

大数据技术在动车组故障预警中的应用

吴臻易

(中车青岛四方机车车辆股份有限公司, 山东 青岛 266111)

摘要: 为提前识别动车组运行过程中的故障隐患, 文章基于故障预测及健康管理理论, 通过整合动车组不同场景的多源异构数据, 构建“车-地”一体化大数据平台; 并将设备故障机理与人工智能算法相结合, 构建一种动车组关键部件故障预警预测模型, 以部件关键物理特性来反映其工作状态并提前识别潜在故障。通过大数据平台及牵引电机的故障预警及温度预测模型的应用, CR400AF 型复兴号动车组牵引电机故障率显著下降, 由平均每百万公里 0.5 件降至约平均每百万公里 0.1 件。

关键词: 动车组; 大数据; 故障预警; 数字仿真; 神经网络

中图分类号: U269.2

文献标识码: A

文章编号: 2096-5427(2021)05-0091-06

doi:10.13889/j.issn.2096-5427.2021.05.300

Application of Big Data Technology in the EMU Fault Early Warning

WU Zhenyi

(CRRC Qingdao Sifang Co., Ltd., Qingdao, Shandong 266111, China)

Abstract: In order to identify potential fault hazards in the operation of EMU in advance, based on the theory of prognostics health management, this paper constructs a vehicle-ground integrated big data platform by integrating multi-source heterogeneous data of different scenarios of the EMU. It uses a method of combining equipment fault mechanism and artificial intelligence algorithm to construct a fault early warning and prediction model of key components of EMU, which reflects the working status of a component with its key physical properties and identify potential faults in advance. Through the application of the big data platform and the traction motor fault early warning and temperature prediction model, traction motor fault rate of CR400AF Fuxing EMU decreased significantly from 0.5 pieces/ 10^6 km to about 0.1 pieces/ 10^6 km.

Keywords: EMU; big data; fault warning; digital simulation; neural network

0 引言

随着互联网、大数据技术的不断发展应用, 智能化目前已成为动车组发展的必然趋势之一^[1]。近年来, 轨道交通行业已初步实现了车辆的远程监控, 运维人员可在地面实时获取动车组的故障信息, 提高了行车故障应急处置效率。但在故障预警方面, 目前的主要

措施仍是各项参数的阈值预警(如最大值、最大差值、差值变化率等), 所面临的问题一是无线传输装置(wireless transport device, WTD)车地通信数据密度高, 海量数据汇集到地面服务器进行运算, 致使地面服务器不堪重负; 二是计算机领域的人工智能算法与传统的动车组故障检测方法之间缺少切实可行的结合点, 应用困难。动车组自身是一个极为复杂的系统, 其运行状态会受外部环境的影响, 在实际的故障预警研究中, 目前尚无法获取全部关联数据。因此, 探讨一种现阶段可行的大数据平台构建及故障预警预测实现方法, 对于进一步提高动车组运营可靠性、保障运输

收稿日期: 2021-04-27

作者简介: 吴臻易(1985—), 男, 高级工程师, 长期从事动车组故障诊断及预防技术研究工作。

基金项目: 山东省重点研发计划(国际科技合作)(2019GHZ004)

效率具有必要性及迫切的现实意义。

本文从“车-地”一体化大数据平台的构建出发，分别阐述了车载和地面故障预测与健康管理（prognostics health management, PHM）平台的技术框架，同时提出了通过数字仿真思路建立动车组关键部件预警预测模型的方法，并以牵引电机温度预警预测模型的建立为例进行了详细说明。

1 “车-地”一体化大数据平台构建

“车-地”一体化大数据平台采用分布式网络思路构建，其列车端在基于4G移动网络的传统WTD数据车地传输的基础上增加了车载PHM系统。这样不仅可以很好地弥补地面PHM平台数据密度低、实时性不足的缺陷，而且对车载状态数据进行分散预处理，可以更加高效地进行早期故障预警。其总体设计方案见图1。

1.1 车载PHM系统

车载PHM系统实时采集车辆数据，并通过搭载的车载故障预警模型提前识别车辆异常状态。该系统对车辆实时数据进行分散预处理，将实时性要求高、计算量小的预警模型放在车辆端完成，并将运算结果传输至地面服务器，从而提升车辆数据实时运算效率^[2]。车载PHM系统由部件级PHM和列车级PHM组成。通过在牵引、制动、轴温等系统的控

