

基于递归二分类法的电力机车 四象限输入过流故障诊断研究

王建华

(朔黄铁路发展有限责任公司 机辆分公司, 河北 肃宁 062350)

摘 要: 四象限输入过流故障是HXD1型机车牵引变流器的主要故障之一, 由于引发该故障的原因较多且故障样本分布不均, 导致该故障难以诊断与定位, 为此提出了一种基于递归二分类的故障诊断方法。基于K-Means与小波包分解对数据相关关键字段进行特征提取, 用基于集成学习的梯度提升决策树(GBDT)构造可靠的分类模型以及用递归二分类法对多标签数据进行递归分类与诊断。对HXD1型机车四象限输入过流故障样本文件进行相关试验, 结果表明该方法对于故障样本分类及故障诊断定位有显著提升效果。

关键词: 递归二分类; 电力机车; 四象限输入过流; 小波包分解; 梯度提升决策树; 机器学习; 故障诊断
中图分类号: U264.3⁺7; U269.32⁺2 文献标识码: A

doi: 10.13890/j.issn.1000-128x.2019.02.016

Research on Four-Quadrant Input Over-current Fault Diagnosis of Electric Locomotive Based on Recursive 2-Classification Method

WANG Jianhua

(Locomotive Branch, Shuohuang Railway Development Co., Ltd., Suning, Hebei 062350, China)

Abstract: Four-quadrant over-current fault is one of the main faults of the HXD1 locomotive traction converter. Due to the class distribution of fault cause is unbalanced and there are a lot of over-current fault causes, four-quadrant over-current fault is difficult to distinguish and diagnose. A fault diagnosis method based on recursive 2-Classification was proposed. Based on K-Means and wavelet packet decomposition, feature extraction of data-related key fields was carried out. Reliable classification model was constructed by gradient boosting decision tree (GBDT) based on ensemble learning, and multi-label data was classified and diagnosed recursively by recursive 2-classification. Experiments on four-quadrant input over-current fault data files showed that this method could improve the classification and diagnosis of fault samples.

Keywords: recursive 2-classification; electric locomotive; four-quadrant input over-current; wavelet packet decomposition; gradient boosting decision tree; machine learning; fault diagnosis

0 引言

变流器是电力机车牵引系统的关键部件, 变流器故障将会导致机车失去部分或全部牵引动力, 动力损失严重时将会导致机破而影响整个铁路干线。因此变流器一旦发生故障, 需进行准确故障定位, 以快速查

明故障原因, 采取相关措施, 这对电力机车牵引系统高效运行及列车的平稳运营至关重要。

1 研究现状

目前国内外对电力机车牵引变流器故障诊断的研究, 主要分成三大类:

①基于专家经验的故障诊断。文献[1]通过对变流

收稿日期: 2018-06-22; 修回日期: 2019-01-15

器发生较频繁的故障进行简要分析，结合经验提出处理建议。

②基于电气模型的故障诊断。文献[2]针对CRH2型动车组三电平逆变器功率开关器件开路故障，提出了基于故障逻辑参数模型，通过对实际模型识别量和故障模型识别量进行残差分析识别出故障。文献[3]针对永磁同步电机牵引系统中的逆变器开路故障，通过分析正常逆变器驱动PMSM与观测器的输出的三相电流残差，分析功率开关器件单管和双管开路故障特征所在区间实现诊断。文献[4]通过将正常和故障条件下系统输入电流进行比较，通过比较其中的差异获得故障状态，并根据故障诊断表对故障开关管进行定位。

③基于数据的故障诊断。文献[5]针对SS₈型机车的主变流器功率开关器件的开路故障，提出了基于小波分析和支持向量机(SVM)的故障诊断方法。文献[6]针对200 km/h的动车组牵引变流器的三相输出电流的不平衡故障，提出了运用小波变换提取特征，结合BP神经网络的故障诊断方法。文献[7]提出了基于SOM的神经网络用于整流功率元件的开路故障诊断。文献[8]针对BP算法存在收敛速度慢的问题，提出了用遗传算法来更新神经网络的权值参数，结合小波分析，用于整流器开路故障诊断。文献[9]利用粗糙集对故障特征数据库中的数据进行处理，提取出有效的故障模式规则用于故障诊断。文献[10]提出了基于改进蚁群神经网络对地铁车辆牵引逆变器的开路和短路故障进行诊断。

通过对以上所述各方法进行分析，发现它们都存在一定的局限性：

①基于专家经验的故障诊断依赖于专家经验，当数据量大时，诊断效率低。

②基于电气模型故障诊断的仿真模型太过理想化，不能反映机车车辆在实际运行过程中受运行工况、环境等复杂多因素影响的情况。

③基于数据的故障诊断在计算机硬件水平和开源环境的发展下，得到了广泛的应用。但目前各大科研院校进行的仿真试验缺乏实际列车运行的数据。复杂的模型在列车实际运行数据样本文件数量少、故障样本分布不均衡的情况下不能很好地解决故障诊断问题。

