

基于数据驱动的高速铁路动车组列车头尾车 车轮踏面缺陷预测方法*

王文琦¹ 宋冬利¹ 李 林² 刘 翊² 张卫华¹ 郑则君¹

(1. 西南交通大学轨道交通运载系统全国重点实验室, 610031, 成都; 2. 株洲国创轨道科技有限公司, 412007, 株洲)

摘 要 [目的] 踏面缺陷是高速铁路动车组列车车轮失效的主要表现形式之一, 严重影响了动车组运行的安全性及乘客的乘坐舒适性。车轮踏面缺陷主要集中在头尾车, 可能是多种因素共同作用的结果, 需要寻找一种能综合多种影响因素的预测方法。[方法] 基于某铁路局动车组列车的车轮镟修数据, 每个数据样本包括 10 个特征(4 个名义特征和 6 个连续特征), 对数据进行预处理。通过合成少数类样本过采样技术对不平衡数据集进行处理, 构建了标准化数据集。建立了 DNN(深度神经网络)模型, 对底层特征进行组合, 形成了特征的高层抽象表示。通过网络结构调整和超参数优化得到了模型的最优学习效果。对模型进行训练并测试, 验证了该模型的预测效果。[结果及结论] 基于数据驱动的头尾车车轮踏面缺陷预测方法具有较高的预测精度和较优的综合性能, 其预测精确率达 92.5%, 可有效预测头尾车车轮踏面损伤的发生概率。

关键词 高速铁路; 动车组; 数据驱动; 踏面缺陷; 过采样技术; 深度神经网络; 超参数优化

中图分类号 U463.34; U269.6

DOI: 10.16037/j.1007-869x.2024.04.006

Data-driven Prediction Method for High-speed Railway EMU Train Front and Rear Car Wheel Tread Defects

WANG Wenqi¹, SONG Dongli¹, LI Lin², LIU Yi², ZHANG Weihua¹, ZHENG Zejun¹

(1. State Key Laboratory of Rail Transit Vehicle System, Southwest Jiaotong University, 610031, Chengdu, China; 2. Zhuzhou Guochuang Track Technology Co., Ltd., 412007, Zhuzhou, China)

Abstract [Objective] Tread defects are a primary manifestation of wheel failures in high-speed railway EMU (electrical multiple unites) trains, significantly impacting both EMU train operation safety and passenger ride comfort. Wheel tread defects are predominantly concentrated on front and rear cars, which may result from a combination of various factors, requi-

ring predictive method that comprehensively integrate various influencing factors. [Method] Based on the wheel reprofiling maintenance data of EMU train operated by a railway bureau, the dataset sample consists of 10 features (including 4 nominal and 6 continuous features) and the data are preprocessed. By treating the imbalanced dataset through synthetic minority over-sampling technique (SMOTE), a standardized dataset is constructed. A DNN (deep neural network) model is established to combine the underlying features and form a high-level abstract representation of the features. The optimal learning performance of the model is achieved through network structure adjustment and hyperparameter optimization. The model is trained and tested to verify its prediction effect. [Result & Conclusion] The data-driven prediction method for wheel tread defects of front and rear cars demonstrates high-predictive accuracy and relatively excellent comprehensive performance, achieving a precision rate of 92.5%. Thus the probability of wheel tread damage of front and rear cars can be effectively predicted. **Key words** high-speed railway; EMU; data-driven; tread defect; over-sampling technology; deep neural network; hyperparameter optimization

随着我国高速铁路的发展, 动车组列车运行速度和运量不断提高, 其轮轨间动态相互作用进一步增强, 进而导致车轮踏面缺陷问题逐渐增多^[1]。通过分析某铁路局动车组列车 2016—2021 年的检修数据(共计 93 组), 发现车轮踏面缺陷主要集中在头尾车, 发生踏面缺陷的主要原因是动车组驶入弯道时, 头尾车率先进入, 头尾车受到的横向冲击力及纵向力较大^[2]。此外, 车轮几何参数、外界环境及运行里程都是车轮踏面缺陷潜在的影响因素。车轮踏面缺陷一般无法由轮对动态检测系统进行识别, 需人工进行检查, 这不仅浪费时间, 且存在人为因素, 可能导致问题车轮无法被及时发现。