制主机内增加部件级PHM单元，对所采集的内部数据进行故障预警、预测及系统状态自诊断，从而完成部件级PHM分析；并利用列车网络将部件的原始数据、状态信息以及故障信息发送给车载PHM主机，从而完成整车自诊断。列车智能显示屏显示诊断信息，并将其通过WTD传输至地面。车载PHM系统架构示意图图2。

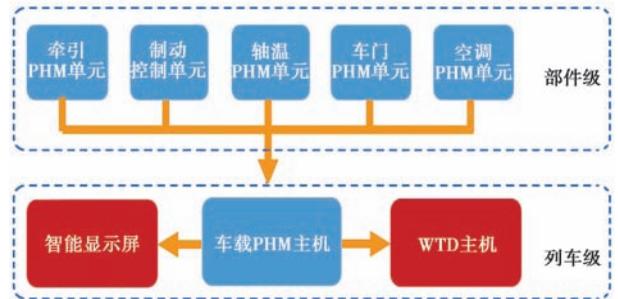


图2 车载PHM系统架构示意图
Fig. 2 Schematic diagram of onboard PHM system

以走行部安全监测为例，其车载PHM系统架构如图3所示。独立的车载PHM系统与列车网络共享走行部安全监测系统数据，对轴温、平稳性、失稳、振动特性等部件级数据及整车级数据进行实时运算分析，并将处理后的预警信息和状态特征传输给列车网络主处理单元，然后统一传输至列车WTD，最终通过移动通信网络发送给地面接收服务器。

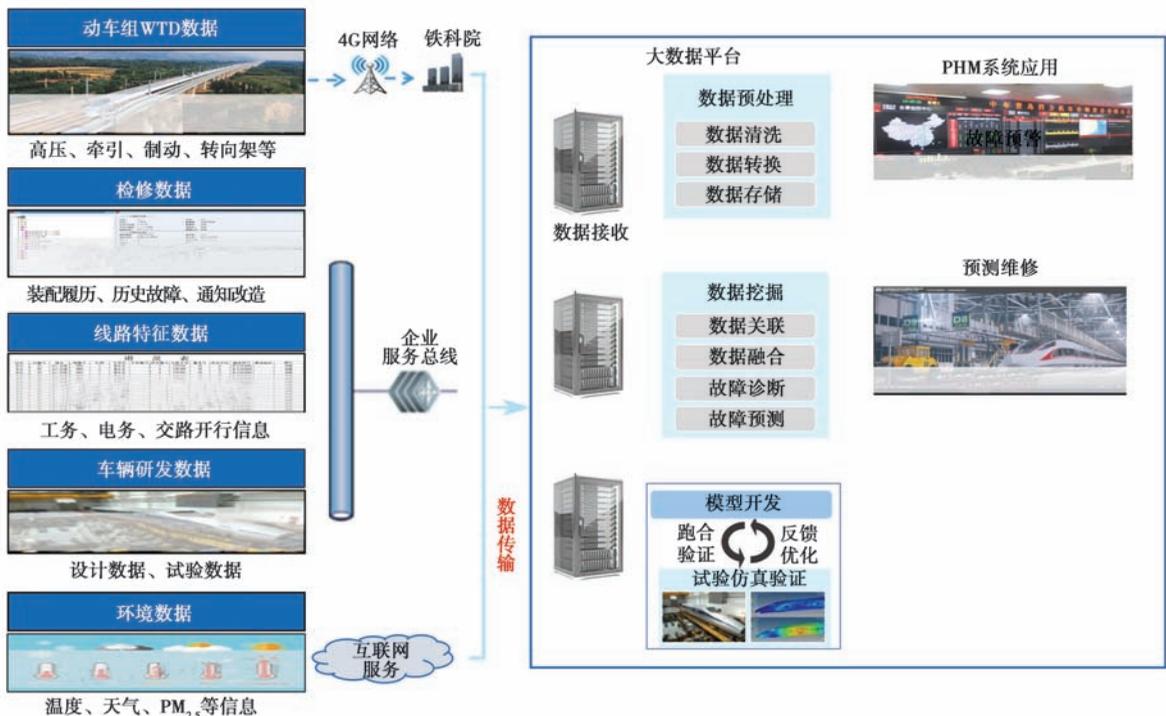


图1 “车-地”一体化大数据平台总体架构
Fig. 1 Overall architecture of the vehicle-ground integration big data platform