④以上大多数研究对象是变流器元器件的开路故障，对于其他类型的故障诊断缺乏研究成果。

基于上述各模型的局限性，本文立足于电力机车实际运行数据，针对故障数据样本少的实际情况和变流器模块内部故障诊断的复杂性，提出了递归二分类法的变流器四象限输入过流故障诊断方法。该算法运用小波包分解进行特征提取，用集成学习模型的梯度提升决策树(GBDT)构造分类模型，对实际的变流器故障样本数据进行试验验证。

2 变流器四象限输入过流

变流器四象限输入过流是变流器故障中一种常见

的故障。引起四象限输入过流故障的原因主要有以下4类：输入异常、元器件异常、检测异常、其他原因。各类的具体原因如表1所示。

表1 四象限输入过流故障原因

原因类型	具体原因
输入异常	①网压波动异常；②受电弓异常；③同步信号线接反；④输入大线接反
元器件异常	①四象限模块故障；②逆变模块故障；③电容故障；④直流连接母排故障
检测异常	①电流传感器故障；②中间电压传感器故障
其他异常	①网侧板卡异常；②脉冲输出异常

导致电力机车四象限输入过流最常见的3种故障原因为供电网压异常、四象限模块故障、逆变模块故障。其故障表征分别为：

①供电网压异常。主要是根据牵引供电网电压瞬时值来判断，正常的牵引网压波形应该是频率为50 Hz的正弦波。供电网发生异常时，牵引网压瞬时值发生大幅度的波动，相邻正向（或负向）峰值相差大。

②四象限模块故障和逆变模块故障均属于模块故障。变流器在正常工作时，若四象限电流突然出现半波波形且过流，或者变流器在充电时，中间直流电压出现直流电压突变伴随输入电流出现过流，可判定为模块故障。四象限模块故障与逆变模块故障时的数据差异很小，较难区分。

由上述的分析可知，在不同工况环境下，属于相同故障原因的样本文件，数据表现可能完全不同，而属于不同故障原因的样本文件数据表现有时却很相似。基于此，本文利用机器学习方法对电力机车四象限输入过流进行故障诊断，其难点是需要对各故障对应的数据进行精细的特征提取，建立有效的分类模型与诊断法则。

3 数据样本的机器学习与故障诊断

电力机车变流器四象限输入过流故障数据文件通过列车传输设备进行数据传送，存放在本地文件、HDFS分布式系统或HIVE数据库中。有监督机器学习的分类算法流程如图1所示。

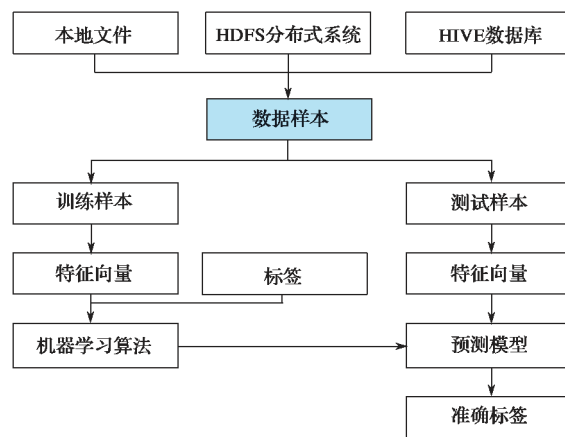


图1 有监督机器学习分类算法流程

3.1 数据特征提取

四象限输入过流故障数据样本相关字段说明如表 2 所示。

表 2 相关字段说明

字段名	含义	字段名	含义
Ud1	中间直流电压 1	lqs3	控制电流
lqs1	第 1 轴四象限电流	Unet	牵引网压
Ud2	中间直流电压 2	Inet	一次侧电流
lqs2	第 2 轴四象限电流	PLL	一次侧锁相值
Ud3	中间直流电压 3		

对样本文件初步统计, 可知每个样本有 9 个关键字段, 每个字段都有 17 500 条记录, 每段数据总共有 157 500 个点。考虑到算法的时效性, 对这样的大规模数据必须进行特征提取, 从大量数据中提取出能代表原始数据的少量关键特征向量。

3.1.1 短窗傅里叶变换 (STFT)

在时域上加窗分段进行 FFT, 提高信号在时域上的分辨率。STFT 存在以下局限性: ①窗口过小会降低数据在频域上的分辨率, 窗口过大会降低数据在时域上的分辨率; ②在提高时域分辨率的同时降低了频域分辨率, 信号在时域分辨率与频域分辨率上不能同时达到最优。因此, 可考虑利用小波包分解进行特征提取。

3.1.2 小波包分解

连续小波变换^[11]定义:

设 $\varphi(t)$ 是平方可积函数, 若满足如下条件:

$$\varphi(t) = \int_{-\infty}^{\infty} \frac{|\varphi^*(w)|^2}{|w|} dw < \infty \quad (1)$$

