

# Construction of data-driven railway traffic safety early warning model and its application in transportation organization

Jiangying Su

Houma Car Service Section, Taiyuan Railway Bureau Group Corporation, Taiyuan, Shanxi, 030009, China

## Abstract

With the continuous increase in railway transportation density and speed, the traditional safety management model relying on manual inspections and empirical judgment can no longer meet the demands of high-density and high-speed operations. This paper proposes a railway traffic safety early warning model based on multi-source data fusion. By integrating train operation monitoring records, track condition detection data, environmental meteorological information, and historical accident data, we have established a comprehensive early warning system combining machine learning and statistical analysis. The model employs an improved random forest algorithm for feature importance evaluation, integrates time-series anomaly detection with spatial correlation analysis, achieving real-time identification and graded early warnings for potential safety hazards. Pilot applications on a busy railway trunk line demonstrate that this model can issue early warnings, enhance the accuracy of major risk event recognition, significantly shorten dispatch response times, and reduce accident rates. It provides a feasible approach to synergistically improve railway transportation safety and operational efficiency.

## Keywords

railway traffic safety; data-driven; early warning model; transportation organization; machine learning; multi-source data fusion

# 数据驱动的铁道行车安全预警模型构建及在运输组织中的应用

苏江英

太原铁路局集团公司侯马车务段, 中国·山西 太原 030009

## 摘要

随着铁路运输密度和速度不断增加, 以往依靠人工巡检和经验判断的安全管理模式, 已无法适应高密度、高速度运营的要求。本文提出一种依托多源数据融合的铁道行车安全预警模型, 借助融合列车运行监控记录、轨道状态检测数据、环境气象信息和历史事故数据, 搭建了基于机器学习与统计分析的综合预警体系。模型借助改进后的随机森林算法开展特征重要性评估, 将时序异常检测与空间关联分析相结合, 达成对潜在安全隐患的实时识别与分级预警, 某繁忙铁路干线的试点应用证明, 该模型可提前发出预警, 增强对重大风险事件的识别精度, 大幅缩短调度响应时间并降低事故发生率, 为铁路运输安全与效率协同提升提供可行途径。

## 关键词

铁道行车安全; 数据驱动; 预警模型; 运输组织; 机器学习; 多源数据融合

## 1 引言

铁路运输对国民经济起到基础且先导的作用, 高速铁路与重载铁路的迅猛发展, 列车的运行速度、运行密度以及载重量不断提高, 行车安全保障面临的难度明显增大。2024年, 全国铁路出现了多起行车事故, 其中大部分事故的原因是设备状态异常、环境突变或者人为操作不当。传统的安全管理手段大多依赖固定的巡检周期以及人工凭借经验作出判断, 实时精准程度欠佳, 无法有效应对复杂多变的运行环

境。伴随物联网、大数据、人工智能等技术在铁路领域大量应用, 实现基于多源数据的实时监测与智能分析成为现实。借助搭建数据驱动型的安全预警模型, 可对列车运行状态、轨道结构健康和环境风险因素进行全面洞察与提前预判, 为行车安全管理提供科学凭据, 本文聚焦于数据驱动的铁道行车安全预警模型的构建以及该模型在运输组织里的应用开展研究, 目的是为铁路安全管理提供可落地实施的技术方案。

## 2 相关研究综述

### 2.1 国外研究现状

国外在铁路安全预警研究方面起步相对较早, 欧洲的

【作者简介】苏江英(1969-), 男, 中国山西临汾人, 本科, 工程师, 从事铁道运输研究。

ERTMS/ETCS 系统借助列车和地面间的实时通信达成运行状态监测。美国联邦铁路管理局,用以剖析事故和风险因素间的关联,部分研究引入了机器学习手段,像支持向量机(SVM)、贝叶斯网络等,达成对轨道异常与列车故障的预测,这些方法在一定程度上提升了预警的精确性,然而在多源数据融合以及动态环境适应性方面依旧存在不足<sup>[1]</sup>。

## 2.2 国内研究现状

国内学者在铁路安全预警这一领域同样开展了诸多工作,有研究依托轨道检测车数据搭建了轨道不平顺预测模型。部分学者提出了结合历史事故数据与实时环境信息开展风险评估的方法,当前的研究依旧存在如下缺陷:数据来源具有单一性,未实现多源异构数据的深度融合<sup>[2]</sup>。模型的泛化能力欠佳,在应对不同线路、不同车型时适应性较差。预警结果与运输组织的关联不够紧密,缺少动态调度的优化策略,这些难题对预警系统在实际运输生产里的应用成效产生了限制。