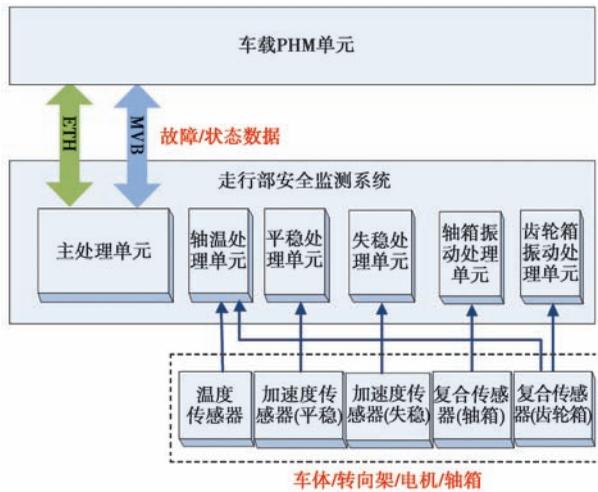


图3 走行部车载PHM架构

Fig. 3 Onboard PHM architecture of running gear

1.2 地面PHM平台

地面PHM平台架构按照云计算的3层框架建设,如图4所示,其包含基础设施层服务(IaaS层)、平台层服务(PaaS层)和软件层服务(SaaS层)3部分内容^[3]。

1.2.1 基础设施层服务

基础设施主要包含计算资源、存储资源及网络资源等硬件设施。计算资源包括虚拟服务器、数据库服务器和数据处理服务器等。存储资源主要为物理的服务器硬盘,负责数据的统一管理存放,保证数据的高可靠性、可用性和冗余性。网络资源主要为高速、大容量和安全稳定运行的企业私有云。

1.2.2 平台层服务

故障预警需要尽可能地贯通动车组全寿命周期的数据,实现不同形式数据间的融合和调用。传统的数据处理方式及系统架构很难满足海量数据的快速处理及运算分析,而“Hadoop+Spark”云平台架构具有跨集群、分布式处理庞大数据的功能,是现阶段组建大数据服务平台的首选^[4]。多源异构数据根据周期、来源、格式分类接入云平台,经过数据清洗、解析后存入到分布式文件系统(Hadoop distributed file system),然后分别通过并行处理框架(MapReduce)及Spark计算引擎对静态数据及实时流数据进行处理,以确保数据服务高效、可靠^[5]。

1.2.3 软件层服务

软件层服务提供车辆运行监控、故障预警预测、维修作业预测及健康评估等功能应用,其将大数据、互联网和人工智能等新技术与动车组运维检修相融合,以充分挖掘动车组全生命周期数据所蕴含的价值,是面向动车组运维人员实现故障预警、维修预测及寿命评估等功能应用的门户。

2 动车组关键部件故障预警模型

动车组故障预警需要基于大数据平台构建的针对具体部件的数据模型来实现。目前行业内的数据模型从技术类型上可以分为两类:一类是基于部件设计原理的机理模型,一类是基于人工智能的算法模型。