其中, $\varphi^*(w)$ 为 $\varphi(t)$ 的傅里叶变换函数, 则称 $\varphi(t)$ 为一个基本小波或小波母函数。

将 $\varphi(t)$ 进行伸缩因子为 a 、平移因子为 b 的伸缩和平移, 令其平移后的函数为 $\varphi_{a,b}(t)$, 则有

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \varphi\left(\frac{t-b}{a}\right) \quad a \neq 0, b \in \mathbf{R} \quad (2)$$

$\varphi_{a,b}(t)$ 为依赖于 a 和 b 的小波基函数, 连续时间信号 $f(t)[f(t) \in L^2(\mathbf{R})]$ 的连续小波变换定义为

$$w(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \varphi_{a,b}^*(t) dt \quad (3)$$

式中: $\varphi_{a,b}^*(t)$ 为共轭函数; 参数 t, a, b 均为连续变量。

小波变换通过小波基函数的伸缩和平移来构成一系列分辨率不同的正交投影空间及其对应的基, 自适应调整窗口大小。在提高时域分辨率的同时, 不降低频域分辨率, 能同时在时频域上得到很高的分辨率。小波变换基本流程如图 2 所示。

小波包 (Wavelet Packet) 分解是小波分析的进一步深化。小波包变换是小波变换扩展, 小波分解只对低频逐层进行分解, 比较关注信号的低频特征。而小波包分解将数据的高频信号段和低频信号段同时逐层

进行分解, 关注信号在整个频段上的特征。小波包分解基本流程如图 3 所示。

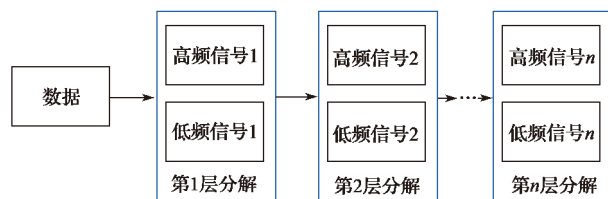


图 2 小波变换流程

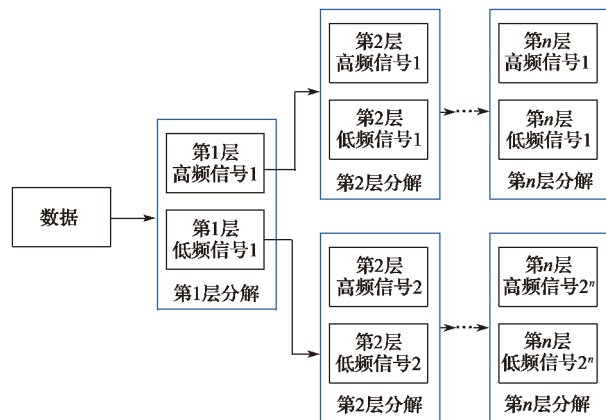


图 3 小波包分解流程

小波包分解特征提取算法流程如图 4 所示。

3.2 基于梯度提升决策树的模型分类

集成学习 (Ensemble learning) 定义为通过构建多个弱分类器, 来最终组成一个强大的分类模型。典型的集成学习模型有 Boosting, Bagging, 随机森林和梯度提升决策树。

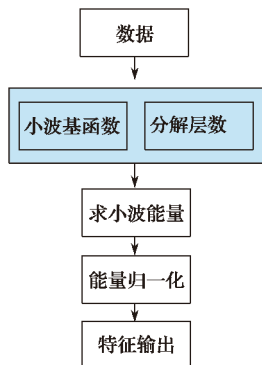
梯度提升决策树 (Gradient Boosting Decision Tree, GBDT) 由 Friedman^[12] 于 2001 年提出, 属于集成学习的一种模型, 其强大的拟合和泛化性能, 使其不仅在世界顶尖的数据挖掘平台 Kaggle 竞赛中屡获名次, 并且在 Facebook 和阿里巴巴等互联网公司中被广泛使用。

GBDT 算法步骤:

步骤 1: 根据输入数据初始化回归树 $T_1, T_2, \dots, T_n$ 。

步骤 2: 分别求回归树 $T_1, T_2, \dots, T_n$ 的损失函数梯度, 并用各回归树的梯度拟合残差回归树 $T_1^*, T_2^*, \dots, T_n^*$, 如图 5 所示。

步骤 3: 用回归树 T_i 的当前值与其他残差回归树 $T_j^* (j \in 1, 2, \dots, n, j \neq i)$ 求和, 得到回归树 T_i 的预测值。将各回归树预测值与分类 (回归) 阈值比较, 得到最终输出, 如图 6 所示。



