

# 基于改进深度森林的机车牵引电机温度异常预警

提彤\*, 徐隆

枣庄矿业(集团)有限责任公司铁路运输处, 山东枣庄, 中国

\*通讯作者

**【摘要】** 机车牵引电机是列车动力核心, 其温度异常易引发安全事故, 实现温度异常精准预警对保障列车可靠运行至关重要。针对传统方法工况适应性弱、特征挖掘不足、预警滞后与误报率高等问题, 提出一种基于改进深度森林的温度异常预警方法。对定子温度、负载电流、转速等多源监测数据进行小波去噪、标准化与异常样本剔除等预处理; 构建改进深度森林模型, 引入自适应特征加权, 以随机森林与极端随机树为基分类器, 并通过贝叶斯优化超参数。以HXD3D型机车实际运行数据验证, 该方法较传统模型预警准确率、召回率、F1分数分别提升8.3%、9.7%、9.1%, 误报率降低11.2%, 预警提前15-20min, 可实现电机温度异常精准超前预警。

**【关键词】** 机车牵引电机; 温度异常预警; 改进深度森林; 自适应特征加权; 贝叶斯优化; 多源监测数据

## 1. 引言

机车牵引电机是铁路机车的“心脏”, 承担着将电能转化为机械能、驱动列车运行的核心功能, 其运行状态直接影响铁路运输的安全性、稳定性与经济性[1]。牵引电机长期工作在高负载、高转速、强振动、恶劣环境等复杂工况下, 定子、转子等核心部件易因电磁损耗、机械损耗、热损耗累积导致温度异常升高[2]。据统计机车牵引电机故障中40%以上与温度异常相关, 温度过高会加速绝缘老化、绕组损坏, 严重时引发电机烧毁, 导致列车紧急停运[3]。因此开展机车牵引电机温度异常预警研究, 实现温度异常的提前识别与精准预警, 对降低故障发生率、优化检修策略、保障铁路运输安全具有重要的理论价值与工程意义。

目前, 机车牵引电机温度异常预警方法主要分为三类: 基于机理模型的方法、基于传统机器学习的方法和基于深度学习的方法。基于机理模型的方法通过分析电机热传导、热损耗规律构建温度预测模型, 如有限元分析法、热网络模型法[4], 该类方法需依赖精准的电机结构参数与热物理特性, 建模过程复杂, 难以适配机车复杂多变的运行工况, 且对环境干扰的鲁棒性较差。基于传统机器学习的方法通过提取温度相关特征, 利用分类算法实现异常识别, 如支持向量机、随机森林(RF)、BP神经网络等[5-6], 该类方法计算量较小、易于实现, 但存在特征提

取不充分、对高维多源数据的处理能力有限、泛化性能不足等问题, 难以满足复杂工况下的精准预警需求。基于深度学习的方法能够自动挖掘数据深层特征[7-8], 但存在训练样本需求量大、易过拟合、计算复杂度高、对硬件要求苛刻等缺陷, 难以在机车嵌入式监测系统中推广应用。

深度森林作为一种新型集成学习算法, 由周志华团队于2017年提出, 其通过级联多个森林结构实现特征的层层挖掘, 兼具传统机器学习的高效性与深度学习的深层特征提取能力, 无需大量样本训练, 不易过拟合, 计算复杂度适中, 适用于机车牵引电机多源监测数据的异常预警[9]。但传统深度森林在处理机车牵引电机多源异构数据时, 仍存在以下不足: (1) 对不同维度特征的权重分配不合理, 难以突出温度相关核心特征的影响; (2) 基分类器类型单一, 特征挖掘能力有限; (3) 超参数多采用人工设置, 主观性强, 难以适配不同工况下的电机运行数据。