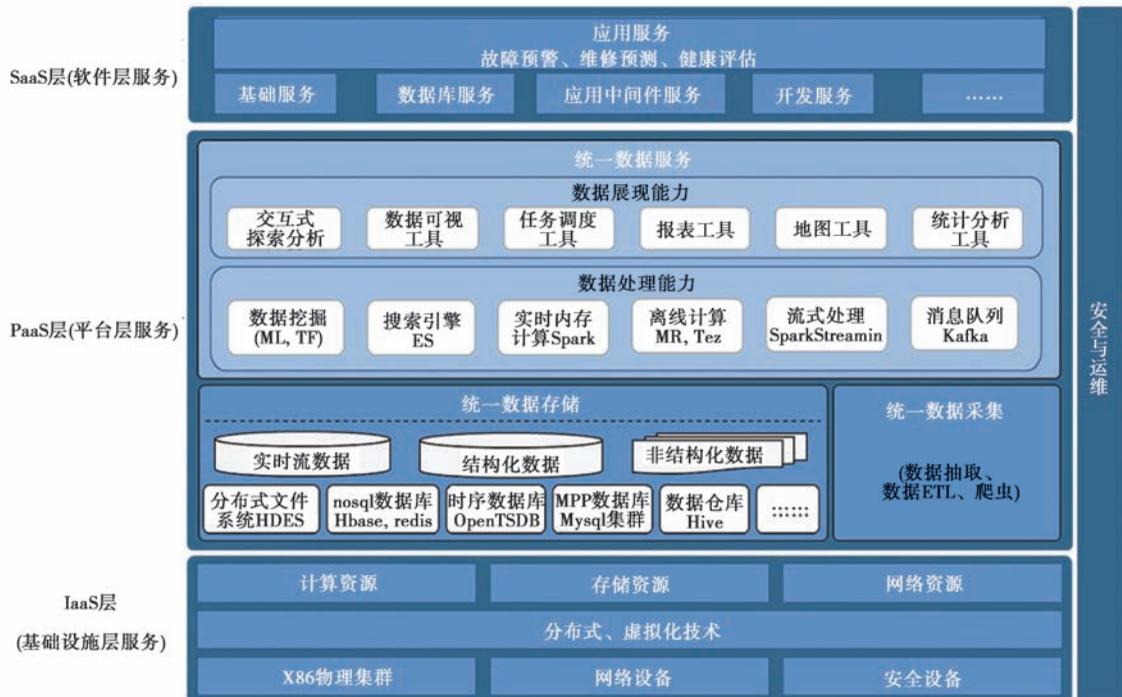


图4 地面PHM平台架构

Fig. 4 Architecture of ground PHM platform

而通过这两种技术方法的相互结合,可以进一步构建数字仿真模型,从而以部件关键物理特性来反映其工作状态并提前识别潜在故障。

2.1 数字仿真模型构建思路

目前,新型动车组在设计阶段针对部分系统或部件实现了数字样机的构建,以数字化方式表达产品的设计信息,从而对其尺寸、功能、特性等进行模拟仿真。在此基础上,我们可以将数字样机的概念进一步拓展,在动车组的服役阶段以数字化方式创建产品物理实体的数字仿真模型,对其在故障发生前出现的异常趋势进行预测^[6]。其思路是以产品设计参数、运行状态数据、造修履历数据及运用维护数据为基础,并结合环境、线路等全寿命关联数据,从部件设计原理及故障影响因素出发,确定反映部件工作状态的核心数据项点及其关联数据范围,然后通过数据拟合方法分析和验证关联数据的有效性,并选用合适的人工智能算法对故障样本及历史数据进行训练,最终建立虚拟化数字仿真模型。在此基础上,需要运用机理分析方法对模型的各项数据权重进行调优,使模型效果符合预期;通过对核心数据项点的监测,预知部件工作状态及演变趋势,从而实现部件故障预警及趋势预测。动车组关键部件数字仿真模型构建路线见图5。

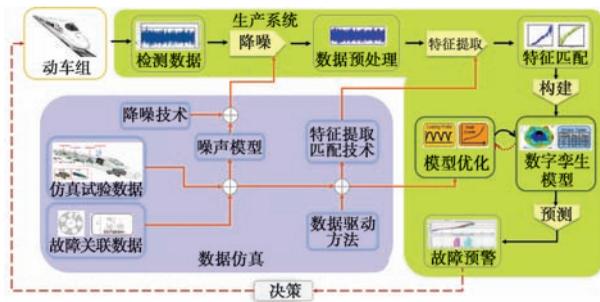


图5 动车组关键部件数字仿真模型构建路线图
Fig. 5 Digital simulation model construction roadmap of EMU key components

现实条件下,无法获取动车组服役期间所涉及的全部数据,甚至单个部件的全部数据也几乎不可能获取得到。因此,目前我们需要有选择地针对特定部件、特定物理维度来实现有限的数字仿真。

2.2 牵引电机的故障预警及温度预测模型

牵引电机是动车组动力系统的核心部件,其在工作会产生热损耗,若因内阻变化、散热不良等原因导致电机温度急剧升高,超过其所能承受的最高温度,则会造成设备故障,导致列车晚点或停车。本文以CR400AF型复兴号高速动车组牵引电机为例,来