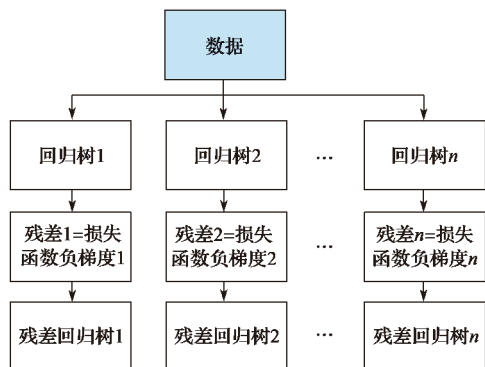


图5 残差回归树生成

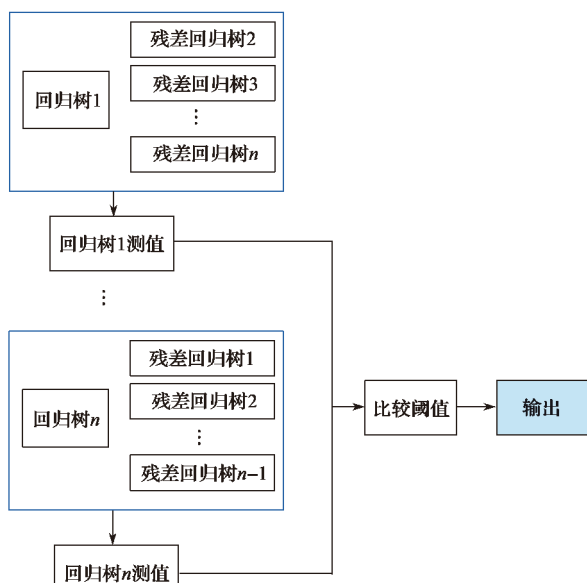


图6 梯度提升决策树预测值输出

3.3 模型评估

3.3.1 聚类模型评估

轮廓系数 (Silhouette Coefficient) 同时兼顾聚类的凝聚度 (Cohesion) 与分离度 (Separation)，用于评估聚类模型的效果。

轮廓系数计算步骤：

步骤 1：对于已聚类数据中第 i 个样本 x^i ，计算 x^i 与其同一个类簇内的所有其他样本距离平均值 a^i ，用于量化簇内凝聚度。

步骤 2：选取 x^i 外的一个簇 b ，计算 x^i 与簇 b 中所有样本的平均距离，遍历所有其他簇，找到最近的平均距离，记作 b^i ，用于量化簇间分离度。

步骤 3：对样本 x^i 轮廓系数进行计算，计算式为 $sc^i = \frac{b^i - a^i}{\max(b^i, a^i)}$ 。

步骤 4：对所有样本 x 求出平均值为当前聚类结果的整体轮廓系数。

轮廓系数取值范围为 $[-1, 1]$ ，数值越接近于 1，聚类效果越好。利用轮廓系数，可确定各特征提取的聚类效果，确定初步特征，特征提取流程如图 7 所示。

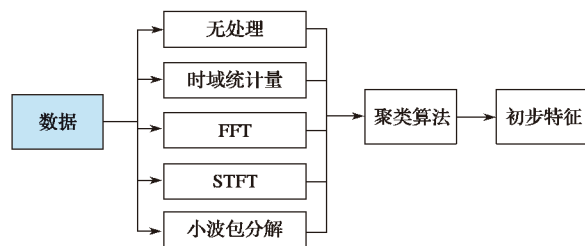


图7 初步特征提取流程

3.3.2 分类模型评估

混淆矩阵用 P 和 N 来表示样本的原标签以及样本经过模型分类后的标签的差异性，如表 3 所示。

设总类数为 n ，

表3 混淆矩阵

分类标签	P	N
P	TP	FN
N	FP	TN

第 i 类分类命中率为 p_i ，分类效果常用指标如表 4 所示。

计算步骤：

表4 分类评估指标

评估指标	计算公式
Accuracy	$\frac{TP + TN}{TP + TN + FP + PN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-score	$\frac{2}{\frac{1}{precision} + \frac{1}{recall}}$
Shoot	$\sum_{i=1}^n \frac{p_i}{n}$

步骤 1：初始化 $t=0$ ，样本集 L 。
步骤 2： $t=t+1$ ，对标签为 $L_1, L_2, \dots, L_n$ 的训练集，将 L_{i_t} 与 $L_1, L_2, \dots, L_{i_t-1}, L_{i_t+1}, \dots, L_n (i_t \in 1, 2, \dots, n)$ 分别标记为 Label1 和 Label2 进行训练，构造 GBDT 模型。

步骤 3：输入测试集样本，若测试样本被分在 Label1，则结果为 Label1。否则，若剩下的训练标签集 Label2 中标签类数只有一类，则结果为 Label2。