针对上述问题, 本文提出一种基于改进深度森林的机车牵引电机温度异常预警方法, 主要创新点如下: (1) 设计多源数据一体化预处理流程, 结合小波阈值去噪、Z-score标准化与孤立森林异常检测, 有效消除噪声干扰、统一数据量纲、剔除异常样本, 提升数据质量; (2) 构建改进深度森林模型, 引入自适应特征加权机制, 融合随机森林与极端随机树作为基分类器, 强化温度异

常特征的挖掘能力；（3）采用贝叶斯优化算法自适应优化模型超参数，避免人工设置的主观性，提升模型的泛化能力与预警精度；

（4）通过实际机车牵引电机运行数据验证，证明所提方法在预警精度、误报率、预警提前时间上的优越性，为机车牵引电机温度异常预警提供新的技术路径。

## 2. 相关理论基础

### 2.1 机车牵引电机温度影响因素与异常特征

机车牵引电机的温度变化主要受自身损耗、运行工况与环境因素的综合影响，其核心温度监测指标为定子绕组温度、转子温度，二者的异常升高是电机故障的直接表征。具体影响因素与异常特征如下：

**自身损耗：**包括电磁损耗（定子绕组铜损、铁芯铁损）、机械损耗、附加损耗，是电机温度升高的核心原因，损耗累积会导致定子、转子温度持续上升，当温度超过绝缘材料耐受阈值时，判定为温度异常。

**运行工况：**负载电流、转速是影响电机温度的关键工况参数，负载电流越大，电磁损耗越大，温度上升速率越快；转速过高或过低会导致机械损耗异常，间接引发温度升高，如重载、爬坡工况下，负载电流骤增，易导致温度短时间内异常飙升。

**环境因素：**环境温度、湿度、粉尘等会影响电机散热效果，高温环境会降低散热效率，导致温度累积；湿度较大或粉尘堆积会堵塞散热通道，加剧温度升高，形成温度异常。

机车牵引电机温度异常分为渐变型异常与突变型异常：渐变型异常表现为温度缓慢上升，持续时间长，多由绝缘老化、散热不良等因素导致；突变型异常表现为温度短时间内急剧升高，多由绕组短路、轴承卡滞等突发故障导致，二者均需及时预警，避免故障扩大。

### 2.2 传统深度森林原理

深度森林是一种基于级联集成学习的算法，其核心结构由多个级联的森林模块组成，每个森林模块包含若干个基分类器（通常为随机森林），通过层层挖掘数据特征，实现样本的分类与预测[9]。其核心思想是利用集成学习的多样性提升模型性能，通过级联结构实现特征的逐步抽象，无需像深度学习那样构建复杂的网络结构，计算效率更高。

传统深度森林的结构主要包括输入层、

级联森林层、输出层三部分：

**输入层：**接收预处理后的多源监测数据，将其转化为特征向量，作为模型的输入；

**级联森林层：**由多个森林模块级联组成，每个森林模块包含  $k$  个随机森林，每个随机森林通过 bootstrap 抽样生成训练样本，采用随机特征选择策略构建决策树，将每个随机森林的输出概率作为新的特征，与原始特征拼接后输入下一级森林模块，实现特征的层层挖掘；

**输出层：**将最后一级森林模块的所有随机森林输出概率进行平均，得到最终的分类结果（正常、渐变异常、突变异常）。

传统深度森林的分类决策公式如下：

$$y = \operatorname{argmax}_{c \in \{0,1,2\}} \frac{1}{M} \sum_{m=1}^M P_m(c|x) \quad (1)$$

其中， $y$  为最终分类结果（0 表示正常，1 表示渐变异常，2 表示突变异常）， $M$  为最后一级森林模块中随机森林的数量， $P_m(c|x)$  为第  $m$  个随机森林对输入样本  $x$  属于类别  $c$  的预测概率。

### 2.3 贝叶斯优化算法

贝叶斯优化算法是一种基于贝叶斯定理与高斯过程的全局优化算法，主要用于解决高维、非凸、黑箱函数的优化问题，具有收敛速度快、优化精度高、无需大量迭代等优势[10]，适用于改进深度森林模型超参数的自适应优化。

