

# 随机森林在轨道交通车辆门系统亚健康状态诊断中的应用

严 俊<sup>1</sup> 支有冉<sup>2</sup> 许志兴<sup>2</sup> 史 翔<sup>2</sup>

(1. 上海地铁维护保障有限公司, 200070, 上海; 2. 南京康尼机电股份有限公司, 210013, 南京//第一作者, 高级工程师)

**摘 要** 提出了一种基于随机森林(RF)的车门亚健康状态诊断方法。首先,从车门驱动电机的电流、转速和转矩输出信号中提取时频域特征指标作为表征车门运行状态的特征向量;然后,RF通过对分类器的集成来提高车门亚健康状态诊断的精度,并且从特征指标对分类结果影响的角度评估特征的重要性。利用台架调门试验模拟车门的典型亚健康工况,采集数据并分析验证。试验结果表明,与传统的单分类器(决策树、支持向量机、贝叶斯、KNN近邻)的诊断结果相比,RF方法在车门的亚健康状态诊断中具有更高的诊断精度,并可对特征的重要性进行评估。

**关键词** 轨道交通;车辆门系统;亚健康诊断;随机森林

**中图分类号** U279.3+23

**DOI**:10.16037/j.1007-869x.2019.12.027

## Application of Random Forest in the Sub-health Diagnose of Railway Vehicle Door System

YAN Jun, ZHI Youran, XU Zhixing, SHI Xiang

**Abstract** The sub-health status diagnose of railway vehicle door system (RVDS) based on Random Forest (RF) is proposed. Firstly, features in time domain and frequency domain are extracted from the output signal of the door driving motor, including current, speed and torque as the feature vectors of RVDS operation; secondly, by integrating the classifiers, RF will improve the accuracy of sub-health diagnosis and evaluate the importance of features, from the perspective of feature vectors impact on the final classification. And finally, the bench adjustment experiment is used to simulate the typical sub-health conditions of RVDS, data are collected for analysis and verification. The results show that compared to the traditional single classifier (decision tree, support vector machine, Bayes, KNN algorithm), RF method has higher diagnostic accuracy in RVDS sub-health diagnose, and can assess the importance of features.

**Key words** rail transit; railway vehicle door system; sub-health diagnose; random forest

**First-author's address** Shanghai Metro Maintenance and Guarantee Co., Ltd., 200070, Shanghai, China

在列车运营过程中,由于车门开闭频繁及乘客干扰等众多因素,导致客室车门故障频发<sup>[1]</sup>。据相关数据统计,车门故障发生次数约占列车系统总故障的30%以上。在故障发生之前,车门通常会在亚健康的状态下运行一段时期。因此,及时发现车门系统常见的早期亚健康征兆,进行实时的健康状态监测并提供亚健康状态判别,对提高车门系统的安全性、可靠性以及降低故障率具有重要意义。

目前,多种算法在轨道交通车辆门系统的状态检测领域取得了较好的诊断效果,如人工神经网络、Petri网、贝叶斯网络<sup>[2-3]</sup>、聚类分析<sup>[4]</sup>等。文献[1]以轨道交通车门为研究对象,提出了以故障树为基础的混合结构树,实现了自动门的远程智能诊断。文献[2]提出了一种利用贝叶斯网络对车门驱动电机异常状态进行可靠性分析的方法。文献[5]提出以支持向量机作为辨识算法建立设备的状态评估模型。上述方法均通过训练数据建立单一分类器,然后对测试样本集进行状态识别,从而达到诊断的目的。已有研究表明,与单一分类器相比,分类器群能够达到更高的正确率<sup>[6]</sup>。文献[1]提出的随机森林(RF),就是一种利用多个随机树分类器组合进行分类的方法,RF从特征集中随机选择部分特征量形成特征子集,利用不同的特征子集构造分类器,有效地提高了分类器之间的差异,具有较好的泛化性和准确性。

然而,轨道交通车门数据采集系统通过配备的智能电机返回的数据参量只有电机转矩、转速和角度,原始数据无法直接应用分类器群进行训练。因此,RF在轨道交通车辆门系统的亚健康诊断领域中具有重要的研究意义。

# 1 轨道交通车辆门系统

## 1.1 车门结构与典型的亚健康状态

本文以电动双开塞拉门系统为研究对象。其车门系统的运行由电机驱动,门控器控制驱动电机的电流、转速和位移,电动机带动丝杆,驱动螺母,螺母带动门扇运动,从而实现车门的自动开关。

随着列车的持续运营,各部件尺寸由于元件老化及磨损,车门一直在亚健康状态下运行,直接影响着车门的正常使用性能。本文以车门的安全列出体系规范及实际运营维修返回的相关技术统计为依据,研究了车门的 7 种典型亚健康工况。表 1 列出了全部 7 种亚健康状态的具体情况。