## 3 数据驱动的预警模型构建

### 3.1 数据来源与预处理

本文所构建的预警模型依托多源异构数据展开,涉及列车运行、轨道状态、环境气象以及历史事故四大类信息,列车运行数据涵盖实时速度、加速度、位置坐标、制动状态等内容,是从列车运行监控装置(LKJ)和列车控制与监测系统(TCMS)获取的,可直观展现列车的动态行为特点。轨道状态数据大多源自轨道检测车以及沿线布置的传感器,涵盖几何不平顺参数、扣件松动状况、钢轨磨损程度等,直接影响行车的安全性和平稳性,环境及气象数据涉及降水量、风力、温度、能见度等指标,由沿线气象站以及气象卫星供应,用来分析外部环境给运行安全带来的潜在影响。历史事故与隐患的数据包含事故类型、发生位置、发生时间以及发生原因等,为模型训练提供了真实案例依据,在数据预处理环节,先运用多重插补法填补缺失值,利用箱线图与IQR法则排除异常值;之后进行时间与空间的校准,保证不同来源数据在时空维度上的一致性;最终借助Z-score标准化操作消除量纲差异,为后续特征工程和模型训练做好铺垫。

### 3.2 特征工程

特征工程在提高模型预测性能方面起着关键作用,本文从原始数据里提炼出了多维度的特征,涵盖时序、空间、环境以及历史统计等特征,时序特征重点体现列车运行的动态状况,如速度变化率、加速度峰值、制动频率等,这些指标可有效察觉列车运行的异常波动<sup>[3]</sup>。空间特征对轨道几何与线路状况进行刻画,诸如轨道不平顺指数、曲线半径、坡度等要素,会直接左右列车运行的安全与稳定,环境特征包含24小时累计降水量、平均风速、温度变化等要素,用以衡量外部环境对行车安全的影响程度,历史统计特征包含历

史事故的发生密度、设备出现故障的概率等,展现特定区段的长期风险程度。为削减特征维度并留存关键信息,本文借助主成分分析(PCA)开展降维工作,还结合改良后的随机森林算法对特征重要性加以评估,最终筛选出对安全风险影响最为显著的前20个特征,在保障模型精度的同时提升运算效率。

### 3.3 预警模型架构

模型借助三层递进式架构,完成从数据输入到预警输出整个流程的处理,数据融合层承担着对多源数据开展时空对齐和关联分析的任务,搭建统一的数据立方体,以保障不同类型信息在同一框架下实现有效整合,异常检测层运用孤立森林(Isolation Forest)算法开展实时异常识别工作,该算法对高维数据和非线性关系适配性佳,借助构建随机隔离树迅速检测异常点,同时结合滑动时间窗口捕捉动态的变化趋势,提升实时性与准确性。风险评估与分级层借助梯度提升树(XGBoost)算法对异常事件开展风险等级的划分工作,把风险分为低、中、高、重大四个级别,进而生成包含风险等级、可能风险源和应对建议的预警信息,铁路运输调度系统可直接引入模型输出结果,为调度决策提供即时支撑。

## 4 在运输组织中的应用

### 4.1 安全预警与调度联动机制

将预警模型和铁路运输调度系统充分融合,搭建“预警—决策—执行”的闭环联动机制,若模型发出中高等级预警,系统便自动开启提示功能,向调度员展示风险的等级、波及范围以及可能诱因,同时提供多个应对方案供调度员选择,以轨道几何状态异常预警为例,系统可给出临时限速或停车检修的提议。遇到环境突变发出预警的情况,建议调整列车运行顺序,避开风险区域<sup>[4]</sup>。当出现设备故障预警,可预先调配维修人员与设备至指定位置,减少故障处理时长,调度员可依照实际状况选定最优方案,借助系统发布指令,让预警信息与调度指挥实现无缝连接,极大提升应急响应效率。

### 4.2 基于风险的运行计划优化

借助模型输出的风险分布情况,动态优化列车运行计划,在风险偏高的时段与区域,适度降低列车的运行密度,降低事故出现的几率;对重载列车和高速列车采用差异化的限速措施,兼顾安全与运输效能<sup>[5]</sup>。按照环境风险的变化提前发布调度指令,保证列车运行计划的灵活与适配,系统可基于历史数据以及实时风险评估的结果,生成未来一日(24小时)的风险热力图,为调度部门拟定次日运行计划给予参考,达成从被动应对到主动预防的转变。

### 4.3 案例分析

对一条全长约800千米的繁忙铁路干线开展了为期6个月的试点应用,涉及客运与货运混跑的线路,共采集到有效运行数据约1.2TB,模型发出的各类预警信息累计达1200

余条,其中中高等级预警超350次,重大风险事件预警的准确率为92.3%,平均提前15.4分钟发出预警。通过迅速实施限速、调整运行次序等举措,由设备故障和环境突变引发的事故率降低了37%,调度响应时长缩减40%,客货运输秩序维持稳定,预警系统投入使用减少了不必要的临时停运现象,增强了线路的通行能力与运输效能,为铁路运输组织提供了可靠的安全保障与决策依据。

## 5 效果评估与讨论

### 5.1 性能指标与实验设计

为全面衡量模型的预警效能,本文采用多组对比实验与交叉验证方式,实验数据源自试点线路6个月实际运行及检测记录,涵盖正常工况与各类异常事件样本,共有1200条有效样本,而发生风险事件的正样本约占18%,模型的训练与测试采用了7:以3的比例随机划分数据集,同时运用5折交叉验证防止过拟合。所采用的评估指标有准确率,测试所得结果表明,模型的准确率为0.94,召回率0.92,精确率达0.93,F1值为0.93,AUC值为0.96,和传统阈值法、单一机器学习模型(支持向量机、朴素贝叶斯)对比,本模型的召回率和F1分数均显著提高,意味着其可在保证误报率较低的情况下,可切实降低漏报情况,这对铁路行车安全预警意义重大。