* 国家重点研发计划项目(2019YFB1405401); 西南交通大学学科交叉基础研究项目(2682021ZTPY099)

目前,大多数学者关注的是车轮踏面缺陷的形成机理及其对车辆系统动力学的影响,而对检修数据的分析仍停留在简单的统计分析阶段^[3]。本文提出了一种基于数据驱动的高速铁路动车组头尾车轮踏面缺陷预测方法,该方法采用了数据集预处理、SMOTE-NC(名义特征和连续特征结合的合成少数过采样技术)、扩充不平衡数据集、原始数据集的抽象特征提取及 DNN(深度神经网络)等技术和算法,对动车组列车车轮踏面缺陷进行预测。现场检修数据测试结果表明,此预测方法具有较高的预测准确率,其预测结果可指导车轮的镟修作业,提高检修效率,减少人为误差。

1 特征选择与数据预处理

动车组列车车轮每运行 20 万~30 万 km 后需要进行镟修,以恢复车轮的标准廓形,避免车轮滚动接触疲劳、踏面剥离、多边形等因素对车辆动力学性能产生影响。在不补轮缘的情况下,动车组列车车轮的镟修量一般在 2 mm 左右。当车轮踏面存在缺陷时,其镟修量可以超过 4 mm,因此可以根据头尾车轮镟修量的大小来判断车轮是否存在踏面缺陷^[4]。本文以某铁路局动车组列车的 93 组检修数据为研究对象,将这些列车头尾车轮(共计 5 098 个)的镟修数据组成数据集。数据集内每个数据样本均包括 10 个特征,其中:名义特征包括车号、轴号、轮位及季节,分别考虑了车轮位置和气候条件对车轮踏面缺陷的影响^[5];连续特征包括轮径、轮缘厚、轮缘高、轮径差、累计走行里程及间隔走行里程,分别考虑了车轮几何尺寸、踏面磨耗、踏面偏磨和车辆走行里程对踏面缺陷的影响。

用 (x_i, y_i) 表示第 i 个车轮的数据集,其中 x 为特征(包括名义特征和连续特征), y 为标签值($y=0$ 表示无缺陷, $y=1$ 表示有缺陷)。对于名义特征,其数据是离散和无序的,需要使用 One-hot 编码来处理,每个整数值将被表示为一个二进制向量,除了整数的索引位置取 1 外,其他索引位置均取 0^[6]。使用 One-hot 编码后,离散特征的取值将被扩展到欧式空间,任一离散特征取值均与欧式空间的某个点对应,使得特征之间的距离计算更加合理。此外,One-hot 编码在一定程度上也起到了扩充特征的作用,例如,对于轴号这一特征,使用 One-hot 编码后将形成一个四列的矩阵。对于连续特征,由于各数据具有不同的量纲和数量级,直接使用原始数据进行分析会强调较大数

值的特征在分析中的作用,从而产生学习误差。因此,对于连续特征,需要对其数据进行标准化处理。本文采用 min-max 标准化方法,即通过线性变换将结果映射到 $[0,1]$ 区间,其计算式为:

$$x_{N,ij} = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j} \quad (1)$$

式中:

$x_{N,ij}$ ——经过标准化处理过后的第 i 个车轮的第 j 个特征,为量纲一量;

x_{ij} ——标准化处理前第 i 个车轮的第 j 个特征;

x_j ——所有车轮第 j 个特征的集合。

2 SMOTE-NC 数据集增强

在上文构建的数据集中,车轮踏面缺陷数据仅占全部样本的 5% 左右,正常数据样本占主导地位,因此,该数据集为典型的不平衡数据集。在这种情况下,大部分算法无法完整学习到少数类样本信息,而故障数据一般为少数类样本,这将导致故障的识别精度较低。

