基于多源数据融合的光伏阵列复合故障诊断与全生命周期治理研究

代金超，魏　浩，马洪敏

（唐山学院，河北　唐山　063000）

摘　要： 针对光伏阵列复合故障诊断中单一数据源难以全面反映复杂运行状态、传统方法诊断精度有限，以及治理方案缺乏成本与可靠性协同优化的问题，提出一种基于多源数据融合的复合故障诊断与全生命周期治理方法。通过小波分解提取电流、电压信号时频特征，并对多源特征进行融合，采用支持向量机实现复合故障分类；综合考虑维修、停机及预防性维护成本与系统可靠性，利用布谷鸟算法优化全生命周期治理方案。实验结果表明，在由5个子阵列、1 000块组件构成的光伏系统中，该方法能够较准确地识别不同运行状态及故障类型，且受试者工作特征曲线整体优于两种对比方法，表现出较好的故障诊断性能。

关键词： 光伏阵列；复合故障诊断；多源数据融合；小波分解；支持向量机

中图分类号： TM615　　**文献标识码：** A

Multi-source data fusion-based composite fault diagnosis and lifecycle management for photovoltaic array

DAI Jinchao, WEI Hao, MA Hongmin

(Tangshan University, Tangshan, Hebei 063000, China)

Abstract: To address the problems that a single data source cannot comprehensively reflect complex operating states in composite fault diagnosis of a photovoltaic array, conventional methods have limited diagnostic accuracy, and management schemes lack coordinated optimization of cost and reliability, a composite fault diagnosis and lifecycle management method based on multi-source data fusion is proposed. Wavelet decomposition is used to extract time-frequency features from current and voltage signals, and multi-source features are fused. A support vector machine is then employed to classify composite faults. Considering maintenance cost, downtime cost, preventive maintenance cost, and system reliability, the cuckoo search algorithm is used to optimize the lifecycle management scheme. Experimental results show that, in a photovoltaic system consisting of five subarrays and 1 000 modules, the proposed method achieves a fault diagnosis misclassification rate of 0.21% and a true positive rate of 0.90, outperforming the two comparison methods, whose true positive rates are 0.65 and 0.74, respectively. The results demonstrate that the proposed method can effectively improve the diagnostic accuracy of composite faults in a photovoltaic array.

Key words: photovoltaic array, composite fault diagnosis, multi-source data fusion, wavelet decomposition, support vector machine

1　引言

光伏阵列是光伏发电系统的重要组成部分，其运行状态直接影响系统发电效率与运行可靠性。由于长期暴露于复杂自然环境中，光伏阵列易受温度、湿度及辐照变化等因素影响而产生不同类型故障。当多种故障同时发生时，不同故障特征相互耦合，会进一步增加状态识别与故障诊断难度。因此，提高光伏阵列复合故障诊断精度，并结合诊断结果优化后续治理方案具有重要意义。

目前，已有较多研究将不同算法应用于光伏阵列故障诊断。蔡煌钰等[1]将正常运行及不同故障状态下的电气参数作为源域数据，利用迁移学习实现了故障特征迁移，但该方法对目标域特征的适应能力仍有提

基金项目：唐山市社会科学界联合会立项课题（TSSKL2025-203）；唐山学院“揭榜挂帅”项目（2025BSC002）

升空间。凌文青[2]对光伏阵列图像和电气信号进行预处理，利用卷积神经网络完成了特征提取与融合，但该方法在训练样本有限时易受噪声和异常值影响。李睿智等[3]利用关联规则和聚类分析挖掘不同参数之间的关系，结合规则库实现了故障诊断，但该方法对复杂故障特征的提取能力相对有限。赵军等[4]通过相关性分析筛选关键故障特征，采用机器学习算法完成了模型训练与验证，但该方法在大规模数据条件下仍存在过拟合风险。

总体来看，现有方法在光伏阵列故障识别方面取得了一定成果，但对多源数据关联信息的利用仍不充分，且研究多集中于故障诊断，对诊断后的治理决策以及全生命周期成本与可靠性协同优化关注较少。基于此，本文围绕光伏阵列复合故障识别与治理方案优化展开研究，通过融合多源运行信息提升复杂故障状态下的识别能力，并结合诊断结果对治理方案进行优化，以兼顾运维成本与系统可靠性，为光伏阵列复合故障诊断与全生命周期治理提供参考。

2 光伏阵列复合故障诊断与全生命周期治理方法

2.1 光伏阵列电气信号多层分解

为提取光伏阵列不同运行状态下的电气信号特征，采用小波分解方法对采集的电流、电压信号进行多层处理，通过低通和高通滤波获得不同尺度下的近似系数与细节系数，实现信号在不同频段上的分解。