说明动车组关键部件数字仿真模型的构建思路。

2.2.1 建立实时温度预警子模型

首先从牵引电机工作原理出发,梳理分析电机状态核心判定指标——电机定子温度及其影响关联因素,定义特征变量范围。判定指标既要涵盖电机电流、列车速度、牵引级位、天气情况(温度、雨雪)等与温度直接相关的特征变量数,也要包括间接影响变量,例如线路区段的坡道、曲线半径,其会影响电机的牵引功率,从而影响电机温度。

然后,采用抽样方式提取每列动车组、每自然月中两天的运行数据,提取时需尽量覆盖不同线路。将抽样数据按照时间顺序进行排列并隐藏列号和车厢信息。随机抽取1 000组数据,通过SPSS软件对每个变量与电机定子温度的相关性进行Pearson分析验证。选取相关度大于0.2、显著性(双侧)小于0.01的数据指标来构建模型,符合条件的数据指标见表1。

表1 电机温度关联数据相关性分析
Tab. 1 Correlation analysis of motor temperature linked data

电机定子温度	速度	转子频率	直流电压	档位	恒速状态	环境温度	线路坡度	电机电流
Pearson相关性	0.584	0.584	0.494	0.362	0.362	0.275	0.238	0.207
显著性(双侧)	0	0	0	0	0	0	0	0

注:两个变量的相关性在-1~1,趋向-1为负相关,趋向0不相关,趋向1为正相关;显著性小于0.05为显著相关,小于0.001显示为0。

再者,筛选出CR400AF复兴号动车组7起牵引电机历史故障信息,其均为因电机绕组绝缘不良而导致的接地常见故障。根据设计原理及数据分析可知,这类故障发生前后均有电机温度的较大变化。选取故障发生前0.5 h内的电机温度及表1的关联数据作为故障样本,同时选取50列动车组共计100 h正常历史数据作为正常样本,与故障样本组成样本库。当前数据采集周期为10 s,所以样本库共包含37 260组数据,每6组数据记为一个样本,共生成6 210个样本,其中正常样本6 000个,故障样本210个。将样本数据随机打乱并按照7:3的比例划分为训练集和测试集。训练集共有4 347个样本,其中正常样本4 200个,故障样本147个;测试集为1 863个样本,其中正常样本1 800个,故障样本63个。采用BP神经网络学习算法建立1个3层的牵引电机温度预警神经网络模型^[7],输入神经元包括表1中的数据指标变量,模型的输出神经元为牵引电机温度预警信息,

模型示意如图 6 所示。

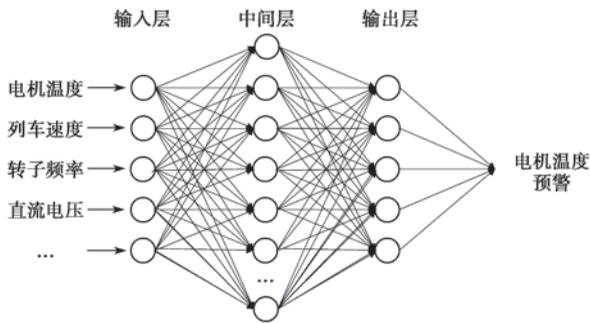


图 6 牵引电机温度神经网络模型示意图

Fig. 6 Schematic diagram of traction motor temperature neural network model

最后，使用训练集样本数据以 0.1 的学习率进行模型训练，目标相对误差为 1×10^{-2} ，最大迭代数为 5 000 次。训练完毕后，利用测试集数据进行泛化能力检测，当泛化能力不再提高时，终止训练，完成基于神经网络算法的牵引电机温度实时预警模型的构建。模型测试集预测结果如表 2 所示，模型准确率为 99.14%，召回率为 93.65%。

表 2 模型测试集预测结果

Tab. 2 Prediction results model test set

预测指标	预测结果
TP (将正类预测为正类数)	59
FN (将正类预测为负类数)	4
FP (将负类预测为正类数)	12
TN (将负类预测为负类数)	1 788
准确率 (accuracy) = (TP+TN)/(TP+FN+FP+TN)	99.14%
精确率 (precision) = TP/(TP+FP)	83.10%
召回率 (recall) = TP/(TP+FN)	93.65%