对于不平衡数据集,一般从算法和数据两个角度进行改进:①算法层面上,一般通过改进分类算法精度,以提高故障识别率,但通过改进算法层来提高识别率的应用范围较小^[7]。②数据层面上,最基本的方法有欠采样技术和过采样技术。欠采样技术一般通过随机删减部分多数类样本来平衡数据集,但可能造成重要信息丢失;过采样技术采用随机复制少数类样本的方法来平衡数据集,但容易造成数据的过拟合。2002 年,文献[8]提出了 SMOTE(合成少数类过采样技术)算法,该算法能更加科学地进行数据的过采样。本文采用的即是名义特征和连续特征相结合的 SMOTE-NC 算法。

SMOTE-NC 算法首先需要计算少数类样本的欧式距离矩阵,找到距离最接近的前 k 个样本(以下简称“ k 近邻”), k 近邻的欧氏距离 d_{diff} 的计算式为:

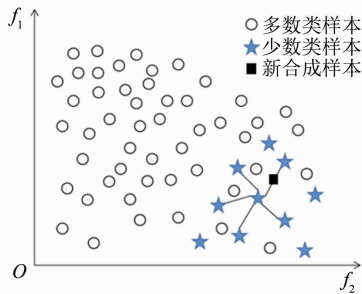
$$d_{\text{diff}} = |d_c - d_{c,k}| = \sqrt{(d_c - d_{c,k})^2} \quad (2)$$

式中:

d_c ——少数类样本的连续特征的欧式距离;

$d_{c,k}$ —— d_c 对应 k 近邻的欧式距离。

根据数据的不平衡情况设置采样倍率 M ,从其 k 近邻中随机选择多个样本,新合成样本的产生原理如图 1 所示。



注: f_1 —特征一的特征值; f_2 —特征二的特征值; 特征一、特征二均泛指数据集的某个特征。

图1 新合成样本产生原理图

Fig. 1 Diagram of new compound sample generation principle

新合成样本的欧式距离 $d_{\text{new},c}$ 的计算式如下:

$$d_{\text{new},c} = d_c + d_{\text{diff}} \text{rand}(0,1) \quad (3)$$

式中:

$\text{rand}(0,1)$ ——用于生成 0~1 间随机数的函数。

对于离散数据,计算其 k 近邻中 z 个样本在该属性上的分布,获取出现次数最多的属性值,并将属性值赋予新合成样本。SMOTE-NC 算法流程图如图 2 所示。

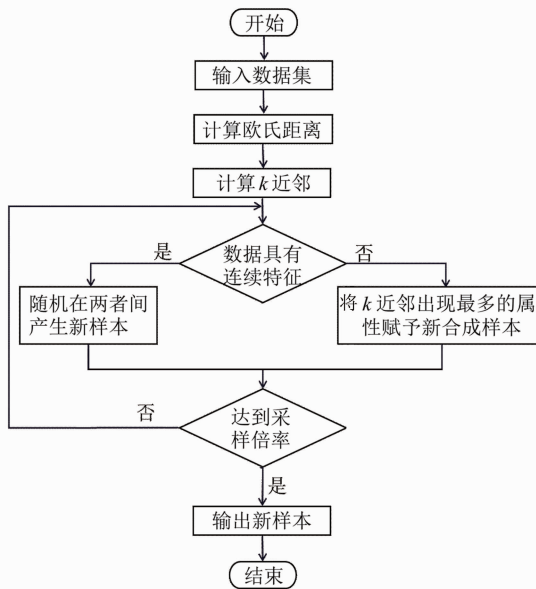


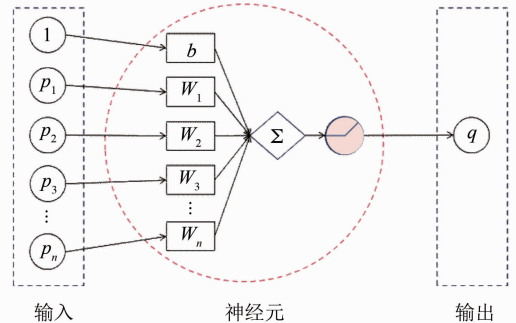
图2 SMOTE-NC 算法流程图

Fig. 2 SMOTE-NC algorithm flowchart

3 建立深度神经网络模型

DNN 与一般神经网络的区别在于其具有多层非线性网络结构,具有更强的学习能力。DNN 可根据网络编码和解码方式分为三类:前馈 DNN、反馈 DNN 及双向 DNN。本文采用误差反向传播的前馈 DNN 模型进行动车组头尾车车轮路面损伤的预测。

