

文章编号: 1005-8451 (2016) 09-0083-05

# 基于大数据技术的铁路安全管理研究

杨连报, 李平, 徐贵红, 刘俊, 刘彦军

(中国铁道科学研究院 铁路大数据研究与应用创新中心, 北京 100081)

**摘要:** 大数据技术为分析铁路安全管理领域的海量历史数据, 发现其事故规律提供了有效的手段。本文分析了铁路安全管理系统的现状, 并结合大数据技术的特点, 按照“人-设备-环境”3个方面, 给出了铁路安全管理大数据应用的需求分析模型。该模型为实现设备的风险分析、状态维修, 给决策者提供辅助决策, 提高铁路安全预警水平, 有重要指导意义。

**关键词:** 大数据; 安全管理; 状态维修; 辅助决策; 安全预警

**中图分类号:** U29-39      **文献标识码:** A

## Big data based railway safety management research

YANG Lianbao, LI Ping, XU Guihong, LIU Jun, LIU Yanjun

(Research and Application Innovation Center for Big Data Technology in Railway, China Academy of Railway Sciences, Beijing 100081, China)

**Abstract:** Big data provides an effective way for analysis massive historical data in the field of railway safety management to find the regularity of accidents. This article analyzed the present situation of Railway Safety Management System and data, combined with the characteristics of big data, was in accordance with the “people-equipment-environment” three aspects, gave a requirements analysis model for railway safety management of big data applications. This model is an effective way to implement the risk analysis and maintenance of the equipment, provide auxiliary decision-making for decision makers, improve railway safety warning level, which has important guiding significance.

**Key words:** big data; safety management; condition-based maintenance; auxiliary decision-making; safety warning

随着铁路信息化的不断发展, 已相继建成车、机、工、电、辆等各专业安全监测/监控系统近60余个, 产生和存储了PB级的文本、图纸、视频、图像、声音等多种类型的监控/监测数据。铁路安全领域迎来了大数据时代, 由于传统数据库无法对这些海量的数据资源进行有效的利用, 从而使这些数据占据了大量的存储空间而不能被挖掘使用, 造成了数据资源的浪费。伴随着大数据技术的快速发展, 针对海量数据的存储、分析和可视化展示等问题, 出现了新的技术来支撑, 例如: 分布式文件存储、并行处理<sup>[1-2]</sup>等。铁路运输安全是一个复杂的安全系统工程, 其故障(事故)之间存在着相关关联、相互影响的耦合关系。通过运用大数据分析技术不仅可以对海量的历史数据进行学习和挖掘, 从而预测安全风险发现事故规律和事故间的相关关系, 对高实时性要求

的监控数据进行实时在线分析, 提高预警水平。

## 1 铁路安全管理现状

根据我国铁路的监控/监测系统的现状, 给出了铁路安全监控体系的架构框图, 如图1所示。

## 2 铁路安全管理数据现状

铁路安全管理数据集呈现“海量数据+复杂类型数据”的状态, 已具有大数据的4V特征, 即规模性(Volume)、多样性(Variety)、快速性(Velocity)、和价值性(Value)。

### 2.1 规模性

按易安信公司(EMC)的界定, 大是指大型数据集, 一般在10TB规模左右, 多用户把多个数据集放在一起, 形成PB级的数据量。目前, 铁路安全领域已相继建立各专业安全监测/监控系统近60余个, 产生和积累了PB级的数据。比如, 在“5T”系统中, TFDS产生的图像信息是数据量最大的, 一趟

收稿日期: 2016-06-15

基金项目: 中国铁路总公司科技研究开发计划课题(2015X003-F); 中国铁道科学研究院基金重大项目(1551DZ8004)。

作者简介: 杨连报, 在读博士研究生; 李平, 研究员。



图1 铁路安全监控体系构成

车经过，产生大概 700~800 M 的图像数据；高速铁路综合视频监控系统产生的视频数据规模每年约为 142 PB。高速铁路自然灾害及异物侵限系统数据以及视频监控数据仅是高速铁路安全数据集中一部分，可见铁路安全管理领域具有大数据的规模性特征。

## 2.2 多样性

多样性是指数据来自于多种数据源，数据种类和格式冲破了以前所限定的结构化数据范围，囊括了半结构化数据和非结构化数据。铁路安全数据分类如图 2 所示。铁路安全管理监控/监测系统涵盖车、机、工、电、辆等专业系统，并且每个专业都是个庞大而复杂的系统，数据格式包含设备台账、人工检查记录本等纸质类数据，轨检车的 geo 和 iic 格式的数据，道岔铺设图、桥梁布置图、线路设备综合图、车站竣工图等 CAD 图纸数据，以及视频监控录像、图像、音频等数据。此外，铁路安全管理大数据应用过程中还存在着与天气、地震等多类型数据的关联分析。