步骤 4：若 Label2 中原标签数大于 2，则重复步骤 2 与步骤 3，直到训练样本被分在 Label2 的最后一级。

递归二分类流程如图 8 所示。

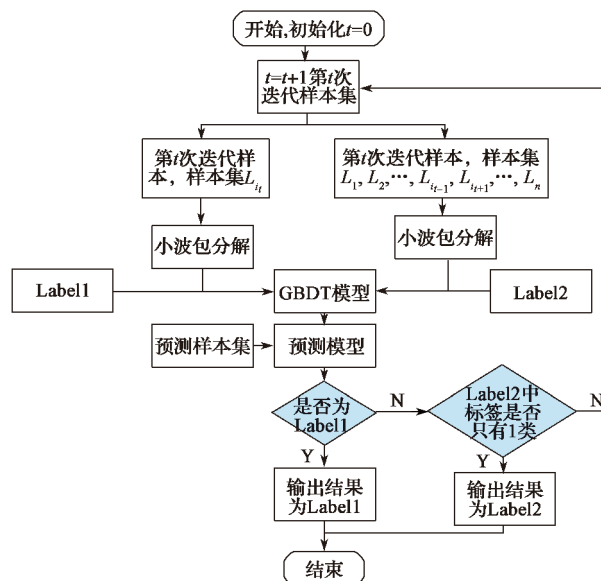


图8 递归二分类流程图

3.3.3 交叉验证

运用 Python 机器学习库 sklearn 中集成的 StratifiedKFold 函数, 将数据集按照标签比例分批提取一定数量的样本作为训练和测试集, 不断重复地验证模型在不同数据集上的精确度, 以此来测试模型的泛化性能。

4 试验验证

4.1 故障样本统计

故障样本统计如表 5 所示。

表 5 故障样本分类统计

故障定位标签	故障名称	样本数量
LO-TS-P	网压波动异常	223
LO-TS-TC-CM-4QS	四象限模块故障	78
LO-TS-TC-CM-INV	逆变模块故障	24
LO-TS-TC-DC	中间直流环节支撑电容故障	7
LO-TS-TC-BCC	直流回路故障	5
LO-TS-TC-RC	谐振电容故障	2
LO-TS-TC-TCU-LCC	插件异常	1

对数据文件概览, 统计信息特征如下:

①样本文件数量少。HX_{d1} 型机车牵引变流器四象限输入过流故障样本文件只有 340 个。

②样本文件数据维度高。在单个样本文件中共有 9 个变量, 每个变量对应的数据量有 17 500 个。

③不同标签类别的样本文件数量分布差异大。HX_{d1} 型机车的中间直流环节支撑电容故障 (LO-TS-TC-DC) 的样本数量仅为电网压异常 (LO-TS-P) 故障的 3.13%。做类别数大于 3 的分类时, 少数类样本将被分错, 导致数据成为了噪声, 这对整个系统的准确性指标影响较大, 此时的分类准确率将失去指导价值。

4.2 聚类进行特征初步确定

将数据样本进行各特征工程处理后聚类处理, 以评测数据经过各特征提取方法后的特征向量对数据原属性的表征程度。

设定聚类数目为 3~10, 对样本文件进行无处理、时域统计量、FFT 及 STFT 工程处理后, 经过 K-Means 聚类后的轮廓系数见表 6。

表 6 轮廓系数表

聚类数	无处理	时域统计量	FFT	STFT (win=4)	小波包分解 (DB25,LEV5)
3	0.24	0.33	0.56	0.51	0.62
4	0.23	0.34	0.51	0.42	0.61
5	0.24	0.38	0.43	0.42	0.60
6	0.25	0.41	0.44	0.42	0.56
7	0.25	0.43	0.45	0.37	0.42
8	0.23	0.43	0.39	0.32	0.41
9	0.27	0.44	0.38	0.31	0.41
10	0.22	0.45	0.28	0.32	0.38
平均值	0.24	0.40	0.43	0.39	0.50

分析表 6 可知, 在 8 层小波包分解下, 基函数为 DB25 平均轮廓系数达到了 0.50。小波包分解对数据有着比较好的表征, 初步确定特征提取方法为小波包分解。

4.3 三分类

4.3.1 小波基函数及分解层数

针对样本集每个样本取后 5 000 个采样点 (故障数据出现在数据段后 2/7 段), 对数目在前 3 的标签—牵引供电网压异常、四象限模块故障、逆变模块故障对应的样本集进行分类。设定小波基函数的集合为正交小波基 DB1~DB38, 分解层数为 1~9 层, 对 HX_{d1} 型机车变流器数据采样频率为 6 250 Hz, 对应的频域分辨率为 12.207~3 125 Hz。统计在各基函数与层数组合下的全类平均命中率 (Shoot), 如表 7 所示。

