

# 基于大数据的新能源汽车电池管理系统故障诊断研究

赵磊

湖州耀宁固态电池研究院有限公司宁波分公司 浙江 宁波 315336

**摘要:** 本文聚焦基于大数据的新能源汽车电池管理系统故障诊断研究。阐述新能源汽车电池管理系统(BMS)与大数据技术基础,介绍BMS功能及大数据“5V”特征。详细说明数据采集与预处理方案,探讨故障特征提取与选择方法。分析故障诊断模型选择、训练优化及评估指标。通过研究,旨在利用大数据提升BMS故障诊断能力,保障新能源汽车动力电池安全稳定运行,推动产业高质量发展。

**关键词:** 新能源汽车; 电池管理系统; 大数据; 故障诊断

引言: 新能源汽车发展迅猛, 电池管理系统(BMS)作为保障动力电池安全、稳定、高效运行的核心, 其故障诊断至关重要。传统故障诊断方法存在局限性, 难以满足需求。大数据技术具备海量、高速、异构数据处理能力, 与BMS数据需求高度契合。利用大数据技术对BMS数据进行深度挖掘与分析, 可实现更精准的故障诊断, 为新能源汽车智能化运维提供有力支持, 因此开展相关研究具有重要现实意义。

## 1 新能源汽车电池管理系统与大数据技术基础

### 1.1 新能源汽车电池管理系统概述

新能源汽车电池管理系统(BMS)是保障动力电池安全、稳定、高效运行的核心控制系统, 贯穿于电池充放电、存储、运维全生命周期。其核心功能涵盖电池状态监测、参数调节、故障预警、能量优化分配等, 直接影响车辆续航里程、动力性能及电池使用寿命。当前主流动力电池以锂离子电池为主, BMS需实时采集单体电池电压、电流、温度及总电压等关键参数, 通过算法精准估算SOC(剩余电量)、SOH(健康状态)、SOP(功率状态)三大核心指标。随着新能源汽车向智能化、网联化升级, BMS已从传统独立控制单元进化为多模块协同的智能系统, 融合了硬件采集、软件算法与整车通讯技术<sup>[1]</sup>。同时, BMS需应对低温续航衰减、高温热失控、电池一致性差等行业痛点, 通过动态调节充放电策略、均衡控制单体电池状态, 降低故障风险。在此基础上, BMS已实现与车载智能系统、云端平台的深度联动, 为动力电池全生命周期管理提供精准数据支撑, 已然成为新能源汽车产业高质量发展的重要基石。

### 1.2 大数据技术基础

大数据技术是支撑海量、高速、异构数据高效处理与价值挖掘的核心技术体系, 主要涵盖数据采集、存储、处理、分析与可视化五大核心模块, 其“5V”特征

(Volume、Velocity、Variety、Value、Veracity)与新能源汽车BMS的数据需求高度契合。在数据存储层面, 分布式存储技术(如Hadoop HDFS、HBase)可实现TB级甚至PB级BMS数据的安全存储与高效访问, 解决传统集中式存储的容量瓶颈与并发压力。数据处理环节, 批处理技术(如MapReduce)适用于离线数据分析, 流处理技术(如Spark Streaming)可满足BMS实时数据的即时处理需求, 保障故障预警的时效性。数据挖掘算法(如聚类、分类、回归)则是提取数据价值的关键, 能从海量BMS数据中挖掘电池状态变化规律与故障关联特征。大数据技术还依托云计算平台实现资源弹性扩展, 降低BMS数据处理的硬件成本, 结合人工智能技术构建预测模型, 为电池故障诊断、寿命预测提供技术支撑, 推动BMS从“被动监测”向“主动预判”转型, 赋能新能源汽车智能化运维。

## 2 基于大数据的电池管理系统数据采集与预处理

### 2.1 数据采集方案

基于大数据的BMS数据采集方案以“全维度、高精度、实时性”为核心目标, 构建多源异构数据采集体系, 覆盖电池本体、车载终端、云端平台三个层级。硬件层面, 通过电池包内置的电压传感器、电流传感器、温度传感器实现单体及整包电池参数采集, 确保数据精准度达毫伏级、毫安级。同时通过车载T-Box终端采集车辆行驶工况(车速、加速度、制动状态)、环境参数(外界温度、湿度、气压)及充电数据(充电电流、充电功率、充电时长), 实现跨模块数据融合<sup>[2]</sup>。采用CAN总线、以太网结合5G技术构建传输网络, CAN总线保障车载内部数据的稳定传输, 以太网与5G实现车载数据向云端平台的高速上传, 解决海量数据传输的延迟问题。数据采集策略采用“常态采样+触发采样”结合模式, 可在一定程度上满足常规数据采集需求: 常态下按固定频率采集基础