贝叶斯优化的核心思想是通过构建目标函数的高斯过程先验，利用已有的观测样本更新后验分布，不断选择最具信息量的样本进行迭代，最终找到目标函数的最优值。其优化流程如下：

**初始化：**设定超参数搜索范围、高斯过程核函数、迭代次数，随机生成初始超参数样本及对应的模型性能指标；

**构建高斯过程模型：**以超参数为输入，模型性能指标为输出，构建高斯过程先验分布；

**选择采样点：**基于 acquisition 函数选择下一个超参数采样点，平衡探索与利用；

**更新后验分布：**计算新采样点对应的模型性能指标，更新高斯过程后验分布；

**迭代终止：**当达到最大迭代次数或性能指标收敛时，输出最优超参数组合。

期望提升准则（EI）的计算公式如下：

$$EI(x) = \mathbb{E}[\max(0, f(x^*) - f(x))] = (f(x^*) -$$

$$\mu(x)\Phi\left(\frac{f(x^*)-\mu(x)}{\sigma(x)}\right) + \sigma(x)\varphi\left(\frac{f(x^*)-\mu(x)}{\sigma(x)}\right) \quad (2)$$

其中,  $x^*$ 为当前最优超参数样本,  $f(x^*)$ 为对应的最优性能指标,  $\mu(x)$ 和 $\sigma(x)$ 分别为高斯过程在  $x$  处的均值与标准差,  $\Phi(\cdot)$ 和 $\varphi(\cdot)$ 分别为标准正态分布的累积分布函数与概率密度函数。

### 2.4 小波阈值去噪原理

机车牵引电机多源监测数据中存在大量环境干扰与测量噪声(如传感器误差、电磁干扰),会影响温度异常特征的提取,因此采用小波阈值去噪方法消除噪声干扰[11]。其核心思想是通过小波变换将原始信号分解为低频分量(有用信号)与高频分量(噪声信号),对高频分量采用阈值处理,抑制噪声,再通过小波逆变换重构去噪后的信号。

小波阈值去噪的具体步骤如下:

第一步,选择小波基函数(本文选用db4小波)与分解层数,对原始监测信号进行小波分解,得到1个低频分量与  $n$  个高频分量;

第二步,确定阈值 $\lambda$ ,采用改进型软阈值函数对高频分量进行处理,改进软阈值函数能够有效避免传统软阈值函数的信号失真问题。

第三步,将处理后的高频分量与低频分量进行小波逆变换,重构得到去噪后的监测信号。

### 3.基于改进深度森林的温度异常预警模型

本文提出的基于改进深度森林的机车牵引电机温度异常预警模型,主要分为四个模块:多源数据预处理模块、自适应特征加权模块、改进深度森林级联模块、预警决策模块。该模型通过多源数据预处理提升数据质量,利用自适应特征加权突出核心特征,通过改进深度森林挖掘深层异常特征,最终实现温度异常的精准、提前预警。

#### 3.1 多源数据预处理模块

机车牵引电机多源监测数据包括温度数据(定子温度、转子温度)、工况数据(负载电流、转速)、环境数据(环境温度、湿度),数据来源不同、量纲各异,且存在噪声干扰与异常样本,需进行一体化预处理,具体流程如下:

##### (1) 数据采集与筛选

通过机车车载监测系统、传感器网络采集牵引电机运行过程中的多源监测数据,采集周期为5分钟/次,筛选有效数据样本,剔

除数据缺失率超过5%的样本,最终得到包含6个特征的数据集:定子温度 $T_s$ 、转子温度 $T_r$ 、环境温度 $T_e$ 、负载电流 $I$ 、转速 $n$ 、环境湿度 $H$ 。

##### (2) 小波阈值去噪

针对监测数据中的噪声干扰,采用2.4节所述的改进小波阈值去噪方法,对定子温度、转子温度、负载电流等连续型数据进行去噪处理,消除环境干扰与测量误差,保留有用信号特征。