2.2.2 建立温度趋势预测子模型

同一车次的发车和到站时间以及线路顶棚速度完全相同，所以动车组在同一区间的自身运行情况、线路坡度等影响牵引电机温度的因素基本一致，仅有环境温度属于不稳定变量，其可通过沿线天气预报数据获取。基于此，我们可以实现牵引电机温度预测。

首先，从最近一年的 CR400AF 复兴号动车组正常运行数据中按照不同季节、不同线路对表 1 中 8 项数据指标变量进行抽样，对牵引电机定子温度进行回归分析，以此构建电机温度预测模型^[8]。

然后，通过当天开行计划、实时 GPS 定位信息与历史数据的关联来获取该列动车组在后续区间的最近一次正常运行状态数据，包括速度、转子频率、直流电压、挡位、恒速状态、线路坡度及电机电流，并通过互联网沿线天气预报数据获取环境温度。所有数据关联获取周期为 30 min，通过温度预测模型预

测各牵引电机正常工作状态下后续 30 min 的温度变化趋势。

为反映电机温度预测情况，用某列车组某天的实际电机温度与当天预测的电机温度进行对比验证，结果见图 7。从数据分析结果来看，动车组行驶工况相对简单（如处于恒速、恒挡位状态）时预测效果很好；动车组行驶工况复杂（如反复调速、牵引制动切换频繁）时预测偏差有所增大。导致此种情况的原因是，复杂工况下预测模型相关挡位、电机电流等指标变化剧烈，而 WTD 车地数据传输周期较长（10 s），因此无法完全记录其变化过程。

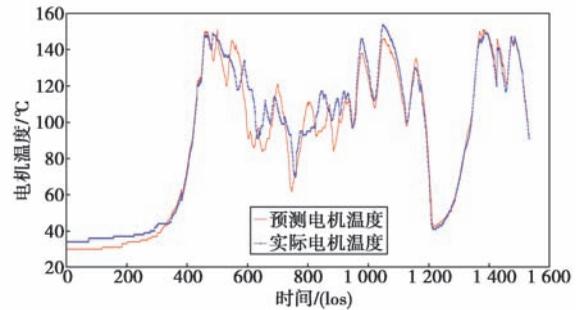


图 7 电机实际温度与模型预测温度对比

Fig. 7 Comparison between actual motor temperature and model predicted temperature

为进一步评价预测效果，将模型预测值与实际值进行拟合分析，见图 8。可以看出，实际温度与预测温度趋势一致性较好，拟合系数 R2 值为 0.97，基本符合预期要求。

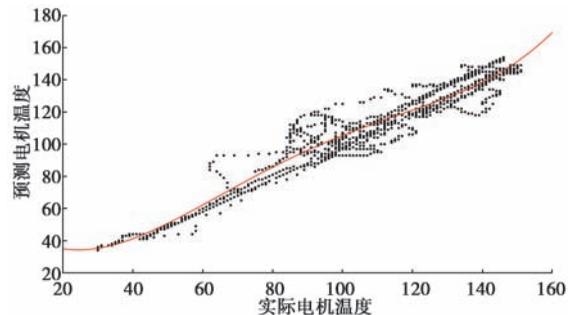


图 8 牵引电机定子温度预测子模型的预测值与实际值拟合分析

Fig. 8 Fitting analysis between predicted values and actual values of traction motor stator temperature prediction sub model

至此，牵引电机定子温度预测的数字仿真模型完成搭建。其功能共包括两部分：一是来判定牵引电机工作状态正常与否，以识别异常并提醒应急指挥人员及时介入；二是获取电机正常工作状态下后 30 min 内的温度变化趋势信息，并与电机故障后的实时温度进行比对，使应急指挥人员能较为直观地了解电机异常程度，为后续的分析决策提供数据支撑。

2.2.3 模型应用

2017年CR400AF复兴号动车组上线初期,其牵引电机接地故障率较高,约每百万公里1.4件;2019年2月采取绝缘测试、改进绝缘工艺等措施加强故障预防后,故障率逐渐下降,与同期整车平均水平持平,约每百万公里0.5件。牵引电机数字仿真模型于2020年6月份上线应用后,累计预防牵引电潜在机故障40余起,使牵引电机故障率下降至每百万公里0.1件左右,显著低于以往整车平均水平,详见图9。