神经元(又称为“简单感知器”)是构成神经网络的最小单元,可以模拟人脑处理许多实际问题,其结构如图 3 所示。图 3 中:假定神经元的输入向量为 $\mathbf{P}=[p_1 p_2 \cdots p_n]$ 。每个输入 p_i 均有相应的权值 W_i 。对于每个神经元,还有自带偏置 b 。



注: q —神经元输出。

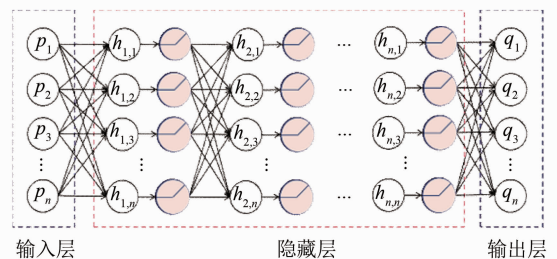
图3 神经元结构

Fig. 3 Neuron structure

假设采用的激活函数为 $\sigma(\cdot)$,则 q 的计算式为:

$$q = \sigma\left(\sum_{i=1}^n W_i p_i + b\right) \quad (4)$$

DNN 是由多个神经元级联而成的网络模型,又被称为多层感知器。DNN 分为输入层、隐藏层及输出层三部分,层与层之间采取全连接的形式,其网络结构如图 4 所示。



注: $h_{s,t}$ —第 s 层的第 t 个神经元; q_n —第 n 个神经元输出。

图4 深度神经网络的基本结构

Fig. 4 Basic structure of DNN

激活函数负责将神经元特征进行非线性映射,是神经网络设计的重要组成部分。常见的激活函数有双曲正切函数 \tanh 、ReLU 函数、sigmoid 函数等^[9],本文采用了 \tanh 和 ReLU 函数。

本文使用的深度神经网络模型中,网络层数为四层,其中:第一层的神经元个数为 64 个,激活函数为 ReLU;第二层的神经元个数为 128 个,激活函数为 ReLU;第三层的神经元个数为 64 个,激活函数为 ReLU;第四层的神经元个数为 16 个,激活函数为 \tanh 。

4 模型优化与结果分析

损失函数的大小可以作为判断 DNN 精度的依据。在模型的训练过程中,损失函数值一般会局部振荡而总体降低,通过优化参数可以降低损失函数的局部振荡并加速其收敛。DNN 中,除连接权值和偏置外,其他参数均为超参数,合理优化算法,进行参数优化,可以得到最优的学习效果。本文所用到的超参数有学习率、训练迭代次数及批次,采用网格搜索法进行参数优化,得到不同参数下训练集的准确率如图 5 所示。由图 5 可直观地发现:训练集的准确率和批次成反比,和迭代次数成正比,其与学习率间则呈现出先增大后减小的变化趋势。

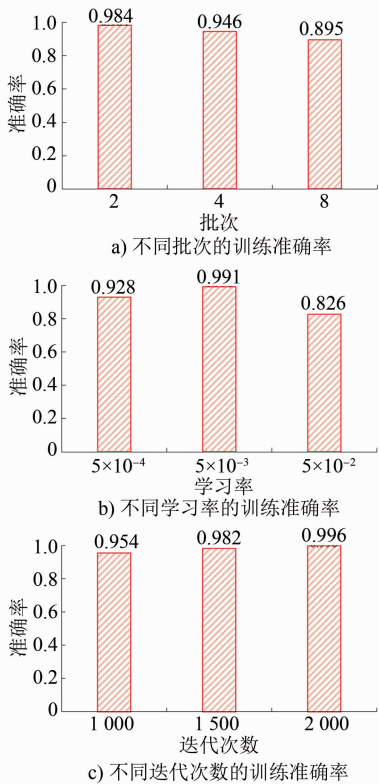


图 5 不同参数对训练集准确率的影响

Fig. 5 Impact of different parameters on training set accuracy

图 6 为损失函数随迭代次数变化图。由图 6 可知:随着迭代次数的增加,损失函数值先急剧下降后趋于平缓;经过 2 000 次迭代计算后,损失函数值下降至 0.02,这表明该模型具有较高的精度和可靠性。

