

融合多源监测数据的贝叶斯网络光伏阵列故障诊断方法优化研究

喻 聪 龚伟红 念 阳

中国电建集团贵州工程有限公司 贵州 贵阳 550003

摘 要：中东沙漠地区光伏电站长期处于极端高温、强沙尘和强紫外线环境，光伏阵列易发生沙尘遮蔽、局部热斑、组件隐裂等单一故障及电气-热学复合故障。现有方法多依赖单一监测数据，复合故障识别能力不足；传统全连接贝叶斯网络还存在节点关联冗余、推理耗时较长的问题。针对上述不足，本文提出融合红外温度、组串电流和逆变器运行参数的改进贝叶斯网络诊断方法。首先通过异常剔除、降噪、时序对齐和归一化完成异构数据预处理；其次依据光伏系统电热耦合机理重构网络拓扑，剔除无物理依据的冗余路径，并结合最大似然估计与专家先验知识对条件概率表进行自适应校正。基于沙特某 1000MW 沙漠光伏电站连续 1 年采集的 200 组实测样本验证，优化模型对沙尘诱发复合故障的诊断准确率为 94.7%，综合平均准确率为 94.9%；单次诊断收敛时间缩短 40.1%，并发数据处理吞吐量提升 68%。结果表明，该方法能够兼顾诊断准确性与实时性，可为大型沙漠光伏电站智能运维提供技术参考。

关键词：沙漠光伏；多源监测数据；贝叶斯网络；复合故障诊断；条件概率表

引言

中东沙漠地区光照资源丰富，已成为大型光伏电站集中建设区域^[1]。光伏组件长期暴露在户外，极端高温与强紫外线会加速封装材料老化，高频沙尘则会引起表面积尘遮蔽、局部热斑和组件隐裂。上述缺陷既可能独立出现，也可能因温升、功率失配和局部电流集中相互耦合，形成征兆交叠的电气-热学复合故障，增加现场定位和运维处置难度。

现有基于单一红外图像或单一电气参数的诊断方法，可对热斑或电流异常进行识别，但难以同时捕捉复合故障的热学与电气耦合特征，识别精度和泛化能力仍受限制^[2-3]。贝叶斯网络能够利用先验知识和监测证据开展不确定性推理，适合处理多故障、多征兆并存的场景；但传统光伏诊断模型常存在输入征兆单一、全连接拓扑冗余、推理搜索空间大等问题，难以支撑 GW 级光伏电站海量监测数据的并发实时诊断。

因此，本文以典型单一故障和复合故障为对象，融合红外温度、组串电流和逆变器输出三类异构数据，从拓扑结构与条件概率赋值两方面优化贝叶斯网络，构建“数据采集-预处理-模型推理-故障判定”全流程诊断框架，并利用实测数据验证模型的准确性、计算效率及应用边界。其中，拓扑优化用于减少无效计算，概率参数校正用于增强模型对传感器漂移和工况变化的适应性。

1 沙尘环境下光伏阵列故障特征与多源监测体系构建

1.1 典型故障类型及多维度征兆关联分析

结合沙特沙漠光伏电站运行数据和现场运维经验，选取沙尘大面积遮蔽、局部热斑、组件隐裂、逆变器 MPPT 异常及沙尘遮蔽+组件隐裂复合故障五类典型故障。不同故障的主导征兆存在差异：积尘通常造成组串电流和功率同步下降，热斑表现为局部温升，隐裂会导致输出波动与局部温差，MPPT 异常则主要反映为逆变器输出偏离。多传感器协同监测有助于补足单一征兆的信息缺失^[4]。据此，将红外温度、组串电流和逆变器输出设置为观测节点，把故障状态设置为目标节点，建立故障类型与多源征兆的映射关系。

1.2 多源监测数据预处理方法设计

三类监测数据在采集频率、量纲和噪声来源上存在明显差异。多源数据融合研究^[5]表明，若直接拼接，时钟偏移、传感器漂移和异常尖峰会形成伪相关，扭曲节点条件概率。为此，设计分层式预处理流程（见图 1），依次完成异常剔除、降噪、时序对齐、基准校准和归一化；预处理结果按统一时间戳写入样本矩阵，并保留数据质量标记，以提高模型输入的一致性和可追溯性。



图 1-预处理流程

各环节具体处理规则如下:

红外温度数据:采用阈值法剔除强光反射、拍摄角度变化和设备抖动造成的畸变值,以同一阵列相邻组件温度作为局部基准,并通过极值归一化统一量纲;

组串电流数据:采用滑动窗口均值滤波抑制瞬时波动,窗口宽度依据采样频率和正常波动周期确定,并结合标准工况完成电流基准值校准;