表 7 各基函数在不同层数组合下的命中率

基函数	1 层	2 层	3 层	4 层	5 层	6 层	7 层	8 层	9 层
DB1	0.680 5	0.638 6	0.653 2	0.664 3	0.629 5	0.659 3	0.751 0	0.645 6	0.644 6
DB2	0.634 7	0.690 2	0.619 4	0.649 6	0.642 2	0.665 8	0.682 1	0.599 9	0.610 0
DB3	0.694 7	0.640 6	0.624 9	0.656 4	0.600 7	0.647 5	0.656 8	0.618 0	0.640 7
DB4	0.669 8	0.708 4	0.689 7	0.634 6	0.658 6	0.658 4	0.631 7	0.580 3	0.602 5
DB5	0.702 0	0.681 0	0.676 9	0.706 9	0.695 8	0.629 6	0.653 8	0.671 0	0.641 1
DB6	0.625 4	0.705 5	0.636 1	0.686 4	0.675 2	0.622 7	0.684 4	0.634 1	0.690 3
DB7	0.597 4	0.680 6	0.682 4	0.664 5	0.666 0	0.652 3	0.592 8	0.615 4	0.656 4
DB8	0.611 6	0.743 9	0.701 9	0.685 3	0.721 5	0.697 7	0.636 0	0.604 6	0.698 9
DB9	0.668 4	0.675 2	0.636 1	0.731 2	0.718 7	0.655 9	0.672 3	0.649 7	0.661 8
DB10	0.705 9	0.702 6	0.675 1	0.666 1	0.702 7	0.661 6	0.688 4	0.647 4	0.629 6
DB11	0.683 9	0.688 8	0.715 5	0.674 9	0.706 9	0.681 5	0.654 5	0.659 5	0.661 0
DB12	0.629 1	0.676 6	0.661 3	0.705 2	0.697 1	0.656 2	0.617 7	0.648 2	0.614 4
DB13	0.622 0	0.660 2	0.682 2	0.716 0	0.694 5	0.672 5	0.665 6	0.638 8	0.642 1
DB14	0.666 7	0.684 6	0.633 7	0.678 4	0.699 1	0.699 5	0.633 5	0.666 6	0.647 6
DB15	0.638 2	0.654 8	0.662 1	0.694 7	0.720 0	0.705 6	0.636 9	0.667 6	0.651 5
DB16	0.618 3	0.663 8	0.630 1	0.672 7	0.668 1	0.691 3	0.681 1	0.628 2	0.642 6
DB17	0.682 3	0.752 0	0.598 1	0.655 5	0.668 3	0.702 5	0.658 1	0.685 9	0.643 8
DB18	0.678 5	0.653 1	0.674 3	0.704 0	0.736 3	0.700 1	0.651 7	0.646 6	0.612 9
DB19	0.655 8	0.646 0	0.655 9	0.628 6	0.762 7	0.688 9	0.669 6	0.712 5	0.655 0
DB20	0.678 2	0.722 7	0.686 8	0.655 0	0.651 1	0.692 5	0.644 4	0.614 5	0.653 9
DB21	0.639 7	0.660 9	0.652 1	0.719 8	0.692 8	0.688 9	0.653 0	0.628 2	0.648 1
DB22	0.685 1	0.693 2	0.610 2	0.693 5	0.731 9	0.678 1	0.697 5	0.638 4	0.628 3
DB23	0.644 5	0.720 1	0.663 6	0.708 0	0.693 2	0.707 6	0.716 8	0.660 1	0.648 2
DB24	0.658 6	0.708 6	0.696 5	0.724 7	0.719 5	0.693 6	0.644 1	0.675 4	0.662 6
DB25	0.664 3	0.687 8	0.661 6	0.688 8	0.769 2	0.715 9	0.675 0	0.631 1	0.706 8
DB26	0.663 0	0.697 6	0.630 2	0.736 7	0.682 1	0.683 2	0.703 0	0.604 9	0.616 5
DB27	0.624 5	0.673 0	0.678 8	0.655 3	0.667 5	0.708 4	0.638 5	0.638 7	0.620 7
DB28	0.645 7	0.660 6	0.637 8	0.690 6	0.692 7	0.680 7	0.631 7	0.667 5	0.676 2
DB29	0.639 1	0.683 3	0.697 5	0.701 7	0.725 6	0.714 7	0.664 8	0.660 7	0.608 0
DB30	0.622 4	0.680 4	0.633 9	0.670 5	0.635 5	0.662 8	0.703 7	0.632 9	0.655 0
DB31	0.664 3	0.668 1	0.641 2	0.692 1	0.673 4	0.691 2	0.722 8	0.646 4	0.663 7
DB32	0.657 6	0.729 7	0.749 9	0.740 8	0.687 5	0.681 7	0.706 5	0.672 5	0.719 8
DB33	0.672 3	0.690 5	0.672 9	0.694 3	0.629 5	0.678 2	0.660 7	0.647 3	0.653 1
DB34	0.611 8	0.653 2	0.673 8	0.703 5	0.661 3	0.682 4	0.681 7	0.660 2	0.657 3
DB35	0.650 1	0.713 4	0.678 1	0.658 3	0.701 3	0.699 0	0.688 1	0.682 1	0.723 6
DB36	0.650 1	0.693 4	0.667 5	0.690 4	0.676 3	0.724 6	0.714 1	0.658 3	0.713 7
DB37	0.617 4	0.660 4	0.620 9	0.643 7	0.636 8	0.655 2	0.638 1	0.667 5	0.690 1
DB38	0.638 2	0.659 6	0.663 9	0.659 5	0.690 0	0.690 0	0.736 3	0.698 7	0.705 0

