

# 大数据技术在 LKJ 设备分析系统中的应用研究

杨 献, 言 圣

(湖南中车时代通信信号有限公司, 湖南 长沙 410100)

**摘要:** 目前列车运行控制装置(LKJ)设备分析系统存在分析不全面、效率低、故障定位难等问题。文章分析了LKJ设备分析系统未来发展需求和大数据技术的应用情况,提出一种基于大数据技术的LKJ设备分析系统方案,其基于大数据平台,采用连续变量突变、相关系数法、复杂关系网络等分析方法对LKJ设备大量历史数据进行挖掘分析,形成了LKJ设备的异常数据和故障关键因素及支持度。大数据技术的应用为LKJ设备故障及时查找、故障分析和维护提供数据支撑和决策指导。

**关键词:** 大数据; 异常事件; 趋势分析; 故障关联; LKJ设备

中图分类号: U285

文献标识码: A

文章编号: 2096-5427(2020)04-0044-05

doi:10.13889/j.issn.2096-5427.2020.04.009

## Application of Big Data Technology in LKJ Equipment Analysis System

YANG Xian, YAN Sheng

(Hunan CRRC Times Signal & Communication Co., Ltd., Changsha, Hunan 410100, China)

**Abstract:** At present, there are some problems in LKJ equipment analysis system, such as incomplete analysis, low efficiency and difficult fault location. This paper analyzed the future development requirements of LKJ equipment analysis system and the application of big data technology, and proposed a scheme of LKJ equipment analysis system based on big data technology. Based on the big data platform, it uses continuous variable mutation, correlation coefficient method, complex relationship network and other analysis methods to mine and analyze a large number of historical data of LKJ equipments, forming abnormal data and key fault causes of LKJ equipments and support. The application of big data technology can provide data support and decision guidance for problem finding in time, fault analysis and maintenance of LKJ equipments.

**Keywords:** big data; abnormal event; trend analysis; fault association; LKJ equipment

## 0 引言

列车运行控制装置(LKJ)为铁路行车安全保障设备。虽然其记录了丰富的列车运行数据,同时设备的监测、生产、返修、故障记录数据都有相应的系统进行管理,但目前仍存在以下问题:现场LKJ设备文件分析不全面且分析效率低,致使故障查找定位难;设备维护管理采用定期修和故障修两种模式,设备故障易导致事故,且设备故障修影响列车运行效率;既有设备质量分析、生

产管理、LKJ设备运行监测管理(LMD)等系统的数据独立,未关联融合,综合数据的价值有待发掘。

大数据技术随着第四次产业革命而发展,在生产、供、销、创新等环节发挥了重要作用<sup>[1]</sup>,推动企业向精益生产和服务型制造业的快速转型升级;工业大数据和人工智能技术的出现正是着眼于产品全生命周期内服务能力的提升,以全面提升产品和服务品质,实现降本增效,提高企业的核心竞争力。如何利用大数据技术实现LKJ设备的智能分析,提升产品可靠性,提高维修效率,降低维护成本,是当前亟待解决的问题。为此,本文通过研究分析大数据关键技术,提出一种基于大数据技术的LKJ设备分析系统方案及其功能实现方法,并进行了实验验证。

收稿日期: 2019-09-04

作者简介: 杨献(1973—),男,高级工程师,主要从事信号系统地面软件开发工作。

基金项目: 湖南创新型省份建设专项(2019GK4015)

## 1 大数据关键技术

大数据技术是信息化社会的代表技术之一,其应用范围广泛<sup>[2]</sup>。根据大数据的信息处理流程,基于大数据技术的LKJ设备分析系统信息处理流程可分为数据采集、处理、存储、挖掘分析和结果展示5部分(图1),其中数据采集、处理和挖掘分析是关键技术。下面主要介绍数据处理和挖掘分析技术。

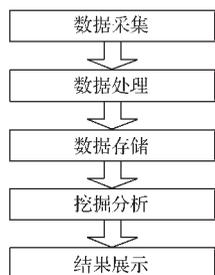


图1 信息处理流程

Fig. 1 Information processing process

### 1.1 数据处理技术

传统的数据处理方法是以前端处理器为中心;而在大数据环境下,需要采取以数据为中心的模式,以减少数据移动带来的开销。大数据处理模式包括流处理和批处理,其中批处理模式采用先数据存储后处理方式,流处理模式采用直接处理方式。根据大数据的数据特征和计算需求,数据处理方法还包括内存计算、图计算和迭代计算等。