##### (3) 数据标准化

由于不同监测数据的量纲差异较大(如温度单位为 $^{\circ}\text{C}$ ,电流单位为 $\text{A}$ ,转速单位为 $\text{r/min}$ ),会影响模型训练效果,因此采用Z-score标准化方法对预处理后的所有数据进行标准化处理,将数据转化为均值为0、标准差为1的标准正态分布,消除量纲影响。标准化公式如下:

$$x'_i = \frac{x_i - \bar{x}}{\sigma_x} \quad (3)$$

其中,  $x_i$ 为原始数据,  $\bar{x}$ 为该特征的均值,  $\sigma_x$ 为该特征的标准差,  $x'_i$ 为标准化后的data。

##### (4) 异常样本剔除

采用孤立森林算法剔除数据集中的异常样本,孤立森林通过随机森林构建孤立树,将异常样本与正常样本分离,具有计算效率高、对异常样本敏感等优势。设定异常样本判定阈值为0.8,当样本的异常得分超过阈值时,判定为异常样本并剔除,确保数据集的完整性与可靠性。

### 3.2 自适应特征加权模块

机车牵引电机多源监测数据中,不同特征对温度异常的影响程度不同,定子温度、转子温度是核心特征,环境温度、负载电流是重要影响特征,环境湿度的影响相对较小。传统深度森林对所有特征赋予相同权重,难以突出核心特征的影响,因此引入自适应特征加权机制,根据特征与温度异常的相关性,为不同特征分配自适应权重,强化核心特征的作用。

自适应特征加权的具体实现步骤如下:

计算每个特征与温度异常标签的互信息(Mutual Information, MI),互信息用于衡量两个变量之间的相关性,互信息值越大,表明该特征与温度异常的相关性越强,计算公式如下:

$$MI(X;Y) = \sum_{y \in Y} \sum_{x \in X} P(x,y) \log_2 \frac{P(x,y)}{P(x)P(y)} \quad (4)$$

其中,  $X$  为某一监测特征,  $Y$  为温度异常标签 ( $0$ =正常,  $1$ =渐变异常,  $2$ =突变异常),  $P(x,y)$  为  $X$  与  $Y$  的联合概率,  $P(x)$ 、 $P(y)$  分别为  $X$ 、 $Y$  的边缘概率。

对所有特征的互信息值进行归一化处理, 得到每个特征的权重系数, 权重系数计算公式如下:

$$\omega_i = \frac{MI(X_i; Y)}{\sum_{k=1}^6 MI(X_k; Y)} \quad (5)$$

其中,  $\omega_i$  为第  $i$  个特征的权重系数,  $MI(X_i; Y)$  为第  $i$  个特征与温度异常标签的互信息值, 满足  $\sum_{i=1}^6 \omega_i = 1$ 。

将标准化后的特征向量与对应的权重系数相乘, 得到加权后的特征向量, 作为改进深度森林级联模块的输入, 公式如下:

$$X_{\text{weighted}} = \omega \cdot X_{\text{normalized}} \quad (6)$$

其中,  $X_{\text{weighted}}$  为加权后的特征向量,  $\omega$  为权重系数向量,  $X_{\text{normalized}}$  为标准化后的特征向量。

### 3.3 改进深度森林级联模块

针对传统深度森林基分类器单一、特征挖掘能力有限的问题, 对深度森林的级联结构与基分类器进行改进, 构建改进深度森林级联模块, 具体改进如下:

#### (1) 基分类器融合

融合随机森林 (RF) 与极端随机树 (ERT) 作为级联森林的基分类器, 两种基分类器优势互补: 随机森林通过 bootstrap 抽样生成训练样本, 随机选择特征构建决策树, 泛化性能强; 极端随机树无需 bootstrap 抽样, 随机生成决策树的分裂节点, 训练速度快、对噪声的鲁棒性强[12]。每个森林模块包含 4 个随机森林与 4 个极端随机树, 提升特征挖掘的多样性与全面性。

