术层面，国外在数据采集、传输和存储方面拥有先进的硬件设备和成熟的软件系统。高精度的传感器能够实时、准确地采集动车组运行中的各种数据，高速可靠的数据传输网络确保数据能够及时上传至数据中心，高效的数据存储架构则保障了海量数据的安全存储和快速读取。在聚类分析算法的研究和应用上，国外不断探索新的算法和改进现有算法，以适应动车组复杂多变的数据特征和实际应用需求。例如，一些研究机构提出了基于密度峰值的聚类算法，该算法能够在不预先指定聚类数量的情况下，自动识别出数据集中的聚类中心和边界，对于处理动车组质量数据中聚类数量不确定的问题具有很好的效果。

## 1.4.2 国内研究情况

国内在动车组质量数据聚类分析领域的研究近年来取得了显著进展。众多科研机构 and 高校积极开展相关研究，一些铁路运营企业也开始重视并尝试应用聚类分析技术提升动车组质量管控水平。在理论研究方面，国内学者对聚类分析算法在动车组质量数据处理中的应用进行了深入探讨。例如，北京交通大学的研究团队针对动车组质量数据类型多样、维度高且数据量大的特点，对 **Chameleon** 聚类算法进行了改进。通过引入密度因子和距离度量的自适应调整策略，提高了算法在处理复杂数据时的聚类质量和效率。实验结果表明，改进后的算法在动车组质量问题定位和预测方面表现出色，能够更准确地识别出质量异常的数据簇，为后续的故障诊断和维修决策提供有力支持。

在实际应用方面，中国铁路总公司（现中国国家铁路集团有限公司）下属的一些动车段开始尝试运用聚类分析技术对动车组的运行数据和维修数据进行分析。通过对历史故障数据的聚类，找出了一些常见故障的发生规律和影响因素，如通过对制动系统故障数据的聚类分析发现，制动盘磨损程度与列车的运行速度、制动频率以及线路坡度等因素密切相关。基于这些分析结果，动车段制定了更加科学合理的检修计划，针对不同运行条件下的动车组，对制动系统采取差异化的检修策略，有效提高了检修的针对性和效率。此外，一些动车组制造企业也将聚类分析技术应用于产品质量控制中，通过对生产过程中的质量数据进行聚类，及时发现生产线上的潜在质量问题，优化生产工艺，提高产品质量。

然而，与国外先进水平相比，国内在动车组质量数据聚类分析的研究和应用方面仍存在一定差距。在数据采集和传输方面，部分设备的稳定性和可靠性有待提高，数据采集的全面性和准确性也需要进一步加强。在算法研究方面，虽然国内在改进现有算法上取得了一定成果，但在新算法的创新和应用方面相对滞后，缺乏具有自主知识产权的核心算法。在实际应用中，聚类分析技术与动车组运维管理流程的融合还不够紧密，尚未形成完善的基于聚类分析的质量管控体系，导致聚类分析结果在实际决策中的应用效果受到一定限制。

## 1.5 研究方法与创新点

### 1.5.1 研究方法

- **文献研究法**：广泛查阅国内外关于动车组质量管控、数据挖掘、聚类分析等方面的学术文献、技术报告、行业标准等资料。通过对这些文献的梳理和分析，了解相关领域的研究现状、技术发展趋势以及存在的问题，为本研究提供理论基础和研究思路。例如，在研究聚类分析算法在动车组质量数据中的应用时，参考了大量关于聚类算法原理、改进方法以及在其他领域应用的文献，从而确定了适合本研究的算法和改进方向。
- **案例分析法**：选取国内外典型的动车组质量问题案例，对其故障现象、处理过程、原因分析等进行深入剖析。通过实际案例研究，总结质量问题的发生规律和特点，验证聚类分析方法在实际应用中的有效性和可行性。比如，对德国西门子公司和日本 JR 东日本公司利用聚类分析解决动车组质量问题的案例进行详细分析，学习他们在数据采集、算法应用和质量管理方面的成功经验，为我国动车组质量管控提供借鉴。

- **数据挖掘技术**：运用数据挖掘中的聚类分析技术对动车组的海量质量数据进行处理和分析。
  - 聚类分析技术能够在没有先验知识的情况下，将数据集中相似的数据对象归为同一类，从而发现数据中的潜在模式和规律。针对动车组质量数据类型多样、维度高且数据量大的特点，选择合适的聚类算法，并对算法进行改进和优化，以提高聚类分析的准确性和效率。
  - 例如，采用改进后的 **Chameleon** 聚类算法对动车组的运行数据、维修数据等进行聚类，找出不同质量问题所对应的聚类特征，实现质量问题的精准定位和预测。
- **实证研究法**：收集实际运行的动车组质量数据，在真实的数据环境下进行实验和分析。通过对实验结果的评估和验证，检验研究方法和模型的有效性和可靠性。具体来说，利用某动车段提供的动车组历史运行数据和维修数据，进行聚类分析实验，根据实验结果对质量问题进行定位和预测，并与实际发生的质量问题进行对比，验证研究成果的实际应用价值。
  - 同时，根据实证研究中发现的问题，进一步优化研究方法和模型，使其更好地适应实际应用需求。

## 1.5.2 创新点

- **算法改进创新**：针对传统聚类分析算法在处理动车组质量数据时存在的不足，如对数据分布的敏感性、聚类结果的稳定性差以及计算效率低等问题，提出了创新性的改进思路。例如，对 **Chameleon** 聚类算法进行改进，引入自适应的密度度量和距离度量方法，使其能够更好地适应动车组质量数据复杂的数据分布。在密度度量方面，根据数据点的局部邻域信息动态调整密度计算方式，避免了固定密度阈值在不同数据区域适应性差的问题；在距离度量上，结合动车组质量数据的特点，综合考虑数据点的多个属性特征，采用加权距离度量方法，提高了距离计算的准确性，从而提升了聚类质量和稳定性。此外，通过改进算法的合并策略，减少了不必要的计算开销，提高了算法的运行效率，使其能够快速处理海量的动车组质量数据。
- **数据应用创新**

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/628004113001007051>

•