逆变器输出数据:以组串电流时序为基准完成功率、效率和运行状态量的全局时序对齐,并通过相邻时刻变化率与 3σ 统计阈值联合识别、剔除随机离群点。

经过预处理后,原始数据噪声与异常得到有效抑制,数据一致性和可靠性满足贝叶斯网络输入要求。

2 融合多源数据的贝叶斯网络诊断模型优化设计

2.1 贝叶斯网络基本原理

贝叶斯网络是一种基于图论与概率论的有向无环图模型,由网络节点和有向边构成。节点表示故障状态或监测征兆,有向边描述变量之间的条件概率依赖关系。相关多源信息故障诊断研究表明,该模型能够在监测数据不完备时综合样本统计结果与专家经验^[6]。

设网络包含 n 个随机变量 $X_1, X_2, \dots, X_n$, 根据贝叶斯链式法则与条件独立性假设,网络联合概率分布可表示为:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i))$$

式中:

X_i 为第 i 个网络节点; $Pa(X_i)$ 为节点 X_i 对应的父节点集合;

$P(X_i | Pa(X_i))$ 为节点条件概率,由条件概率表(CPT)量化表达。

2.2 贝叶斯网络拓扑结构优化

传统光伏故障诊断贝叶斯网络多采用全连接拓扑,主要依靠数据相关性构建网络,未充分考虑系统物理机理,因此形成大量缺乏实际意义的有向边。冗余路径会扩大推理搜索空间、增加计算耗时,不利于海量数据的实时诊断。

本文依据光伏系统电热传导和电气运行逻辑重构拓扑:故障层设置五类状态节点,征兆层设置温度、电流和逆变器参数节点;仅保留具有明确因果方向的故障-征兆关联,删除“逆变器输出直接关联红外温度”等缺乏物理依据的冗余路径,并避免在同层观测节点之间设置循环依赖。

优化前全连接网络含 48 条有向边,优化后减至 32 条,冗余路径减少 33.3%。在不破坏故障-征兆有效概率依赖关系的前提下,网络结构得到简化,推理空间明显压缩。

2.3 节点条件概率表自适应调整方法

条件概率表(CPT)直接影响贝叶斯网络的推理精度。沙漠环境下传感器长期受高温、沙尘干扰,监测数据存在持续性波动,固定 CPT 对现场工况变化的适应性不足。

本文采用最大似然估计(MLE)与专家先验相结合的策略^[7],实现 CPT 自适应调整:样本数量充足的节点以统计结果为主,稀缺故障节点则提高专家先验的约束权重,避免小样本导致概率估计过度偏移。

首先利用历史带标签故障样本,通过最大似然估计获得各节点的初始条件概率;

再结合现场运维经验和区域故障统计规律修正初始概率,形成适配沙漠工况的动态 CPT;

模型运行过程中依据实时监测数据迭代更新概率参数,实现持续校正。

该策略可减弱传感器噪声和工况波动对诊断结果的影响,提高复杂现场环境下的诊断稳定性。

3 模型训练与对比验证实验

选取沙特某 1000MW 级大型沙漠光伏电站连续 1 年采集的 200 组真实运行数据作为样本,覆盖正常运行、四类单一故障和一类复合故障。样本按 7:3 随机划分为训练集和测试集,并设置分别以单一红外温度、组串电流或逆变器输出为征兆的贝叶斯网络作为对照模型;评价指标包括各类故障准确率、平均准确率、单次诊断收敛时间和并发吞吐量。训练与测试使用相同的时间窗口和故障标签规则,测试集仅用于最终评价,以保证不同模型之间的样本边界一致。

初步测试表明,优化模型对沙尘诱发复合故障的诊断准确率达到 94.7%,明显高于单征兆模型;同时受益于拓扑冗余剔除,平均诊断时间缩短约 40%,具备 GW 级电站海量数据并发诊断能力。

4 计算结果及分析

4.1 诊断准确率对比分析

200 组样本随机划分为训练集 140 组和测试集 60 组。对照模型分别以红外温度、组串电流和逆变器输出作为单一征兆,优化模型采用三类数据融合输入,各模型诊断结果见表 1。

表 1 不同模型对典型故障类型的诊断准确率对比

故障类型	温度	电流	逆变	优化模型
沙尘大面积遮蔽	58.3%	63.2%	55.0%	96.2%
局部热斑	72.4%	48.7%	61.5%	93.5%
组件隐裂	45.2%	71.8%	52.9%	94.1%
沙尘 + 隐裂复合故障	51.6%	54.8%	60.2%	94.7%
逆变器 MPPT 异常	41.3%	67.5%	82.4%	95.8%
平均准确率	53.8%	61.2%	62.4%	94.9%