数据,当检测到参数异常(如温度骤升、电压波动)时自动提高采样频率,精准捕捉故障发生前后的数据变化。

## 2.2 数据预处理方法

数据预处理是提升BMS数据质量、保障后续故障诊断模型准确性的关键环节,主要针对采集数据中的缺失值、异常值、冗余值及异构数据进行处理,核心目标是将原始数据转化为标准化、可用的数据集。针对缺失值,采用分层处理策略:对于随机缺失且缺失率低于5%的数据,若涉及电压、温度等符合渐变规律的物理参数,优先采用动态均值或中位数填充,避免简单均值掩盖局部趋势;对于连续缺失或缺失率5%~20%的数据,若数据含电流非线性变化特征,可采用分段线性插值或三次样条插值法,保留工况突变时的数据波动特征,其他类型参数可根据特性选择适配插值方式;对于缺失率高于20%的数据,无论参数类型,均结合电池运行工况(如充放电阶段)采用基于机器学习的预测填充法(如随机森林填充),通过历史工况模式匹配实现数据重构,避免特征失真。异常值处理采用“检测-分析-修正/保留”流程,通过 $3\sigma$ 准则、箱线图法检测超出合理范围的异常数据,结合电池实际运行场景判断异常类型,对于传感器故障导致的异常值直接剔除,对于工况突变(如急加速)导致的异常值采用滑动窗口中位数滤波修正,保留故障特征的同时抑制噪声。冗余值处理需平衡数据降维与故障诊断需求,通过特征相关性分析(如皮尔逊相关系数)识别高度相关数据(如总电压与单体电压总和),但保留与故障模式强相关的冗余特征(如电压标准差与极差),仅剔除对诊断无贡献的冗余数据,在降低计算成本的同时保留故障诊断关键信息。最后,进行数据标准化与归一化处理,采用Z-Score标准化、Min-Max归一化方法将不同量纲的参数(电压、电流、温度)转化为统一范围,消除量纲差异对模型训练的影响,为后续特征提取与模型构建奠定基础。

## 3 电池故障特征提取与选择

### 3.1 故障特征提取方法

电池故障特征提取是从预处理后的BMS数据中挖掘与电池故障(如容量衰减、热失控、单体短路)相关的特征参数,构建故障特征集的过程,需结合时域、频域及非线性特征提取方法,全面捕捉故障隐含信息。时域特征提取是最基础的方法,通过统计分析提取数据的时域统计量,包括均值、方差、峰值、谷值、峭度、偏度等,可有效反映电池电压波动、电流稳定性、温度变化趋势等故障关联特征,如电池短路时会出现电压骤降、电流骤升,其峰值与方差特征会显著异常。频域特征提

取通过傅里叶变换、小波变换将时域数据转化为频域数据,提取频率分布、幅值、能量等特征,适用于捕捉电池故障初期的微弱信号变化,如电池内阻增大导致的电压频率特性改变,可通过小波变换分解为不同频段信号,提取故障特征频段的能量占比。非线性特征提取针对电池系统的非线性特性,采用分形维数、熵值(如近似熵、样本熵)等指标,反映电池数据的复杂性与不规则性,如电池老化过程中,电压时间序列的熵值会呈现规律性变化,可作为容量衰减故障的特征参数<sup>[3]</sup>。

### 3.2 故障特征选择算法

故障特征选择算法的核心目标是从多维度特征集中剔除冗余特征、噪声特征,筛选出对电池故障诊断贡献度高的关键特征,降低模型计算复杂度、提升诊断精度与泛化能力,主要分为过滤式、包裹式、嵌入式三类算法。过滤式算法基于特征与故障标签的相关性进行筛选,不依赖具体诊断模型,具有计算效率高的优势,常用方法包括皮尔逊相关系数法、互信息法、卡方检验法,通过量化特征与故障类型的关联程度,设定阈值筛选关键特征,适用于特征维度较高的初步筛选。包裹式算法将特征选择与诊断模型性能绑定,以模型预测精度为评价标准筛选最优特征子集,常用方法有遗传算法、粒子群优化算法、逐步回归法,能精准筛选出适配模型的特征,但计算量较大,适用于特征维度精简后的二次筛选。嵌入式算法将特征选择融入模型训练过程,通过模型自身的参数学习自动筛选关键特征,兼顾筛选效果与计算效率,常用方法有L1正则化(Lasso)、决策树特征重要性排序、随机森林特征权重法,其中L1正则化可通过惩罚系数压缩冗余特征的权重至零,实现特征选择与模型训练同步进行。实际应用中,常采用“过滤式+嵌入式”组合策略,先通过过滤式算法初步降维,再通过嵌入式算法精准筛选,平衡筛选效果与计算成本,构建最优故障特征子集。