## 2.3 快速性

大数据的数据挖掘技术与传统的技术不同，在于其处理速度非常快，具体来说数据流往往为高速实时数据流，而且需要快速、持续地实时处理。如高速铁路综合视频监控系统、CTC/TCDS、供电 SCADA 系统、电务微机监测系统、防灾监测系统等，实时

处理着各种监测数据，快速判断现场发生的灾害或事故，及时反馈给铁路运营部门，保障铁路行车安全。

## 2.4 价值性

铁路安全数据集的价值性主要体现为铁路运营部门可利用积累的设备相关历史数据，通过各种统计、算法，对与安全相关的大数据进行分析，从数据中发现关联、发现知识，成为风险控制、安全行车的重要依据。

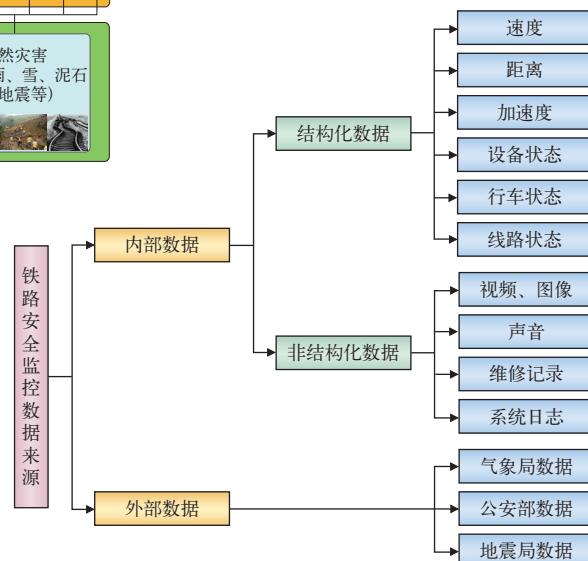


图2 铁路安全数据分类

## 3 铁路安全管理对大数据分析需求

根据目前铁路安全管理数据现状和面临的形式，并结合大数据技术，得出围绕“人-设备-环境”3方面的的铁路安全大数据需求，如图 3 所示。

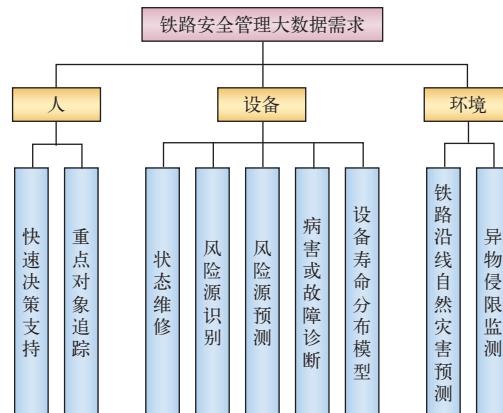


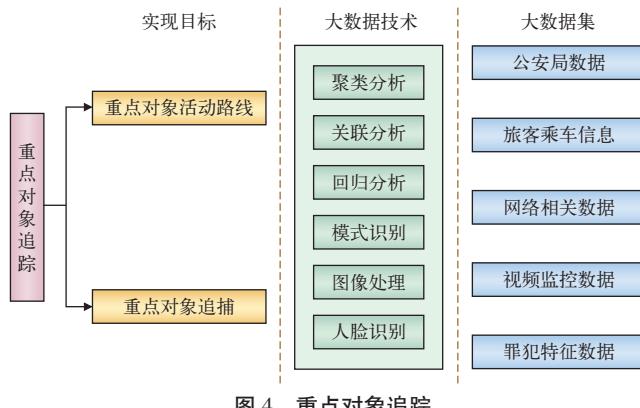
图3 铁路安全管理的大数据需求

### 3.1 铁路人员相关的大数据分析需求

### 3.1.1 重点对象追踪

重点对象追踪以社会安全出发，旨在确保人身和财产安全。重点对象追踪借助大数据技术，并结合人脸识别技术，实现重点对象的活动路线追踪。人脸识别技术是利用图像处理和模式识别技术鉴别或验证身份的一种热门的计算机安全技术，现已在国家安全、军事安全和公共安全等方面实现了初步的应用。铁路车站环境存在人流密度大、治安难度大、环境复杂等特殊的行业应用特点。通过铁路车站人脸识别系统前置摄像头采集到的人脸信息，并与公安局提供的重点对象特征数据进行对比，定位重点对象；同时结合客票系统的乘车信息，挖掘其活动路线，再通过网络（网页、社交软件）数据提取其经常联系的对象，实现重点对象团伙的追捕和打击。