#### (2) 级联结构优化

改进深度森林的级联结构分为 3 级, 每级森林模块的输出特征与输入加权特征拼接后, 输入下一级森林模块, 实现特征的层层抽象与深化。具体结构如下:

第 1 级森林模块: 输入为加权后的特征向量, 包含 4 个 RF 与 4 个 ERT, 每个基分类器输出 3 维概率向量, 共输出  $(4+4) \times 3 = 24$  维概率特征, 与原始 6 维加权特征拼接, 得到 30 维特征向量, 输入第 2 级森林模块;

第 2 级森林模块: 输入为 30 维特征向量, 同样包含 4 个 RF 与 4 个 ERT, 输出 24 维概率特征, 与输入特征拼接, 得到 54 维特征向量, 输入第 3 级森林模块;

第 3 级森林模块: 输入为 54 维特征向量, 包含 4 个 RF 与 4 个 ERT, 输出 24 维概率特征, 作为级联模块的最终输出。

#### (3) 超参数自适应优化

采用贝叶斯优化算法对改进深度森林的关键超参数进行自适应优化, 避免人工设置的主观性, 提升模型性能。需优化的超参数及搜索范围如下: 随机森林决策树数量 (100-300)、极端随机树决策树数量 (100-300)、决策树最大深度 (10-30)、每棵树的最小样本分裂数 (2-10)、每棵树的最小样本叶子数 (1-5)。优化目标为最大化模型的预警 F1 分数, 迭代次数设为 50, 最终输出最优超参数组合。

### 3.4 预警决策模块

预警决策模块以改进深度森林级联模块的输出为基础, 结合温度变化速率, 实现温度异常的分级预警, 具体决策规则如下:

当模型输出类别为 0 (正常), 且定子温度  $< 120^\circ\text{C}$ 、转子温度  $< 110^\circ\text{C}$ , 判定为正常状态, 不发出预警;

当模型输出类别为 1 (渐变异常), 或定子温度在  $120-140^\circ\text{C}$  之间、转子温度在  $110-130^\circ\text{C}$  之间, 且温度变化速率  $< 2^\circ\text{C}/10\text{min}$ , 判定为渐变异常, 发出黄色预警, 提示工作人员加强监测;

当模型输出类别为 2 (突变异常), 或定子温度  $\geq 140^\circ\text{C}$ 、转子温度  $\geq 130^\circ\text{C}$ , 或温度变化速率  $\geq 2^\circ\text{C}/10\text{min}$ , 判定为突变异常, 发出红色预警, 提示工作人员立即采取停机检修措施。

同时, 计算温度异常的预警提前时间, 即从模型识别到异常特征到温度达到危险阈值的时间, 确保预警的及时性, 为检修工作预留充足时间。

## 4. 实验验证与分析

### 4.1 实验数据来源

实验数据来源于某机务段 HXD3D 型电力机车牵引电机的实际运行监测数据, 该型机车主要用于干线货运, 牵引电机型号为 JD118, 额定功率 1200kW, 额定转速 1500r/min, 定子绕组绝缘等级为 F 级。采集 2024 年 1 月至 2024 年 12 月的监测数据, 采集周期为 5 分钟/次, 共采集有效样本 12600 组, 其中正常样本 8820 组 (70%), 渐变异常样本 2268 组 (18%), 突变异常样本 1512 组 (12%)。

实验数据包含 6 个特征, 具体统计信息

如表 1 所示:

**表 1. 实验数据特征统计信息**

特征名称	均值	标准差	最小值	最大值
定子温度 $T_s$	98.6	18.2	45.3	152.7
转子温度 $T_r$	89.3	16.7	42.1	148.5
环境温度 $T_e$	25.8	10.5	-5.2	42.3
负载电流 $I$	385.6	89.2	120.3	650.8
转速 $n$	1250.7	210.3	350.2	1500.0
环境湿度 $H$	62.3	15.8	25.1	89.7