本文选择的超参数如下:批次为 2,学习率为 5×10^{-3} ,迭代次数为 2 000。在当前的试验参数下,训练集的精确率为 99.6%,测试集的精确率为 91.3%。精确率表示测试集中正样本被准确识别为正样本的概率,可以直观反映该模型对故障样本的

识别精度,样本集的精确率达 92.5%,这表明 92.5% 的故障车轮被准确识别。

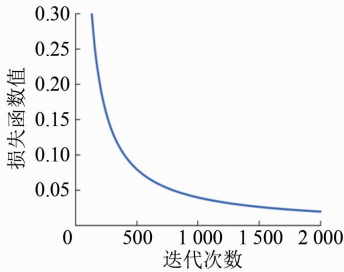


图 6 损失函数随迭代次数变化图

Fig. 6 Graph of loss function changing with the number of iterations

在实际检修工作中,不仅要求所有的故障车轮被准确识别出,还需降低正常车轮的检修频次,以提高检修效率。这就要求模型对两者进行平衡,以取得最优的分类效果。本文选取 ROC(接受者操作特性)曲线作为该性能评估的指标。经计算得到该模型的 ROC 曲线如图 7 所示。图 7 中:横坐标为假正率,代表所有负样本被错误预测为正样本的概率(即假警报率);纵坐标为真正率,代表所有正样本被预测正确的概率(即命中率)。ROC 曲线下的 AUC 值(其值等于 ROC 曲线下方区域的面积)越大,则表明该模型的综合性能越优。图 7 中的 AUC 值为 0.93,这表明本文提出的方法有较高的识别精度和较优的综合性能。

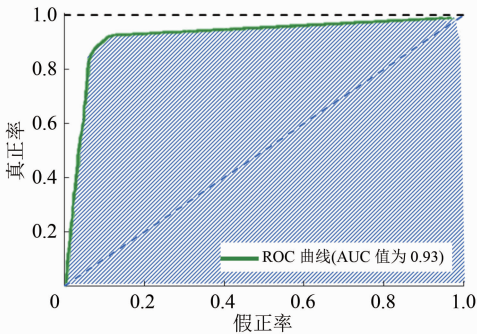


图 7 ROC 曲线图

Fig. 7 ROC curve chart

5 结语

1) 本文提出了基于数据驱动对高速铁路动车组头尾车轮踏面缺陷的预测方法。通过分析车轮的镟修数据,选出 10 个可能对车轮缺陷造成影响特征并组成样本集,将数据分别按名义特征和连续特征进行预处理,通过合成少数类样本的过采样技术对不平衡数据集进行处理,最后构建标准化数

据集。

2) 建立了 DNN 模型,通过网络结构调整和超参数优化,得到模型的最优学习效果。模型的精确率达 92.5%,AUC 值为 0.93,这表明本文提出方法具有较高的预测精度和较优的综合性能,可以有效预测高速铁路动车组头尾车车轮踏面损伤的发生概率。

参考文献

- [1] 敬霖,刘凯. 车轮踏面缺陷引起的轮轨动态响应综述[J]. 交通运输工程学报, 2021, 21(1): 285.
JING Lin, LIU Kai. Review on wheel-rail dynamic responses caused by wheel tread defects[J]. Journal of Traffic and Transportation Engineering, 2021, 21(1): 285.
- [2] 张果. 动车组踏面缺陷的运用分析及控制措施研究[J]. 铁道机车车辆, 2017, 37(6): 103.
ZHANG Guo. Research on wheelset abnormal sound of CRH380AL EMU[J]. Railway Locomotive & Car, 2017, 37(6): 103.
- [3] 顾正隆,葛亦凡. 基于轮对磨耗数据数理统计的轮对镟修策略多目标优化[J]. 城市轨道交通研究, 2021, 24(增刊1): 21.
GU Zhenglong, GE Yifan. Multi-objective optimization of wheelset rotation repair strategy based on mathematical statistics of wheelset wear data[J]. Urban Mass Transit, 2021, 24(S1): 21.
- [4] 马进火,江亚男,陈俊栋,等. 地铁车辆车轮镟修踏面的优化方案[J]. 城市轨道交通研究, 2020, 23(8): 162.
MA Jinhua, JIANG Yanan, CHEN Jundong, et al. Optimization strategy of wheel tread profile lathing for metro vehicles[J]. Urban

Mass Transit, 2020, 23(8): 162.