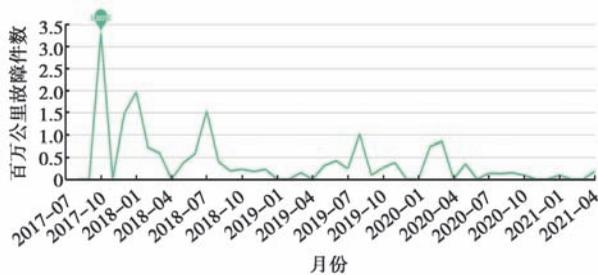


图9 CR400AF型动车组牵引电机故障率
Fig. 9 Failure rate of traction motor for CR400AF EMU

3 结语

本文从动车组故障预警需求出发,提出了一种车-地数据分散处理、分布式大数据平台构建方案;运用基于部件故障机理与神经网络算法相结合的方法,探索构建动车组关键部件故障预警模型,并以CR400AF型复兴号动车组牵引电机的故障预警及温度预测模型构建及应用为例,进行牵引电机故障预警。应用结果显示,采用该预警模型后,牵引电机故障率从每百万公里0.5件下降至每百万公里0.1件。可见,该模型的应用提高了应急处置效率。

由于目前动车组制造、运维及检修的全场景数据尚未全部贯通,故障预警、预测的研究与应用仍存在诸多限制。因此,需要通过不断积累现有数据应用经验,提出多源异构数据研究需求,推动动车组全生命周期数据贯通,逐步优化、完善现有的数据采集、传输方式,进而通过研究成果的推广落地引导动车组

生产制造、故障诊断、运维检修等环节现有机制的不断变革。

参考文献:

- [1] PAULA F L, TIAGO M F C, LUIS C. Towards the internet of smart trains: a review on industrial IoT-Connected railways[J]. Sensor, 2017, 17(6): 1457-1501.
- [2] 梁建英. 高速列车智能诊断与故障预测技术研究[J]. 北京交通大学学报, 2019, 43(1): 63-70.
LIANG J Y. Research on intelligent diagnosis and fault prediction technology for high speed trains[J]. Journal of Beijing Jiaotong University, 2019, 43(1): 63-70.
- [3] 梁建英, 刘韶庆, 范龙庆, 等. 大数据在我国高速动车组运维中的应用[J]. 控制与信息技术, 2019(1): 7-11.
LIANG J Y, LIU S Q, FAN L Q, et al. Application of Big Data Technology in the Operation and Maintenance for High Speed EMUs[J]. Control and Information Technology, 2019(1): 7-11.
- [4] 陈吉荣, 乐嘉锦. 基于Hadoop生态系统的大数据解决方案综述[J]. 计算机工程与科学, 2013, 35(10): 25-35.
CHEN J R, LE J J. Reviewing the big data solution based on Hadoop ecosystem[J]. Computer Engineering & Science, 2013, 35(10): 25-35.
- [5] 赵珂, 王伟, 姜喜民, 等. 分布式高铁动车组PHM大数据架构设计与实现[J]. 软件, 2018, 39(12): 90-94.
ZHAO K, WANG W, JIANG X M, et al. Design and Implementation of PHM Big Data Architecture for Distributed High-speed Rail EMU[J]. Computer Engineering & Software, 2018, 39(12): 90-94.
- [6] 雷亚国, 何正嘉, 瞿艳阳. 基于混合智能新模型的故障诊断[J]. 机械工程学报, 2008, 44(7): 112-117.
LEI Y G, HE Z J, ZI Y Y. Fault Diagnosis Based on Novel Hybrid Intelligent Model[J]. Chinese Journal of Mechanical Engineering, 2008, 44(7): 112-117.
- [7] 王黛, 马卫武, 李立清, 等. 基于BP神经网络的高速动车组牵引能耗计算模型[J]. 中南大学学报(自然科学版), 2017, 48(4): 1104-1110.
WANG D, MA W W, LI L Q, et al. Estimating traction energy consumption of high-speed trains based on BP neural network[J]. Journal of Central South University (Science and Technology), 2017, 48(4): 1104-1110.
- [8] 王惠文, 孟洁. 多元线性回归的预测建模方法[J]. 北京航空航天大学学报, 2007, 33(4): 500-504.
WANG H W, MENG J. Predictive modeling on multivariate linear regression[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(4): 500-504.