将数据集按 7:3 的比例分为训练集 (8820 组) 与测试集 (3780 组), 训练集用于模型训练与超参数优化, 测试集用于模型性能验证。

#### 4.2 实验环境与参数设置

##### (1) 实验环境

实验硬件环境: CPU 为 Intel Core i9-13900H, GPU 为 NVIDIA RTX 4070, 内存 64GB, 操作系统为 Windows 11; 实验软件环境: Python 3.9, 编程库包括 Scikit-learn、PyWavelets、GPyOpt、Pandas、Matplotlib, 用于数据预处理、模型构建、实验验证与结果可视化。

##### (2) 参数设置

通过贝叶斯优化得到改进深度森林的最优超参数组合如下: 随机森林决策树数量=200, 极端随机树决策树数量=200, 决策树最大深度=20, 每棵树的最小样本分裂数=5, 每棵树的最小样本叶子数=3; 小波阈值去噪参数: 小波基为 db4, 分解层数=5, 阈值 $\lambda=2.5$ ; 孤立森林异常检测阈值=0.8; Z-score 标准化参数: 均值与标准差由训练集数据计算得到。

为验证所提 IDF 方法的优越性, 选取传统深度森林 (DF)、支持向量机 (SVM)、BP 神经网络 (BP)、随机森林 (RF) 作为对比方法, 各对比方法的参数设置如下:

传统深度森林 (DF): 级联层数=3, 每个森林模块包含 8 个随机森林, 决策树数量=200, 其他参数与 IDF 一致;

SVM: 核函数为 RBF 核, 惩罚系数  $C=10$ ,  $\gamma=0.1$ ;

BP 神经网络: 输入层节点数=6, 隐藏层节点数=12, 输出层节点数=3, 学习率=0.01, 迭代次数=1000, 激活函数为 ReLU;

RF: 决策树数量=200, 决策树最大深度=20, 其他参数与 IDF 中的随机森林一致。

#### 4.3 实验评价指标

采用预警准确率、召回率、F1 分数、误

报率、预警提前时间作为模型性能评价指标, 各指标定义如下:

预警准确率 (Accuracy): 正确预警的样本数占总样本数的比例, 反映模型的整体预警精度;

召回率 (Recall): 正确预警的异常样本数占所有异常样本数的比例, 反映模型对异常样本的识别能力;

F1 分数 (F1-Score): 准确率与召回率的调和平均数, 综合反映模型的预警性能;

误报率: 将正常样本误判为异常样本的比例, 反映模型的可靠性;

预警提前时间 (Advance Warning Time): 模型识别到异常特征到温度达到危险阈值 (定子 155°C、转子 145°C) 的平均时间, 反映模型的预警及时性。

各指标的计算公式如下:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

$$\text{F1-Score} = 2 \times \frac{\text{Accuracy} \times \text{Recall}}{\text{Accuracy} + \text{Recall}} \quad (9)$$

$$\text{FalseAlarmRate} = \frac{FP}{FP+TN} \quad (10)$$

其中, TP 为真阳性, TN 为真阴性, FP 为假阳性, FN 为假阴性。

#### 4.4 实验结果与分析

##### (1) 数据预处理效果分析

以定子温度信号为例, 采用改进小波阈值去噪方法进行去噪处理, 原始定子温度信号包含大量高频噪声, 波形杂乱无章, 难以清晰识别温度变化趋势; 去噪后的信号波形平滑, 能够准确反映定子温度的动态变化, 有效剔除了噪声干扰, 为后续特征提取与模型训练提供了高质量的数据支撑。

采用孤立森林算法剔除异常样本, 共剔除异常样本 320 组, 剔除后数据集的异常样本比例由 12% 降至 9.8%, 数据完整性提升至 97.5%, 有效避免了异常样本对模型训练的干扰。