大数据的基本处理流程与传统数据处理流程差别不大。由于大数据要处理大量非结构化的数据,所以在各处理环节都采用Map Reduce方式进行并行处理。Map Reduce适合进行数据分析、日志分析、商业智能分析、客户营销、大规模索引等业务。Spark Streaming是构建在Spark上的处理Stream数据的框架,其基本原理是将Stream数据分成小的时间片段(几秒钟),以支持高吞吐量和容错的实时数据流处理。

### 1.2 数据挖掘分析技术

数据挖掘分析是大数据的核心。由于大数据具有海量、复杂多样及变化快等特性<sup>[3]</sup>,需要采用适应大数据特点的分析方法。常用分析方法有数据挖掘方法、概率统计法、相关系数法和机器学习方法。

#### (1) 数据挖掘方法

数据挖掘方法<sup>[4]</sup>主要包括分类分析、关联分析、聚类分析和异常检测等4种。分类分析包括模式识别、决策树、贝叶斯分类及人工神经网络等;关联分析包括Apriori算法、FP增长算法及频繁子图挖掘等方法;聚类分析包括基于原型的聚类、基于密度的聚类以及基于图的聚类等方法;异常检测包括离群点检测等。

#### (2) 概率统计法

概率统计法又称数理统计方法,是研究自然界中随机现象统计规律的数学方法。概率统计法主要研究对象为随机事件、变量和过程,采用概率理论研究大量随机现象的规律性,通过一组样本判定能否以相当大的概率来保证某一判断的正确性,并可以控制出现错误的概率。概率统计中常用的异常检测方法是利用高斯密度函数,计算数据出现的概率,如果概率小于某个阈值的数据,就认为该数据是异常的。

#### (3) 相关系数法

相关系数是最早由统计学家卡尔·皮尔逊设计的统计指标,是研究变量之间线性相关程度的量,一般用字母 $r$ 表示。由于研究对象的不同,相关系数有多种定义方式,较为常用的是皮尔逊相关系数。皮尔逊相关系数是用以反映变量之间相关关系密切程度的统计指标,其按积差方法计算,同样以两变量与各自平均值的离差为基础,通过两个离差相乘来反映两变量之间相关程度,一段着重研究线性的单相关系数。皮尔逊相关系数的定义为

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}}$$

式中:  $\text{Cov}(X, Y)$  ——  $X$  与  $Y$  的协方差;  $\text{Var}[X]$  ——  $X$  的方差;  $\text{Var}[Y]$  ——  $Y$  的方差。

#### (4) 机器学习法

机器学习的本质是使用实例数据或经验训练模型进行分析处理,其包括归纳学习、分析学习、类比学习、遗传算法、联结学习及增强学习等。机器学习是面向任务、基于经验提炼模型以实现最优解设计的计算机程序,其利用经验学习规律,一般应用在缺少理论模型指导但存在经验观测的领域中。

## 2 系统方案

根据上述大数据关键技术分析,本文提出一套基于大数据的LKJ设备分析系统方案。该LKJ设备分析系统通过与其他信息系统接口,获取LKJ设备的运行记录、故障记录、现场反馈记录及检修记录等数据,并对这些数据进行清洗融合、抽取转换加载(extraction-transformation loading, ETL)和事件拦截分析;采用聚类分析、贝叶斯网络<sup>[5]</sup>、复杂关系网络、有向概率图等模型算法,通过数值理论分析、数据挖掘等环节,生成大数据模型与算法,建立故障规则和质量项点分析规则,通过趋势、突变、关联及因素分析,生成LKJ设备的状态图或异常趋势图,形成故障或异



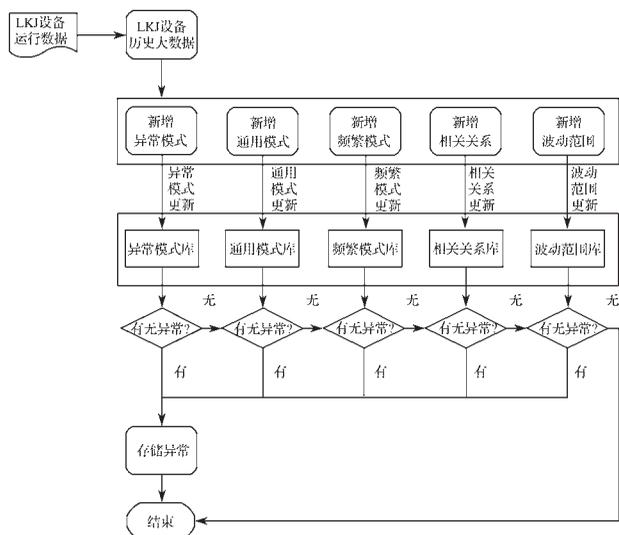


图4 异常拦截流程图  
Fig. 4 Abnormal intercept diagram

### 3.3 设备趋势分析

采用离散序列的偏态分析，总结异常规律，形成故障数据的趋势分析<sup>[9]</sup>，实现设备状态的预警提示。LKJ系统故障/异常趋势分析时，按时间维度分析故障/异常发生的趋势，生成某种故障/异常状态趋势图并总结出相关规律，为LKJ设备的检修形成指导依据。在分析过程中，增加车型（机车型号、机车号）、地域、区间和时间等要素，以分析在特定条件下的趋势规律；同时，针对不同分析维度添加预警线，也能帮助定位到极度异常事件。图5示出异常状态趋势分析结果展示界面。