## 4 基于大数据的电池管理系统故障诊断模型构建

### 4.1 故障诊断模型选择

基于大数据的BMS故障诊断模型选择需结合数据特征、故障类型及诊断需求,兼顾模型精度、实时性与泛化能力,主流模型分为传统机器学习模型与深度学习模型两大类。传统机器学习模型适用于特征工程完善的中小规模数据集,具有模型结构简单、训练速度快、可解释性强的优势,常用模型包括支持向量机(SVM)、逻辑回归、决策树、随机森林。其中,SVM通过核函数映射将数据转化至高维空间,解决非线性故障分类问题,适用于电池轻微故障诊断;随机森林通过多棵决策树集

成学习,降低单一模型的过拟合风险,能有效处理BMS数据中的噪声,适用于多类型故障(如容量衰减、热失控)的综合诊断。深度学习模型适用于海量高维BMS数据,可自动学习深层故障特征,无需复杂人工特征工程,常用模型包括卷积神经网络(CNN)、循环神经网络(RNN)、长短时记忆网络(LSTM)。CNN擅长提取数据的局部特征,可捕捉电池故障的瞬时特征变化;LSTM能有效处理BMS时序数据,记忆故障发生前后的时序关联信息,适用于故障预测与早期预警。

#### 4.2 模型训练与优化

模型训练与优化是提升BMS故障诊断模型性能的核心环节,需平衡模型精度、泛化能力与实时性。训练时,将标准化故障特征数据集按7:2:1比例随机划分为训练集、验证集和测试集,采用分层抽样保证故障样本分布均衡,防止数据泄露引发过拟合。对于传统机器学习模型,优化重点在于超参数调节。利用网格搜索、随机搜索、贝叶斯优化等方法,对SVM的核函数等超参数、随机森林的树数量与深度等超参数进行优化,以提升模型分类精度。针对深度学习模型,则采用梯度下降算法(如Adam、SGD)优化权重,借助批量归一化(BN)、Dropout层抑制过拟合,并合理设置学习率衰减策略,避免训练时梯度消失或爆炸。同时,引入数据增强技术,通过时序数据平移、缩放、添加噪声等方式扩充训练集,提升模型泛化能力,应对不同车型及工况下的电池故障差异。

#### 4.3 模型评估指标

BMS故障诊断模型评估需构建多维度指标体系,全面衡量模型的诊断精度、泛化能力、容错性与实时性,为模型优化提供量化依据,核心指标分为分类性能指标、泛化能力指标与实时性指标三类。分类性能指标是核心评估依据,基于混淆矩阵衍生出准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1分数四大核心指标。准确率反映模型整体诊断正确率,适用于故障与正常样本分布均衡的场景;精确率衡量预测为故障的样本中实际为故障的比例,避免误判导致的过度维修;召回率衡量

实际为故障的样本中被正确诊断的比例,确保故障无遗漏,降低安全风险;F1分数为精确率与召回率的调和平均数,综合反映模型的分类平衡性,适用于样本不均衡场景<sup>[4]</sup>。泛化能力指标通过模型在不同车型、不同工况测试集上的性能差异衡量,常用指标包括泛化误差、标准差,泛化误差越小、标准差越低,说明模型适应性越强。实时性指标针对车载场景需求,衡量模型的诊断延迟与计算效率,常用指标包括单样本诊断时间、每秒处理样本数,要求单样本诊断时间低于100ms,满足实时故障预警需求。另外,还需评估模型的容错性,通过向测试集添加少量噪声数据,观察模型性能变化,确保在数据存在轻微误差时仍能稳定诊断。

#### 结束语

基于大数据的新能源汽车电池管理系统故障诊断研究,为保障电池安全与性能提供了新思路。通过构建多源异构数据采集体系、优化数据预处理方法、提取关键故障特征,结合合适的诊断模型,提升了故障诊断的精度与实时性。随着大数据与人工智能技术进一步发展,故障诊断将更智能高效。但目前研究仍存在局限,如极端温度工况下模型诊断精度有所衰减、多车型数据迁移适配性不足等。未来可聚焦联邦学习与数字孪生技术的融合应用,在保障数据隐私的前提下,构建跨车型、全工况的故障诊断模型,为新能源汽车动力电池安全运维提供更全面的技术支撑。

#### 参考文献

- [1]陈媛媛.基于大数据的新能源汽车故障诊断系统研究[J].内燃机与配件,2025(5):70-72.
- [2]焦连岷.新能源汽车动力电池安全监测与管理技术研究[J].汽车电器,2025(12):25-27.
- [3]吴彦.基于大数据的新能源汽车电池故障诊断研究[J].现代工业经济和信息化,2025,15(7):56-57,60.
- [4]张泽茂.基于大数据的新能源汽车动力电池智能监控与优化策略[J].时代汽车,2024(23):118-120.