由表 1 可知，单特征模型在不同故障类型下表现差异明显，平均准确率仅为 53.8%~62.4%，尤其难以识别沙尘 + 隐裂复合故障。其原因在于单一维度数据无法同时表征红外热学异常与电流、逆变器输出等电气特征。优化模型通过多源数据融合实现特征互补，复合故障准确率提高至 94.7%，平均准确率达到 94.9%，说明该方法能够显著增强沙漠环境下的故障识别能力。

4.2 计算效率与收敛性能

传统全连接拓扑存在大量冗余关联，推理计算量较大。依据光伏阵列电气-热传导逻辑删除无物理支撑的路径后，网络边数由 48 条降至 32 条，冗余消除 33.3%，在保留有效关联的同时显著缩小推理搜索空间。

在相同硬件环境和测试样本下，优化模型平均单次诊断收敛时间由 187.5ms 降至 112.3ms，缩短 40.1%；模拟 200 个组串同时诊断时，数据处理吞吐量提高至传统模型的 1.68 倍，能够满足 GW 级大型光伏电站实时运维需求。对比测试统一处理器线程数、内存配置和采样窗口，计时范围仅覆盖证据输入、后验推理与故障标签输出，不包含现场数据采集和网络传输时间。

4.3 误差溯源分析

60 组测试样本中，优化模型正确诊断 57 组，误判 3 组，误判率为 5.0%。对误判样本逐一回溯原始红外图、组串电流和逆变器日志后发现，3 组误判均发生在清晨或黄昏的极弱光区间，辐照度低于 150W/m²，且不存在设备通信中断或标签错配。

弱光条件下，沙尘造成的组件红外温差不足 2℃，组串电流幅值较低，叠加逆变器弱光控制策略后 SNR 低于 5dB，导致故障热学和电气特征被噪声削弱。

此时观测节点概率分布趋于平滑，不同故障的后验概率差值缩小，模型难以稳定区分轻微沙尘遮蔽与组件正常弱光运行。

因此，本模型在辐照度≥150W/m²的常规光照工况下

适用性较好；在极弱光环境下诊断可靠性有所下降。工程部署时可将低辐照度状态设为诊断置信度修正条件，对低置信样本触发延时复核。后续可增加辐照度观测节点，引入自适应滤波等方法强化弱光信号，并通过时序状态模型提高连续数据中的故障辨识能力。

5 结论

针对沙漠环境下光伏阵列复合故障多发、传统诊断方法精度与实时性不足的问题，本文得到以下结论：

构建了红外温度、组串电流和逆变器输出三位一体的多源监测体系及分层预处理流程，可有效降低异构数据噪声、时序错位和量纲差异影响，表征故障电热耦合特征；

依据光伏系统物理机理重构贝叶斯网络拓扑，并利用最大似然估计与专家知识自适应校正 CPT，网络冗余路径减少 33.3%，兼顾推理效率与抗干扰能力；

实测验证表明，优化模型综合诊断准确率为 94.9%，沙尘诱发复合故障准确率为 94.7%，单次诊断耗时缩短 40.1%，高并发处理能力提升 68%；

辐照度低于 150W/m² 时，弱光会削弱热学和电气故障特征，模型诊断可靠性下降，后续需进一步加强弱光特征约束与时序建模。

研究结果表明，该方法可为高温、高沙尘等极端气候区域大型光伏电站的数字化、智能化运维提供技术参考。

参考文献

[1] 王小宇, 刘波, 孙凯, 等. 光伏阵列故障诊断技术综述 [J]. 电工技术学报, 2024, 39(20): 6526-6543.

[2] 刘圣洋, 冬雷, 王晓晓, 等. 基于高斯核模糊 C 均值聚类的光伏阵列故障诊断方法 [J]. 太阳能学报, 2021, 42(5): 286-294.

[3] 李光辉, 段晨东, 武珊. 基于半监督机器学习法的光伏阵列故障诊断 [J]. 电网技术, 2020, 44(5): 1908-1913.

[4] 贾嵘, 李云桥, 张惠智, 等. 基于改进 BP 神经网络的光伏阵列多传感器故障检测定位方法 [J]. 太阳能学报, 2018, 39(1): 110-116.

[5] 陈凌, 韩伟, 张经纬. 基于数据融合的光伏组件故障诊断 [J]. 电网技术, 2017, 41(6): 1864-1873.

[6] 程江洲, 朱偲, 付文龙, 等. 基于贝叶斯网络的水电站多源信息故障诊断 [J]. 水利水电技术, 2018, 49(12): 103-110.

[7] 李松峰, 宋震, 侯珏, 等. 基于贝叶斯网络的地铁设备故障诊断方法 [J]. 交通运输研究, 2022, 8(4): 65-73.