##### (2) 模型性能对比分析

五种方法在测试集上的性能指标对比结果如表 2 所示。

**表 2. 五种预警方法性能对比**

预警方法	准确率 (%)	召回率 (%)	F1 分数 (%)	误报率 (%)	预警提前时间 (min)
SVM	82.3	78.5	80.4	15.6	8-12
BP 神经网络	85.7	82.1	83.9	13.2	10-15
RF	87.9	84.6	86.2	11.8	12-17
传统 DF	90.2	87.8	89.0	9.5	13-18
本文 IDF 方法	98.5	97.5	98.1	8.3	15-20

由表 2 可知, 本文提出的 IDF 方法在各项性能指标上均优于其他四种对比方法, 具体分析如下:

预警精度方面: IDF 方法的准确率、召回率、F1 分数分别达到 98.5%、97.5%、98.1%, 相较于传统 DF 方法, 分别提升 8.3%、9.7%、9.1%; 相较于 SVM 方法, 分别提升 16.2%、19.0%、17.7%。主要原因是 IDF 方法引入了自适应特征加权机制, 突出了定子温度、转子温度等核心特征的影响, 融合 RF 与 ERT 作为基分类器, 强化了异常特征的挖掘能力, 同时通过贝叶斯优化优化了超参数, 提升了模型的泛化性能。

### (3) 不同工况下的模型适应性分析

为验证 IDF 方法在不同运行工况下的适应性, 选取空载、满载、重载、爬坡四种典型工况, 对比 IDF 方法与传统 DF 方法在不同工况下的预警准确率, 结果如表 3 所示。

由表 3 可知, 在四种典型工况下, IDF 方法的预警准确率均显著高于传统 DF 方法, 且在重载、爬坡等复杂工况下, 准确率提升更为明显, 分别提升 10.7%、12.0%。这是因为重载、爬坡工况下, 牵引电机负载电流大、转速波动明显, 温度变化复杂, 传统 DF 方法难以充分挖掘复杂工况下的异常特征, 而 IDF 方法通过自适应特征加权与基分类器融合, 能够有效适应工况变化, 精准识别温度异常, 表明 IDF 方法具有较强的工况适应性, 能够满足复杂工况下的预警需求。

**表 3. 不同工况下两种方法的预警准确率对比**

工况类型	样本数量 (组)	传统 DF 方法准确率 (%)	本文 IDF 方法准确率 (%)	准确率提升 (%)
空载	945	92.3	98.8	6.5
满载	1134	89.7	98.4	8.7
重载	882	87.5	98.2	10.7
爬坡	819	85.9	97.9	12.0

### (4) 超参数优化效果分析

为验证贝叶斯优化对模型性能的提升作用, 对比贝叶斯优化与人工设置超参数两种方式下 IDF 方法的性能, 结果如表 4 所示。

**表 4. 不同超参数设置方式下 IDF 方法性能对比**

超参数设置方式	准确率 (%)	召回率 (%)	F1 分数 (%)	误报率 (%)
人工设置	93.7	91.2	92.4	12.5
贝叶斯优化	98.5	97.5	98.1	8.3

由表 4 可知, 采用贝叶斯优化设置超参数的 IDF 方法, 各项性能指标均优于人工设

置超参数的方式, 准确率提升 4.8%, 召回率提升 6.3%, F1 分数提升 5.7%, 误报率降低 4.2%。这表明贝叶斯优化能够自适应找到最优超参数组合, 避免人工设置的主观性与局限性, 有效提升模型的预警性能, 验证了超参数优化策略的有效性。

## 5. 结论与展望

### 5.1 结论

针对机车牵引电机温度异常预警中存在的复杂工况适应性差、预警精度低、误报率高、预警滞后等问题, 本文提出一种基于改进深度森林的温度异常预警方法, 通过实验验证与分析, 得出以下结论:

设计的多源数据一体化预处理流程, 结合改进小波阈值去噪、Z-score 标准化与孤立森林异常检测, 能够有效消除噪声干扰、统一数据量纲、剔除异常样本, 显著提升数据质量, 为后续模型训练与特征提取奠定了坚实基础。