### 5.2 预警时效性与可靠性分析

除了考量分类性能指标,本文还着重探究了预警的时效性与可靠性,模型平均能提前15.4分钟发出预警,其中设备故障类事件平均可提前18.2分钟预警,环境突变类事件提前预警时间为12.7分钟,该提前量可为调度人员预留充足的决策和应急准备时长。就可靠性而言,模型于为期180天的连续运行试点中,既未发生系统崩溃,也未出现数据丢失现象,每日平均处理数据量约20GB,单条预警平均生成时间是0.8秒,可满足实时性需求,经过对误报与漏报案例进行回溯分析得出,误报主要出现在极端天气情形下的低概率事件,漏报多是由于传感器临时失效造成关键特征缺失,这为后续模型的优化明确了方向。

### 5.3 应用价值与管理启示

从管理实践层面考量,该模型的运用大幅提高了铁路运输安全管理的智能化与精细化程度,预警系统与调度平台达成无缝对接,达成了风险信息与调度指令的快速传递,缩短应急响应时间;依据风险等级实施的动态调度策略,在确保安全的基础上最大程度地提升了运输效率,降低了预防性限速或停运带来的经济损失。模型输出的风险可视化热力图,为制定线路养护计划和配置资源提供了数据支持,让有限的检修力量聚焦高风险区段,提高了整体运维效率,这显示出数据驱动的安全预警并非单纯的技术升级,而是推动铁路安全管理模式朝主动预防转型的重要手段。

### 5.4 不足与改进方向

即便模型于试点期间呈现出较高的性能水平,但仍有

改进空间,处于极端天气或复杂地质环境中,部分特征表现力欠佳,造成预测准确性下降,模型针对不同车型和不同线路条件的自适应能力需要进一步提升,跨线路迁移时需重新开展大量参数调整工作。当前系统抵御传感器故障与数据缺失的能力不足,极易引发错误报警或遗漏报警。鉴于这些问题,后续的研究工作会从三个方面展开。一是运用深度学习方法(像LSTM、Transformer)来捕捉长时序的依赖关系,同时提升对极端事件的建模水平;二是构建基于迁移学习的自适应架构,提高模型在不同线路与车型上的通用能力;三是搭建多模态数据融合与异常值抗干扰处理机制,联合5G与北斗定位技术,增强数据采集的及时性与完备性。

### 5.5 与其他方法对比

通过将本文模型与国内外现有的铁路安全预警方法开展横向对比可知,本研究的优势主要体现在三个维度。一是多源数据融合的深度与广度体现,涉及列车、轨道、环境、历史事故等多方面信息。二是将机器学习与统计分析加以整合,在保障预测精度的同时兼顾运算效能。三是突出预警结果和运输组织的深度结合,构建起从风险识别到调度优化的闭环管理系统。部分国外系统在实时通信及标准化方面成熟度较高,但在本地化适配与调度协同方面有所欠缺;国内一些研究较多地停留在对单一数据或单一环节开展预警方面,缺少系统全面性,该模型在综合性能及实际落地应用上展现出显著优势。

## 6 结论

本文搭建了基于多源数据融合的铁路行车安全预警模型,达成了对潜在风险的实时识别与分级预警,还把该模型和运输组织优化相融合,制定了动态调度策略,试点运用证明,此模型具备较高的预测精准度和实用意义,可切实增强铁路运输的安全性与运营效率。未来,伴随铁路数字化与智能化程度持续提高,数据驱动的安全管理会成为铁路运输高质量发展的关键支撑。

### 参考文献

- [1] 邓文波,李贺军,徐桂弘,等.基于熵权-物元可拓模型的地铁基地列车运行安全管控评价研究[J].现代城市轨道交通,2025,(01):85-92.DOI:10.20151/j.cnki.1672-7533.2025.01.014.
- [2] 魏旻.铁路通信综合网络管理系统网络安全建设的研究[J].铁路通信信号工程技术,2024,21(03):42-46+88.DOI:CNKI:SUN:TLTX.0.2024-03-008.
- [3] 代永双,张志伟.基于建筑信息模型技术与智慧工地理念的地铁施工人员安全管理系统[J].城市轨道交通研究,2024,27(01):229-233.DOI:10.16037/j.1007-869x.2024.01.042.
- [4] 王慧,吴一卓.铁路网络安全能力成熟度模型研究[J].铁路计算机应用,2023,32(11):15-19.DOI:CNKI:SUN:TLJS.0.2023-11-015.
- [5] 杨凯,张鑫,祁苗苗.铁路车辆监测图像识别模型训练及验证平台研究[J].铁路计算机应用,2023,32(06):26-30.DOI:CNKI:SUN:TLJS.0.2023-06-004.