表 1 车门典型的亚健康状态表

亚健康状态	潜在危险
对中尺寸异常	开门阻力增大甚至无法关闭
V 型尺寸异常	障碍检测功能异常,影响车门的密封性
门扇高度异常	开关门阻力增大,障碍检测功能异常
门扇开度减小	影响开门障碍检测
门扇开度增加	障碍检测功能异常
密封尺寸减小	关门局部阻力增大
密封尺寸增加	开门阻力增大甚至无法关闭

## 1.2 门系统健康状态评估特征集

本文所研究的车门状态评估特征集建立在列车运行过程中的 5 个基本阶段(启动段、升速段、匀速段、减速段及缓行段)之上。分别对各运行阶段提取运行时间、行程、转速、电流的平均值以及有效值、极大值、极小值等指标,作为时域特征向量。本文采用小波变化对原始速度信号进行 5 层小波分解,然后对高频分量进行了多尺度重构,选择容错性强的 Shannon 能量熵值作为时频域的特征指标。上述提取的时频内总计 39 个特征组合成为一组特

征向量  $F=[F_1,F_2,\cdots,F_{39}]$ ,形成表征车门运动状态的特征集。

## 2 RF 方法

RF 是由多个决策树  $\{h(x,\Theta_k),k=1,2,\cdots,n\}$  组成的分类器群,其中,  $\{\Theta_k\}$  是相互独立且同分布的随机向量。通过对基分类器的集成和候选特征子集的修改,使分类性能得到提高,最终由所有决策树综合投票决定输出的结果。考虑到若森林中的多个分类器结构相同,并采用相同的训练集训练,会导致分类器之间相似度极高,以及多个分类器类似于一个分类器。

## 3 试验结果与分析

### 3.1 实验数据采集与特征提取

本文针对轨道交通车门的亚健康诊断进行亚健康台架调门试验,车门的采样周期为 10 ms,一次采样 350 个点,并通过无线传输的方式(由外置监测设备转发)将数据信息传输到数据中心服务器。首先检查台架车门的相关尺寸和功能是否符合检验标准,并确认该车门未存在任何形式的亚健康状态,此时执行一段时期的开关门循环,通过远程系统在服务器上采集这段时期内的电机输出数据共 247 组。采集上述 7 类典型的亚健康状态下车门系统的运行数据,其中,对中尺寸异常为 193 组、V 型尺寸异常为 208 组、门扇高度异常为 196 组、门扇开度减小为 165 组、门扇开度增加为 188 组,密封尺寸减小为 170 组,密封尺寸增加为 192 组;采集到的车门正常健康数据为 147 组。统计得到上述 7 类车门亚健康数据和正常数据样本的总容量为 1 459 组。车门系统中电机的一组输出参数如图 1 所示。

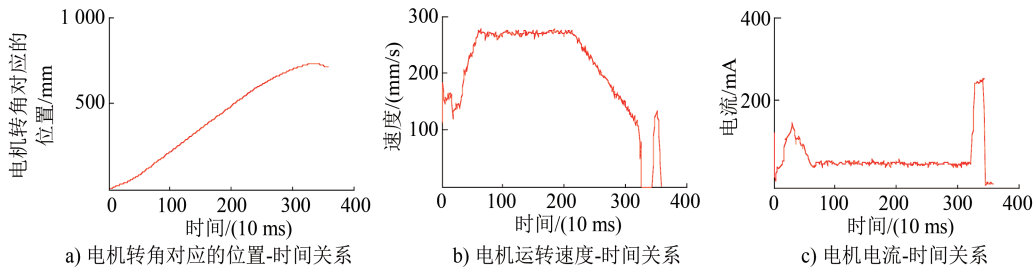


图 1 车门系统中电机的一组输出参数曲线图

### 3.2 方法结果验证

将上述经特征提取后的数据所对应车门的健康、对中尺寸异常、V 型尺寸异常、门扇高度异常、门

扇开度增加、门扇开度减小、密封尺寸增加、密封尺寸减小等 8 种车门的健康和亚健康状态分别标上标签,记为  $C=[0,1,2,3,4,5,6,7,8]$ ,并作为 RF 亚健