- [5] 李晓峰,李国栋,宋春元,等. 高速动车组车轮磨耗线路试验研究[J]. 中国铁路, 2019(4): 65.
LI Xiaofeng, LI Guodong, SONG Chunyuan, et al. Wheel wear test on lines for high-speed EMU trains[J]. China Railway, 2019(4): 65.
- [6] HANCOCK J T, KHOSHGOFTAAR T M. Survey on categorical data for neural networks[J]. Journal of Big Data, 2020, 7(1): 1.
- [7] 宋艳,白治江. 基于扩展近邻 SMOTE 过采样的 SVM 分类器[J]. 现代计算机(专业版), 2018(15): 34.
SONG Yan, BAI Zhijiang. SVM classifier based on extended neighbor SMOTE over-sampling[J]. Modern Computer, 2018(15): 34.
- [8] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321.
- [9] 曲之琳,胡晓飞. 基于改进激活函数的卷积神经网络研究[J]. 计算机技术与发展, 2017, 27(12): 77.
QU Zhilin, HU Xiaofei. Research on convolutional neural network based on improved activation function[J]. Computer Technology and Development, 2017, 27(12): 77.

· 收稿日期:2021-09-29 修回日期:2021-12-30 出版日期:2024-04-10
Received:2021-09-29 Revised:2021-12-30 Published:2024-04-10
· 第一作者:王文琦,硕士研究生,1402778186@qq.com
通信作者:宋冬利,高级实验师,sdl.cds@163.com
· ©《城市轨道交通研究》杂志社,开放获取 CC BY-NC-ND 协议
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license

(上接第 27 页)

- [8] 陆佳. 时速 160 km 刚性悬挂接触网的可行性分析[J]. 科技创新与应用, 2020(13): 65.
LU Jia. Feasibility analysis of rigid suspension catenary with speed of 160 km/h[J]. Technology Innovation and Application, 2020(13): 65.
- [9] ZHOU N, ZHANG W. Investigation on dynamic performance and parameter optimization design of pantograph and catenary system[J]. Finite Elements in Analysis and Design, 2011, 47(3): 288.
- [10] 周宁,邹欢,邹栋,等. 城市轨道交通弓网系统仿真模型适应性研究[J]. 西南交通大学学报, 2017, 52(2): 408.
ZHOU Ning, ZOU Huan, ZOU Dong, et al. Investigation on the applicability of pantograph and catenary model urban railway system[J]. Journal of Southwest Jiaotong University, 2017, 52(2): 408.
- [11] 黄安宁,易智民,张安全. 基于 SIMPACK 的轨道车辆动态

性能仿真分析[J]. 机械与电子, 2013, 31(3): 77.

- HUANG Anning, YI Zhimin, ZHANG Anquan. Simulation research on dynamic performances of railway vehicles based on SIMPACK[J]. Machinery & Electronics, 2013, 31(3): 77.
- [12] LEE J H, PARK T W, OH H K, et al. Analysis of dynamic interaction between catenary and pantograph with experimental verification and performance evaluation in new high-speed line[J]. Vehicle System Dynamics, 2015, 53(8): 1117.

· 收稿日期:2021-09-22 修回日期:2021-11-17 出版日期:2024-04-10
Received:2021-09-22 Revised:2021-11-17 Published:2024-04-10
· 第一作者:董晓,工程师,xiaoxiaoever11@163.com
通信作者:周宁,副研究员,zhouning@swjtu.edu.cn
· ©《城市轨道交通研究》杂志社,开放获取 CC BY-NC-ND 协议
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license