当基函数为 DB25，分解层数为 5 时，Shoot 指标最高接近 0.77。

4.3.2 三分类结果

设定基函数为 DB25，分解层数为 5，对牵引供电网压异常、四象限模块故障、逆变模块故障进行交叉验证，各指标统计结果如表 8 所示。

表 8 指标统计结果

标签	1	2	3	4	5	Shoot
牵引供电网压	1	1	1	0.93	0.95	0.98
四象限模块	0.88	0.69	0.81	0.93	0.73	0.80
逆变模块	0.40	0.20	0.60	0.80	0.25	0.45
Acc	0.92	0.86	0.92	0.92	0.86	0.89

逆变模块的 Shoot 指标仅为 45%。虽然此时的 Acc 指标达到了 89%，但已经失去了指导意义，因为大部分的逆变模块样本被分错。

4.4 递归二分类

4.4.1 供电网压异常与变流器模块故障二分类

采用递归二分类法，对牵引供电网压异常、四象限模块故障、逆变模块故障对应的样本集进行分类。

四象限模块故障与逆变模块故障均为变流器模块故障，在首次分类中，将这两类模块故障标签合并，与供电网压异常进行分类。

时域统计量特征交叉验证结果如表 9 所示。

表 9 交叉验证结果表

标签	1	2	3	4	5	Shoot
牵引供电网压	1	1	0.93	1	0.95	0.98
四象限模块 + 逆变模块	0.90	0.95	1	1	1	0.97
Acc	0.97	0.98	0.95	1	0.96	0.97

Acc 平均指标达到了 97%，Shoot 平均指标也达到了 97%。仅用时域统计量就能比较好地将供电网压异常与其余两个变流器模块故障区别开来。

将四象限模块故障与逆变模块故障进行分类，时域统计量特征交叉验证结果如表 10 所示。

表 10 交叉验证结果表

标签	1	2	3	4	5	Shoot
牵引供电网压	0.94	0.88	0.63	0.93	0.93	0.86
四象限模块 + 逆变模块	0	0	0.60	0.60	0.25	0.29
Acc	0.71	0.67	0.62	0.85	0.79	0.72

牵引供电网压异常和模块故障的二分类 Acc 平均指标仅有 72%，Shoot 平均指标仅有 29%。

4.4.2 四象限模块故障与逆变模块故障二分类

针对数据样本集，采用的小波包分解基函数 (DB1~DB38) 和分解层数 (1~9) 试验参数集，频域分辨率为 12.207~3 125 Hz，统计逆变模块 Shoot ≥ 0.7 的小波参数组合见表 11。

表 11 小波参数组合表

小波基	层数	Acc 平均值	Shoot 平均值
DB5	1	0.80	0.71
DB18	5	0.82	0.75
DB37	7	0.77	0.73
DB35	8	0.78	0.69
DB37	9	0.80	0.73

当基函数为 DB18，分解层数为 5 时，Acc 与 Shoot 指标都达到了最高。该参数组合 5 次交叉验证的详细结果见表 12。

表 12 交叉验证结果表

标签	1	2	3	4	5	Shoot
四象限模块	0.75	0.81	0.88	0.93	1	0.87
逆变模块	0.60	0.60	0.80	0.60	0.50	0.62
Acc	0.71	0.76	0.86	0.85	0.90	0.75

结合牵引供电网压异常与变流器模块故障二分类指标，得到最终递归二分类的结果如表 13 所示。

表 13 递归二分类结果表

指标	Acc 平均值		Shoot 平均值	
	三分类	递归二分类	三分类	递归二分类
牵引供电网压	0.96	0.98	0.98	0.98
四象限模块	0.79	0.84	0.81	0.85
逆变模块	0.55	0.78	0.45	0.80

四象限模块故障的 Acc 指标和 Shoot 指标分别从 79%、81% 提升至 84%、85%。逆变模块故障的 Acc 指标和 Shoot 指标分别从 55%、45% 提升至 78%、80%。

5 总结

本文主要介绍了小波包分解和梯度提升决策树在 HXD1 型机车四象限输入过流故障诊断中的应用，对于四象限过流故障中常见的牵引供电网压异常、四象限模块故障与逆变器模块故障这 3 种故障原因进行了初步分析，并提出一种基于小波包分解的特征提取，结合集成学习模型—梯度提升决策树，对这 3 种故障进行递归二分类。

本文对于牵引供电网压异常和变流器模块故障，采用了时域统计量的方法来进行二分类，对于变流器模块内部的四象限模块与逆变器故障，采用基于正交小波基函数 DB18，分解层数为 5 的小波包分解。通过对实际数据进行试验验证，基于小波包分解和集成学习模型的递归二分类法优于三分类法、时域统计量特征的递归二分类法，能对四象限输入过流故障样本数据进行比较准确的故障定位。

随着数据量快速增长，运用机器学习的方法对机车进行故障诊断，不仅使故障诊断变得更加高效，而且能减轻机车售后运用维护的成本，提高经济效益。

(下转第 138 页)