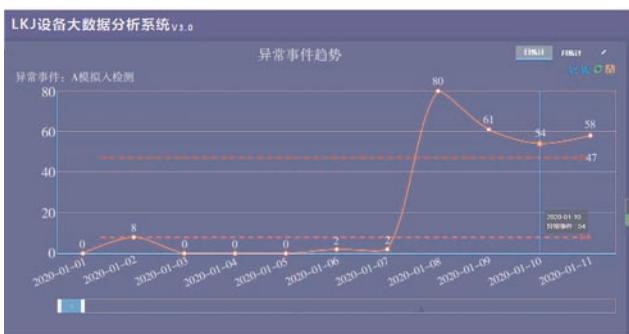


图5 异常状态趋势分析  
Fig. 5 Abnormal state trend analysis

### 3.4 故障关联分析

LKJ设备的运行数据文件、故障反馈单等信息中含有故障检修和故障关联的相关信息。对于故障文本数据，可以通过对其进行数据预处理，筛选出有关故障表征、故障原因的数据，结合专家知识建立自然分词库和人工提词库，并通过机器学习算法和人工筛选分类建立故障表征和故障设备的关联关系；基于故障特征建立故障关联分析，其中提取故障特征是关键的分析环节，其分析结果的好坏直接影响到故障之间关系的映射。针对故障文本特征的提取算法有 Viterbi 算法、N-Gram 算法、

TF-IDF 算法、基于信息熵的新词发现等，通过上述算法实现最初的中文分词，从不同的角度对分词结果进行筛选和优化，分类建立故障表征和故障设备的关联关系；通过参数学习和结构学习，分类建立故障与异常之间关联关系。

采用故障/异常之间关联图，展现故障和异常事件、异常事件和异常事件、故障和故障之间的关联关系，以及关联的精密度，为故障的预测和深度故障的挖掘<sup>[10]</sup>提供数据支撑。图6示出LKJ设备故障关联分析图。

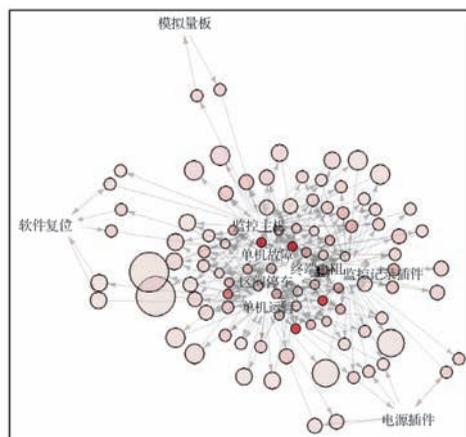


图6 故障关联分析图  
Fig. 6 Fault association analysis chart

### 3.5 故障因素分析

故障因素分析采用复杂网络模型实现。复杂网络模型主要研究事物的关联性，基于异常变量的关系，探索故障关联性；借鉴复杂网络模型，构建故障关联的复杂网络模型，实现从大量运行记录数据中挖掘分析与故障相关的关键要素。对接入大数据分析平台的LKJ设备文件实时进行挖掘模型分析，找出该文件出现故障或异常的关键要素，以便于故障定位，提高分析效率。故障因素分析主要包括单机运行、显示器黑屏、轮对空转、信号异常（绿灯转白灯、灭灯、双黄灯转红黄灯、红黄灯转红灯、绿灯转红黄灯）等项点，其主要实现的功能如下：

- (1) 显示挖掘分析的模型关键要素及当前运行记录数据中该故障或异常出现的关键要素，进行故障定位分析（图7）；
- (2) 显示该故障或异常的全程运行记录数据；
- (3) 显示挖掘分析结论；
- (4) 显示故障前置事件并进行排序；
- (5) 提供指导意见的编辑功能；
- (6) 显示关键前置事件并进行排序；
- (7) 实现按时间、车型、车号查询出现故障的因素分析。