重点对象追踪用的大数据相关技术有聚类分析、关联分析、模式识别等，具体如图 4 所示。



### 3.1.2 铁路实时重大决策支持

铁路实时重大决策支持是针对重大事故时所采取的应急措施，其主要的数据来源为应急救援指挥信息系统提供的历史事故数据、GIS 提供的线路环境数据、以及铁路沿线视频监控数据，经过大数据聚类分析、关联分析等技术发现类似事件的事故规律，供决策者参考，如图 5 所示。

### 3.2 铁路设备相关的数据分析需求

铁路设备全生命周期管理是指对铁路设备基于建设、运营直至报废的全生命周期过程，为确保运输安全及设备处于健康状态而进行的管理活动，为实现在最恰当的时候、利用最恰当的方法、找到解决问题的最恰当资源提供了一种新思路，这种设备管理模式可更有效地统筹铁路建设与运营相互协调

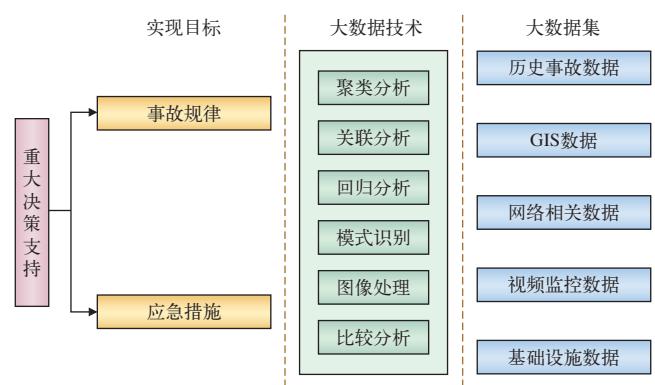


图 5 铁路实时重大决策支持

发展，构建寿命分布模型，实现信息可追溯、病害或故障诊断更科学、维修周期预测更准确，进而最大程度降低运营安全风险和维修成本，如图 6 所示。

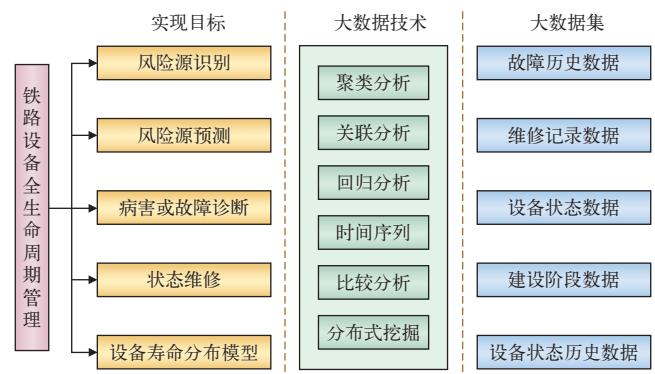


图 6 铁路基础设施健康管理

铁路设备全生命周期管理主要是通过大数据聚类分析、关联分析、时间序列等技术，对设备的历史状态信息进行分析，得出设备的生命周期曲线，然后根据传感器所测得数据，判断设备目前所处的阶段，进而用于故障诊断、故障预测和最佳维修周期的制定。做到防范于未然，规避风险。

### 3.2.1 铁路设备寿命分布模型

在铁路设备的实际使用中，由于受到外界多种因素的影响，设备实际的使用寿命往往与设计寿命、预期使用寿命不同，为保障生产的连续高效进行，以及避免重大的设备故障发生，重点要关注设备的剩余寿命。剩余寿命是指从当前时刻算起，直到设备达到不能使用或者设定的阈值的全部时间。用于剩余寿命预测的方法有自回归移动平均模型、神经网络模型、灰色理论模型、支持向量机模型等。铁路设备较为适用的寿命分布模型有指数分布模型、威

布尔分布模型和伽马分布模型。

通过对铁路设备寿命数据的分析,确定设备的寿命分布模型,掌握设备状态变化规律,获得设备可靠性评价指标,定量地把握铁路设备的性能状态,判断设备状态“好”、“坏”,并可把所获的信息反馈到设计、建造和使用维修中去,以其改善可靠性、科学安排设备维修和更换,保障铁路行车安全。同时精准维修、预防性维修可帮助管理者合理配置维修资源,从而能够大幅度降低设备管理部门的生产成本。