本文利用高精度传感器采集光伏阵列的电流和电压信号，得到离散电流信号序列$I[n]$与电压信号序列$V[n]$，其中n表示采样点数[5]。设$s[n]$为待分解的电流信号或电压信号，$h[k]$和$g[k]$分别为低通滤波器系数与高通滤波器系数，则第一层分解表示为：

$$cA_1[n]=\sum_{k=0}^{L-1}h[k]s[2n-k] \tag{1}$$

$$cD_1[n]=\sum_{k=0}^{L-1}g[k]s[2n-k] \tag{2}$$

其中，$cA_1[n]$为为第一层分解得到的近似系数序列，用于表征电气信号的低频变化特征；L为滤波器长度；$cD_1[n]$为第一层分解得到的细节系数序列，用于表征信号中的突变和瞬态特征，可反映局部故障引起的高频变化[6]。在此基础上，对近似系数继续进行逐层分解，可获得不同尺度下的近似系数与细节系数，共同构成光伏阵列电气信号的时频特征。

2.2 光伏阵列复合故障分类诊断

本文对电流、电压信号经多层分解得到的时频特征进行融合，并基于融合特征构建支持向量机分类模型，通过核函数映射实现光伏阵列运行状态及故障类型的分类诊断。

设电流信号$I[n]$对应的第j层近似系数和细节系数分别为$cA_j^I[n]$与$cD_j^I[n]$，电压信号$V[n]$对应的第j层近似系数和细节系数分别为$cA_j^V[n]$与$cD_j^V[n]$。采用加权融合方式对电流和电压特征进行整合，多源数据融合后的综合特征向量$\boldsymbol{F}[n]$表达式为：

$$\boldsymbol{F}[n]=\alpha\sum_{j=1}^{J}\left(w_{A_j}cA_j^I[n]+w_{D_j}cD_j^I[n]\right)+\beta\sum_{j=1}^{J}\left(w_{A_j}cA_j^V[n]+w_{D_j}cD_j^V[n]\right) \tag{3}$$

其中，w_{A_j}和w_{D_j}分别为第j层近似系数与细节系数对应的融合权重；α和β分别为电流特征权重与电压特征权重。基于融合后的特征向量，采用支持向量机构建分类决策函数，通过寻找最优超平面最大化不同类别样本之间的间隔，实现光伏阵列运行状态分类[7]。具体原理如图1所示。

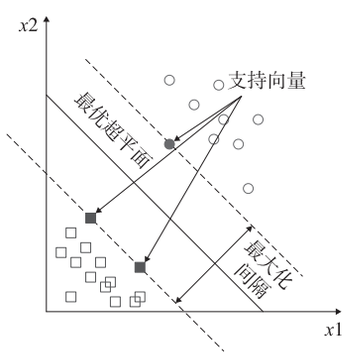


图1 光伏阵列复合故障支持向量机诊断原理

考虑到光伏阵列复合故障特征在原始特征空间中具有线性不可分特性，本文引入核函数，将低维特征映射至高维特征空间。决策函数表示为：

$$f(\boldsymbol{F})=\mathrm{sgn}\left(\sum_{i=1}^{N}\alpha_i y_i K(\boldsymbol{F}_i,\boldsymbol{F})+b\right) \tag{4}$$

其中，$\boldsymbol{F}$为待分类样本的融合特征向量；$\boldsymbol{F}_i$为第i个训练样本的融合特征向量；N为训练样本数量；α_i为拉格朗日乘子；y_i为第i个训练样本的类别标签；b为偏置项。将待诊断样本的融合特征向量代入决策函数，当$f(\boldsymbol{F})\geqslant 0$时判定为正常状态，当$f(\boldsymbol{F})<0$时判定为故障状态。在此基础上，根据不同故障类别对应分类器的输出结果进一步确定具体故障类型。

2.3 光伏阵列复合故障治理方案优化

综合考虑治理方案的全生命周期成本与系统可靠性，构建治理方案适应度函数，并采用布谷鸟搜索算法对决策变量进行优化，通过莱维飞行更新个体位置，在复杂方案空间中搜索较优治理方案。

首先，对光伏阵列复合故障治理方案进行编码，将每个治理方案视为一个“鸟巢”。治理方案由多个决策变量组成，包括维修方式m_i、预防性维护频率等[8]。将各决策变量组合形成治理方案决策向量$\boldsymbol{X}$，并构建适应

度函数。该函数综合考虑光伏阵列全生命周期成本与可靠性，表示为：

$$P(\boldsymbol{X})=-[C_{\text{total}}(\boldsymbol{X})+\lambda\cdot(1-R_{\text{total}}(\boldsymbol{X}))] \tag{5}$$

$$R_{\text{total}}(\boldsymbol{X})=\prod_{i=1}^{k}(1-P_{\text{fault}}(t_i)z_i)\cdot\eta(m_i) \tag{6}$$