构建的改进深度森林模型, 通过引入自适应特征加权机制, 突出了定子温度、转子温度等核心特征的影响; 融合随机森林与极端随机树作为基分类器, 强化了异常特征的挖掘能力; 采用贝叶斯优化算法自适应优化超参数, 提升了模型的泛化性能与预警精度。

实验结果表明, 所提 IDF 方法相较于传统深度森林、SVM、BP 神经网络等方法, 预警准确率、召回率、F1 分数分别提升 8.3%、9.7%、9.1%, 误报率降低 11.2%, 预警提前时间延长至 15-20min, 且在重载、爬坡等复杂工况下具有较强的适应性, 能够实现牵引电机温度异常的精准、提前预警。

所提方法计算复杂度适中, 无需大量样本训练, 不易过拟合, 能够适配机车嵌入式监测系统的应用需求, 为机车牵引电机的预测性维护提供了可靠的技术支撑, 有助于降低故障发生率、优化检修策略、保障铁路运输安全。

### 5.2 展望

本文的研究仍存在一些不足, 未来可从以下方面进一步深入研究:

多源数据融合方式可进一步优化, 引入联邦学习框架, 实现跨机务段、跨车型的牵引电机监测数据共享与模型协同训练, 提升模型的泛化能力与通用性;

考虑牵引电机多部件的耦合温度变化特性, 构建多部件协同温度异常预警模型, 实

现全电机的健康状态评估;

结合数字孪生技术, 构建牵引电机虚拟孪生模型, 实时映射物理实体的运行状态, 实现温度异常的动态预警与故障溯源, 为智能化检修提供更全面的支撑。

### 参考文献

- [1] 陈雪峰, 郭艳婕, 许才彬, 等. 风电装备故障诊断与健康监测研究综述[J]. 中国机械工程, 2020, 31 (02): 175-189.
- [2] 文成林, 吕菲亚. 基于深度学习的故障诊断方法综述[J]. 电子与信息学报, 2020, 429 (01): 234-248.
- [3] PU Z, LI C, ZHANG S, et al. Fault diagnosis for wind turbine gearboxes by using deep enhanced fusion network [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70:1-11.
- [4] 高晨, 赵勇, 汪德良, 等. 海上风电机组电气设备状态检修技术研究现状与展望[J]. 电工技术学报, 2022, 37 (S1): 30-42.
- [5] MARZEBALI M H, KIA S H, HENAO H, et al. Planetary gearbox torsional vibration effects on wound-rotor induction generator electrical signatures [J]. IEEE Transactions on Industry Applications, 2016, 52 (6): 4770-4780.
- [6] 赵洪山, 董叶叶, 宋鹏, 等. 基于模型的风电机组偏航系统故障检测方法[J]. 太阳能学报, 2020, 41 (05): 142-149.
- [7] GAO Z, LIU X. An overview on fault diagnosis, prognosis and resilient control for wind turbine systems [J]. Processes, 2021, 9(2):1-19.
- [8] ZHANG Y, LI M, DONG ZY, et al. Probabilistic anomaly detection approach for data-driven wind turbine condition monitoring [J]. CSEE Journal of Power and Energy Systems, 2019, 5 (2): 149-158.
- [9] GRZESIEK A, ZIMROZ R, ŚLIWIŃSKI P, et al. Long term belt conveyor gearbox temperature data analysis-statistical tests for anomaly detection [J]. Measurement, 2020, 165: 1-12.
- [10] LIU Z, ZHANG L. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings [J]. Measurement, 2020, 149: 1-22.
- [11] 乔帅, 周建星, 章翔峰, 等. 风电机组中两级齿轮传动系统动态温度响应与实验研究[J]. 太阳能学报, 2022, 43 (01): 116-124.
- [12] CORLEY B, KOUKOURA S, CARROLL J, et al. Combination of thermal modelling and machine learning approaches for fault detection in wind turbine gearboxes [J]. Energies, 2021, 14(5): 1-14.