### 3.2.2 铁路设备病害/故障诊断

设备病害或故障产生有深刻的产生或发生机理,有可能是设备本身的特点所决定的,有可能是不同设备之间的相互影响造成的,也有可能是维修不当造成的。设备故障诊断是指在设备运行中或在基本不拆卸情况下,通过各种手段掌握设备运行状态,判定产生故障的部位和原因,并预测、预报设备未来的状态,从而找出对策。故障诊断技术的发展经历了感性阶段、量化阶段、诊断阶段以及人工智能和网络化等阶段,设备故障诊断常用的方法包括故障树分析、数理统计、专家系统、模糊集、人工神经网络、遗传算法、数据挖掘等方法。铁路设备的数据集呈现“海量数据+复杂类型数据”的状态,具有大数据的“4V”特征,可以采用数据挖掘的方法进行设备故障诊断。

数据挖掘的目的是从大量数据中发现规律,揭示数据特征、预测发展趋势、辅助决策。基于数据挖掘的一种故障诊断方法是利用设备故障时的动态事件序列信息的时空特性,形成故障过程的时间序列信息模型,并运用基于事件序列相似性的数据挖掘原理挖掘出事件序列信息中隐含的关联性;基于数据挖掘的另一种诊断方法是采用距离函数来度量事件序列的相似性,用动态规划算法实现相似性的最优代价运算,将故障诊断问题转化为寻找与实际故障事件模式最相似的标准故障事件模式。因此,采用数据挖掘技术能够实现对设备的故障诊断,发现其中有用的设备故障特征和故障模式。

### 3.2.3 铁路设备安全风险源识别

安全风险管理是通过知识生产经营活动中存在的危险、有害因素,并运用定性或定量的统计分析方法确定其风险严重程度,进而确定风险控制的优

先顺序和风险控制措施,以达到改善安全生产环境、减少和杜绝安全生产事故的目标。铁路设备的安全风险管理主要是针对设备状态的安全风险管理,包括安全风险源识别、安全风险评估和安全风险控制。

基于故障诊断结果,可进行风险源识别。与故障设备类型相同、具有相同条件(故障原因)尚未发生故障的设备安全风险最大,如一根正在服役的钢轨因钢轨材质发生断裂,与其同一炉生产的钢轨都是潜在的风险源,应针对性地进行处理。同时,与设备故障特征相近或故障模式相似的设备也具有较高的安全风险。此外,应根据单个设备的个性化状态变化规律,预测未来设备状态达到或超出安全管理容许值的时间序列,把剩余寿命短的设备作为安全风险进行管理。

### 3.2.4 铁路设备安全风险评估

基于大数据技术进行铁路设备安全风险评估应构建2个层面:(1)基于设备故障特征和故障模式的设备安全风险评估方法,根据设备故障诊断结果,对故障特征和故障模式进行数字化表达,建立设备的特征、模式与故障特征、故障模式的距离函数来度量设备状态与标准故障状态的相似程度,距离函数的值越小,则设备安全风险越高。(2)基于设备健康状态的设备安全风险评估,根据设备当前的状态数据,划分不同的安全风险等级,设备状态越差,其安全风险等级则越高。

### 3.2.5 铁路设备状态维修

状态维修(CBM, Condition Based Maintenance)包含以可靠性为中心的维修(PCM, Reliability Centered Maintenance)和预测性维修(PM, Predictive Maintenance)。RCM是在对元件的可能故障对整个系统可靠性影响评估的基础上决定维修计划的一种维修策略。PM是根据对潜伏故障进行在线或离线测量的结果和其他信息来安排维修的技术。其关键是依靠先进的故障诊断技术对潜伏故障进行分类和严重性分析,以决定设备(部件)是否需要立即退出运行或应及时采取的措施。

针对安全风险等级极高的设备,采用个性化维修策略,及时安排设备维修,消除安全隐患,降低安全风险等级;针对安全风险等级高的设备,采用

个性化的设备状态监视策略，增加状态检测、监测的频率，重点监测设备状态变化，有针对性地安排维修计划，将安全风险控制在允许范围内；针对风险等级一般的设备，适当增加状态检测、监测的频率，重点监测设备状态变化速率，防止设备状态突变，将安全风险控制在允许范围内。