其中，$C_{\text{total}}(\boldsymbol{X})$为治理方案对应的光伏阵列全生命周期总成本，包括维修成本、停机损失成本和预防性维护成本；$R_{\text{total}}(\boldsymbol{X})$为相应的全生命周期可靠性；$P_{\text{fault}}(t_i)$为第$i$类故障在维修时间间隔$t_i$内发生的概率；$z_i$为第$i$类故障对应的状态变量；$\eta(m_i)$为第$i$类故障采用维修方式$m_i$后的修复效果系数；$\lambda$为成本与可靠性之间的权重系数；$k$为故障类型数量。适应度值越大，表明治理方案的综合性能越优。

在布谷鸟搜索过程中，利用莱维飞行更新个体位置，其表达式为：

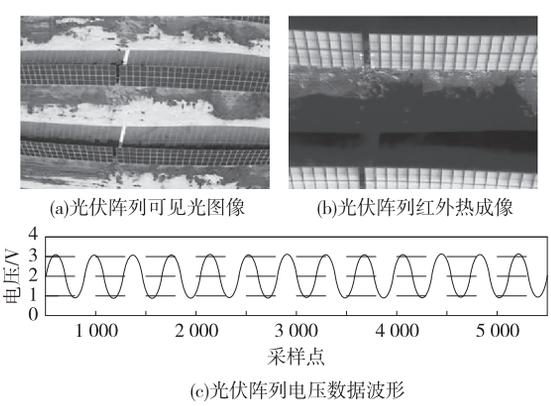
$$\boldsymbol{X}_i^{\text{new}}=\boldsymbol{X}_i+\delta\otimes L(\mu) \tag{7}$$

其中，δ为步长控制因子；$L(\mu)$为莱维飞行的概率密度函数；$\boldsymbol{X}_i$和$\boldsymbol{X}_i^{\text{new}}$分别为更新前后的治理方案决策向量。通过迭代更新种群位置并比较各治理方案的适应度，最终选择适应度最高的个体，其对应方案作为光伏阵列复合故障治理的优化结果。

3　实验与结果分析

3.1　实验设置

选取由5个子阵列构成的光伏阵列作为研究对象，每个子阵列包含20串光伏组件，整体共计1 000块光伏组件，并配置5台逆变器，用于模拟不同功率转换及运行状态。分别在各子阵列直流侧输出线路和逆变器交流侧输出线路安装电流传感器，采集不同位置的电流信号；在直流侧和交流侧输出端安装电压传感器，以获取对应位置的电压数据。部分采集数据如图2所示。



(a)光伏阵列可见光图像　(b)光伏阵列红外热成像

(c)光伏阵列电压数据波形

图2　光伏阵列多源数据示例

为验证本文方法的故障诊断性能，采用支持向量机对光伏阵列运行状态进行分类，其参数配置如表1所列。同时，选取文献[1]方法和文献[2]方法作为对比方法，并在相同数据集和实验条件下进行测试。

表1　支持向量机参数配置

参数名称	取值
核函数类型	径向基函数
惩罚参数	10
核函数参数	0.1
最大迭代次数	1 000
收敛精度	1×10^{-3}
类别权重	{0:1,1:2}（0表示正常，1表示故障）

3.2　复合故障诊断结果

基于支持向量机对光伏阵列运行状态进行分类，所得诊断结果的混淆矩阵如图3所示。

实际/输出	正常	组件短路故障	组件开路故障	局部阴影故障	逆变器故障
正常	198		1	1	1
组件短路故障	3	195	2	1	1
组件开路故障	2	1	179	1	1
局部阴影故障	1	2	1	196	2
逆变器故障	2	1	1	3	195

图3　光伏阵列故障诊断混淆矩阵

由图3可知，各类样本主要集中于混淆矩阵对角线位置，表明本文方法能够较准确地识别不同运行状态及故障类型。其中，不同类别之间仍存在少量误分类，但总体分类结果较为稳定，体现出较好的复合故障识别能力。

3.3　不同方法诊断性能对比

为评价本文方法相较于现有方法的故障诊断性能，绘制本文方法、文献[1]方法和文献[2]方法的受试者工作特征曲线，结果如图4所示。

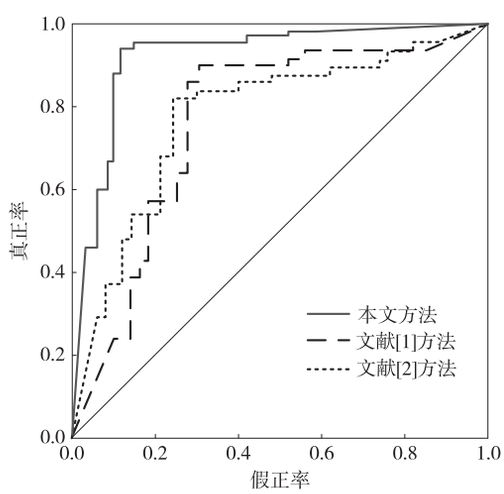


图4　不同诊断方法的受试者工作特征曲线

由图4可知，本文方法的受试者工作特征曲线整体

（下转第273页）