康诊断模型的输入。图 2 显示了 RF 的诊断错误率随森林中决策树数量增加的变化曲线。

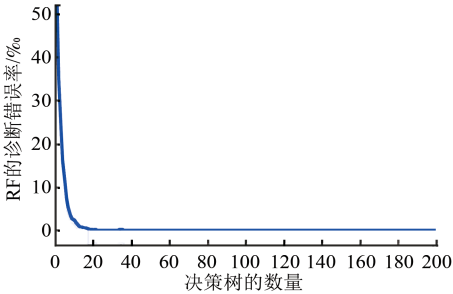


图 2 RF 的诊断错误率与森林中决策树数量的关系曲线

由图 2 可知,前期随着决策树数量的增加,错误率大幅降低;当决策树数量超过 40 后,错误率趋于稳定,约为 0.16%,但模型的复杂度和训练时间会随决策树数量的增加而增加。因此,综合考虑模型的准确率与复杂度,本文选取 RF 决策树的数量为 40。通过上述方式确定决策树的数量后,采用式(3)对构成模型输入数据的 39 维特征进行重要性评估。表 2 列出了按重要性对特征进行排序的结果。

表 2 部分特征重要性排序表

特征序号	特征值	重要性评分
4	电流有效值	1.198 3
22	匀速段最大电流	1.150 0
3	最大电流	1.056 0
28	缓行段时间	0.990 3
2	总行程	0.975 5
10	初始段时间	0.959 1
20	匀速段最大转速	0.947 3
35	到位反转角度	0.906 4

由表 3 可知,排名前列的特征依次是全程电流的有效值、匀速段最大电流、全程最大电流等。由于列车亚健康运行会导致车门运行中阻力发生变化,而电流则是直接和阻力相关的观测量,基于 RF 的特征重要性评估得出的结论与实际情况相符,证明了该方法的可行性。因此,在阵线列车检测系统中,可对表 3 中排名靠前的特征进行重点监控。本文采用 10 次 10 折交叉验证的平均值作为模型最终精度评价方式,结果如图 3 所示。

通过上述多种诊断模型的对比结果发现,对于车门亚健康数据的诊断,贝叶斯网络和 KNN 近邻诊断的效果较差,1 000 条数据的错误诊断数量在 10 条以上;支持向量机和决策树错判的数量为个位数,决策树略好于支持向量机。而 RF 的诊断错误率比贝叶斯网络和 KNN 近邻低了两个数量级,比

决策树和支持向量机的诊断错误率低 0.5%~0.8%,由此可知其诊断效果在 5 种方法中效果最好。

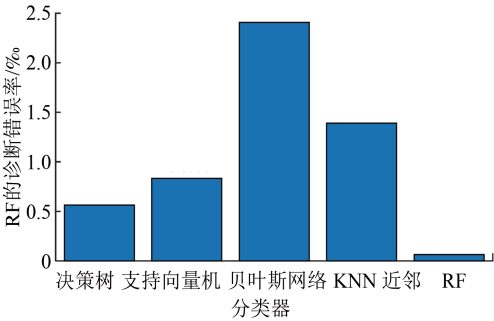


图 3 各分类器模型最终的诊断结果比较图

4 结论

1) 分析车辆门系统亚健康的根源,建立典型的亚健康模式与其表现形式之间的关联。对车门的运动曲线进行合理的分段,并基于分段结果提取了一组能全面表征车门健康状态的特征集。

2) 利用台架调门亚健康试验,采用 RF 的方法对车门健康状态和 7 种典型的车门亚健康状态数据进行模型的训练和诊断,并基于 RF 提出了一种评估特征重要性的方法,结合车门的工作机理,证实该评估结果的可靠度较高。

3) 对多种分类器算法进行了对比试验,结果表明针对车门的亚健康诊断,与其它传统的算法相比,RF 的诊断效果最好,错误率可降低 1% 左右。因此,RF 方法在轨道车辆门系统的亚健康诊断方面有着较好的实际推广意义。

参考文献

[1] 高文明,李志保,唐谦,等.轨道车辆门远程监控与故障诊断系统软件设计[J].机械设计与制造工程,2016(10): 49.

[2] 周巧莲,金碧筠,冒玲丽,等.基于贝叶斯网络地铁车门系统可靠性分析和故障诊断[J].沈阳工业大学学报,2014(4): 441.

[3] 徐霖.城轨列车客室车门系统故障诊断方法研究——基于改进的 TOPSIS 法与贝叶斯网络[D].北京:北京交通大学,2015.

[4] FINK O, ZIO E, WEIDMANN U. Fuzzy classification with restricted boltzman machines and echo-state networks for predicting potential railway door system failures [J]. IEEE Transactions on Reliability, 2015(3): 861.

[5] 洪杰,韩磊,苗学问,等.基于支持向量机的滚动轴承状态寿命评估[J].北京航空航天大学学报,2010(8): 896.

[6] 胡青,孙才新,杜林,等.核主成分分析与随机森林相结合的变压器故障诊断方法[J].高电压技术,2010(7): 1725.

(收稿日期:2019-06-11)