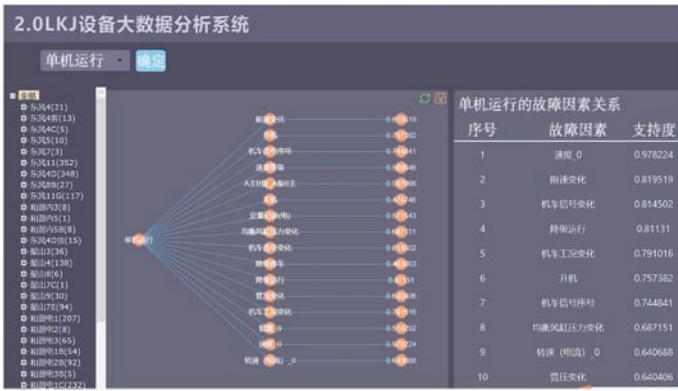


图 7 故障因素分析界面  
Fig. 7 Fault factor analysis interface

### 3.6 统计分析

基于设备故障的经验规则方面的研究，全面综合地分析利用各类 LKJ 设备数据和与设备相关的数据，对大量历史运行记录数据文件进行检索分析并对同一 LKJ 设备状态或故障进行累计，按路局、电务段、时间、线路、故障类型等维度对设备故障或异常进行分析统计，对 LKJ 设备使用里程、时间、按键次数、开关情况进行累计分析，实现设备维修预警和使用寿命管理，并及时提醒检修和设备更换，以减少设备故障。

## 4 实验验证

该系统已在实验室进行部署并应用；通过在实验室搭建大数据分析平台，将大数据挖掘分析算法部署到大数据平台，将近 3 年 LKJ 运行记录数据和故障反馈数据导入大数据分析平台，进行挖掘分析，并对分析结果进行展示。图 8 示出综合分析结果数据的展示界面。



图 8 分析结果数据展示界面  
Fig. 8 Data display of analysis results interface

通过该系统的实验室应用，找出 LKJ 设备运行过程中出现的速度、管压、缸压、转速、电流的突变情况，

发现转速记录异常、速度跳变等问题，并及时进行处理；建立设备故障趋势统计指标，以可视化方式展示各 LKJ 设备状态的历史趋势和设备故障的分布情况，用于针对性地指导维护策略的制定；形成现场各 LKJ 设备的使用时长、使用里程、按键、开关等寿命管理数据，为设备检修和维护提供指导；得到 LKJ 设备的单机运行、显示器黑屏等典型故障的关键因素及支持度，为分析设备故障原因提供数据支撑。

## 5 结语

本文给出基于大数据技术的 LKJ 设备分析系统设计方案及其功能实现。通过部署与试用，所设计的系统可以从大量历史运行记录数据中找出设备运行记录的异常情况，形成设备累积里程和使用时间，生成典型设备故障相关因素汇总、设备状态历史趋势曲线、设备故障异常趋势曲线以及故障指标统计，以辅助维护人员及时发现并处理，从而降低设备故障率，并为设备维护提供数据支撑和决策指导。

随着 LKJ 设备相关数据的不断完善及异常故障数据的不断历史累积，后续将通过深入研究设备故障特征、故障因素及故障规律，挖掘分析模型，不断更新智能化分析技术，以进一步提升设备的故障定位和故障预测的准确率。

## 参考文献:

- [1] 毛文彬. 面向大数据的分布式系统设计关键技术研究 [J]. 无线互联科技, 2014(11): 210.
- [2] 张志坚. 城市轨道交通中的大数据分析 [J]. 信息系统工程, 2017(10): 20-21.
- [3] 王继业, 季知祥, 史梦洁, 等. 智能配用电大数据需求分析与应用研究 [J]. 中国电机工程学报, 2015(8): 1829-1836.
- [4] 王国胤, 刘群, 于洪, 等. 大数据挖掘及应用 [M]. 北京: 清华大学出版社, 2018.
- [5] 李玉兰. 基于贝叶斯网络的列控车载设备故障诊断研究 [D]. 北京: 北京交通大学, 2016.
- [6] 林刚. 基于大数据云计算的铁路智能运维系统技术研究 [J]. 铁道通信信号, 2019(5): 37-41.
- [7] 曾怡. 基于大数据平台的自动化运维及监控技术研究 [J]. 科技创新导报, 2018, 15(25): 7-8.
- [8] 靳继红, 刘淑芝. 大数据在铁路货运信息系统中的应用框架探讨 [J]. 电脑知识与技术, 2015(13): 5-6.
- [9] 黄康, 孙旺. 基于数据挖掘的车载信号设备智能维修方法的实现 [J]. 铁道运输与经济, 2017(4): 73-77.
- [10] 朱超平, 白雪. 基于大数据的电务智能运维平台方案研究 [J]. 铁道通信信号, 2017(4): 74-78.