套动力单元同时运行,完成不同挡位运行试验,冷却水温和液力变速箱油温符合设计要求。

2) 司机室噪声测试

经过试验,在司机室在门窗密闭的情况下,检测车自行速度为 120 km/h,柴油机以额定功率输出时,司机室中央位置处稳态噪声为 76.5 dB(A),小于规定的 88 dB(A);司机座位上方 1 m 处稳态噪声为 73.2 dB(A),小于相关规定的 90 dB(A)。司机室噪声符合 TB/T 2180—2006《电气化铁道接触网综合检修作业车技术条件》中的相关规定。

5.4 检测系统低温试验

2016 年 11 月检测系统完成了低温性能试验,试验项目包括低温工作、低温贮存 2 项,试验温度为 $(-40 \pm 2)^\circ\text{C}$ 。试验过程前、中、后,样品外观无变形异常现象,且能正常工作。

5.5 试运行考核试验

2016 年 10 月至 2017 年 1 月,在西宁供电段完成运用考核,其运行里程累计超过 1 300 km,作业时间超过 40 h。运用考核结果表明,检测车整车技术状态正常,检测系统各项功能运用良好,整车主要功能及主要部件的质量能满足高原地区、高寒条件下的运用要求。

2017 年 2 月底,检测车通过了中国中车组织的技术评审。

截至 2018 年 12 月底,检测车运行里程累计超过 8 000 km,作业时间超过 150 h,运用情况良好。

6 结语

高速铁路接触网检测车的研发以智能化接触网检

测车为基础,并针对青藏铁路兰新高速铁路线路高海拔的特殊环境条件进行研发设计,经车体静强度试验、整车型式试验、检测系统低温试验、高原试运行试验和兰新高速铁路段实际运用的验证,整车满足高海拔、低气温等特殊环境的 120 km/h 速度等级的运行和接触网智能化检测、监测的需求。

参考文献:

- [1] 赵国栋.智能化接触网检测车[J].今日轨道交通,2014(9): 83-89.
- [2] 孙竹生.内燃机车总体及机车走行部[M].北京:中国铁道出版社,1984.
- [3] 张成贵,蒋世明.内燃机车总体[M].北京:中国铁道出版社,1982.
- [4] 铁道部运输局供电部,中国铁道科学研究院,西南交通大学.高速铁路供电安全检测监测系统(6C系统)总体技术规范:铁运[2012]136号[S].北京:中华人民共和国铁道部,2012.
- [5] 中国铁路总公司.接触网悬挂状态检测监测装置(4C)暂行技术条件:TJ/GD 006—2014[S].北京:中国铁道出版社,2014.
- [6] 国家铁路产品质量监督检验中心.CT21000型车体检测报告[R].北京:国家铁路产品质量监督检验中心,2016.
- [7] 国家铁路产品质量监督检验中心.接触网检测车(JX300型)检验报告[R].北京:国家铁路产品质量监督检验中心,2016.
- [8] 国家铁路产品质量监督检验中心.接触网检测车(JX300 II型)检测报告[R].北京:国家铁路产品质量监督检验中心,2016.

作者简介:王文聪(1990—),女,现从事轨道工程机械总体设计工作。

(上接第 79 页) 相关故障诊断方法已部署于朔黄机务段地面诊断系统,极大地提高了地面检修工作效率及机车运行效率。

参考文献:

- [1] 吴勇忠.浅析HX₀₁型电力机车牵引变流器突出故障与处理[J].科技展望,2016,26(33):28.
- [2] 戴晨曦,刘志刚,胡珂珏.关于高铁CRH2型动车组逆变器故障诊断研究[J].计算机仿真,2016,33(4):217-222.
- [3] 张雅琴,滕青芳,李国飞,等.基于电流残差的逆变器开路故障诊断方法[J].兰州交通大学学报,2016,35(3):112-118.
- [4] 陈怡恬,陈特放,成庶,等.基于输入电流的牵引变流器开关管开路故障诊断方法[J].电气工程学报,2016,11(5):11-20.
- [5] 陈特放,钟艳科.基于小波分析和SVM的主变流器故障诊断[J].机车电传动,2009(1):57-59.
- [6] 李玉超,田永洙,高沁翔.基于BP神经网络的200 km/h牵引变流器故障诊断技术的研究[J].铁道机车车辆,2007,27(2):4-7.
- [7] 李全林,何忠韬.SOM神经网络在电力机车主变流器故障诊断中的应用[J].铁道机车车辆,2009,29(2):13-16.
- [8] 魏志磊,苏宏升.基于遗传神经网络的电力机车主变流器故障诊断[J].电子质量,2009(12):39-42.
- [9] 岳生吉,王焕民,牛彩云.基于粗糙集的机车牵引变流器故障诊断方法[J].铁路计算机应用,2012,21(3):62-65.
- [10] 朱琴跃,叶双挺,谭喜堂,等.基于改进蚁群神经网络的牵引逆变器故障诊断[J].机电一体化,2014,20(11):52-57.
- [11] MALLAT S, HWANG W L. Singularity detection and processing with wavelets[J]. IEEE Transactions on Information Theory, 1992, 38(2): 617-643.
- [12] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189-1232.

作者简介:王建华(1982—),男,工程师,长期从事电力机车运用检修和技术管理工作。