### 3.3 铁路环境相关的大数据分析需求

铁路环境相关的大数据分析需求主要是铁路沿线自然灾害预测（如风、雨、雪、地震、滑坡等）和异物侵限监测。传统的自然灾害预测方法有神经网路、灰色预测、马尔科夫模型等方法，由于铁路沿线视频监控等监测数据量很大，传统的方法出现了局限性。因此可以利用大数据技术，诸如聚类分析、关联分析、回归分析等，分析铁路沿线监测装置、线路环境数据，得出自然灾害发生规律，提前预防，保障行车安全；异物侵限事件的发生具有突发性等特点，目前，主要采用双电缆、光缆、红外线、微波、视频等技术进行异物侵限的监测，可以采用大数据技术分析异物侵限相关监测数据，进行异物侵限的预防维护，如图 7 所示。

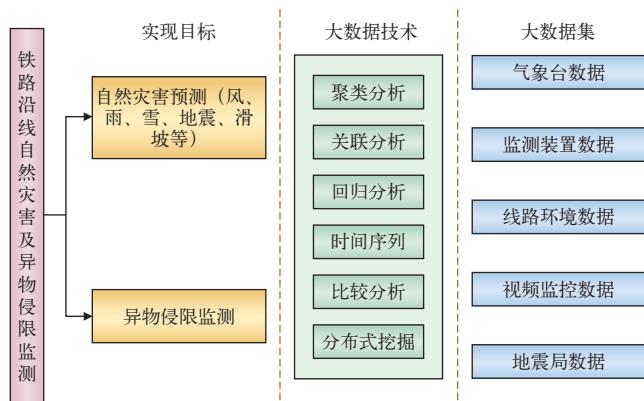


图 7 铁路沿线自然灾害及异物侵限监测

#### 3.3.1 铁路沿线自然灾害预测

由于自然灾害因子众多，动力机制复杂，强风、大雨、大雪等自然灾害发生的范围和程度呈现出趋势性、周期性和随机性的叠加状态；地震灾害等具有突发性的特点。针对不同类型的自然灾害，需要采用不同的预测方法来应对。在大数据技术应用之中，对风、雨、雪等自然灾害的预测可以采用数学建模的方法，分析出铁路沿线发生大风、大雨等自然灾害的模型，进而进行预测，指挥行车；对于地震、

滑坡、泥石流等突发性自然灾害，则要利用大数据技术对海量历史数据进行分析，解决传统数据分析无法分析的问题，得出其发生的大致规律。实现提前预防，减少自然灾害对行车安全的影响。

#### 3.3.2 铁路异物侵限监测

异物侵限监控系统是异物侵限防灾系统研究、发展的主要形式，一旦发生异物侵限，系统应能及时发出报警，并将侵限信息实时传送到行车调度中心，为下达行车控制、维修管理等指令提供依据，有效地避免重大行车事故发生。目前异物侵限监测技术主要有双电缆传感器、光缆传感器、红外线、微波和视频监控等。

通过大数据技术，结合异物侵限发生的时间、发生环境、造成危害等数据，利用数据挖掘技术，挖掘出异物侵限发生的大致规律，为异物侵限监测提供提前的预防措施，实现防范于未然。

## 4 结束语

本文根据业务需求确定了铁路安全管理领域中大数据应用的需求模型。铁路安全管理大数据的应用应综合考虑数据存储、数据分析和可视化展示等关键技术和问题。需要借鉴互联网公司对大数据应用的成熟经验，构建大数据在铁路安全管理应用的技术框架。未来的铁路安全管理，从“人-设备-环境”3个方面，在充分对海量的历史数据学习和挖掘的基础上，实现提前预防，超前防护，做到防范未然。

## 参考文献：

- [1] 陶雪娇,胡晓峰,刘洋.大数据研究综述[J].系统仿真学报,2013, 25 (S) : 142-146.
- [2] 维克托·迈尔·舍恩伯格.大数据时代：生活、工作与思维的大变革 [M].杭州：浙江人民出版社，2012.
- [3] 余振国.铁路安全监督管理信息系统的应用与实现 [D].北京：北京工业大学，2011.
- [4] 肖维斯.物联网在铁路安全监控领域应用研究 [D].北京：北京交通大学，2012.
- [5] 徐超.高速铁路综合防灾安全监控系统的研究 [D].北京：中国铁道科学研究院，2010.
- [6] 彭宇,刘大同,彭喜元.故障预测与健康管理技术综述 [J].电子测量与仪器学报, 2010, 24 (1) : 1-7.

责任编辑 徐侃春