DOI：10.13234/j.issn.2095-2805.2026.7.147 **中图分类号**：TM615 **文献标志码**：A

考虑时序聚类的分布式光伏异常数据辨识方法

毕校然[1]，王 亮[1]，梁宏杰[2]，左 帅[1]，徐彬红[1]，王梓涵[1]
(1.沈阳工程学院电气工程学院，沈阳 110136；2.国网甘肃省电力公司超高压公司，兰州 730030)

摘要：针对现有检测方法忽视或无法有效捕捉光伏时序数据复杂长期非线性特征，导致异常数据难以准确识别和定位的问题，提出了考虑时序聚类的分布式光伏异常数据检测方法。首先，采用动态时间规整 DTW (dynamic time warping)方法计算光伏电站时序数据的相似度，构建相似度矩阵；接着，在此矩阵基础上应用谱聚类得到特征矩阵，分析特征矩阵光伏数据点间的内在联系，将相似光伏数据聚类成集合；然后，建立基于光伏数据聚类集合的子孤立森林联合检测模型，采用局部离群因子算法获得异常分数特征，作为联合检测模型输入，得到各相似聚类集合的光伏数据异常分数矩阵；最后，对异常分数矩阵进行选择重构，经过决策分析判别并定位光伏数据的异常。实验结果表明，该方法可以有效捕获光伏异常数据，提取异常样本精准率与召回率均可达到0.9，提高了光伏异常数据识别的准确性。

关键词：异常辨识；聚类算法；分布式电源；子孤立森林框架；光伏数据异常；长期趋势验证

Distributed Photovoltaic Anomaly Data Identification Method Considering Time Series Clustering

BI Xiaoran[1], WANG Liang[1], LIANG Hongjie[2], ZUO Shuai[1], XU Binhong[1], WANG Zihan[1]
(1. School of Electrical Engineering, Shenyang Institute of Engineering, Shenyang 110136, China; 2. Ultra-high Voltage Company of State Grid Gansu Electric Power Co., Ltd., Lanzhou 730030, China)

Abstract: In order to address the shortcomings of existing detection methods, which are unable to effectively capture the complex nonlinear features of photovolatic time series data, a distributed photovolatic anomaly data detection method considering time series clustering is proposed. This method is designed to facilitate the accurate identification and location of abnormal data. Firstly, the similarity of photovolatic power plant time series data is calculated using the dynamic time warping (DTW) method, and a similarity matrix is constructed. Subsequently, spectral clustering is applied to obtain the feature matrix, and the intrinsic connection between photovolatic data points in the feature matrix is analysed. This enables similar photovolatic data to be clustered into a collection. Then, a joint detection model of sub-isolated forest is established based on the collection of photovolatic data clusters. The abnormal score geatures are obtained using the local outlier factor algorithm. The features are used as inputs to the joint detection model to obtain the anomaly score matrix of photovolatic data for each similar clustering set. Finally, the anomaly score matrix is selected and reconstructed, and the anomalies of photovolatic data are identified and located after decision analysis. The experimental results demonstrate that the method is an effective means of capturing photovolatic anomaly data. The precision and recall rates of extracted anomaly samples reach 0.9, thereby enhancing the accuracy of photovolatic anomaly data identification.

Keywords: Anomaly identification; clustering algorithms; distributed generation; sub-isolated forest framework; photovolatic data anomalies; long-term trend verification

收稿日期：2025-06-30；**修回日期**：2025-08-16，2025-09-08；
录用日期：2025-09-17；**网络首发日期**：2025-10-15
基金项目：新型电力系统下的分布式光伏网络安全防护方法研究资助项目(SGLNJZ00HUS2311442)
This work is supported by Research on Distributed Photovoltaic Network Security Protection Methods under the New Power System under the grant SGLNJZ00HUS2311442

近年来，光伏风电资源产业规模及应用逐步扩大，运行数据量迅速增加[1]。然而，分布式光伏运行状况复杂，可能存在底层运行数据缺失、防护薄

弱导致的运行风险，数据采集量、处理量剧增带来的数据管理风险[2]，宽频振荡与电压失稳风险[3]，以及高比例电力电子设备引发的连锁故障等问题，易产生大量异常数据。光伏数据可用于评估系统发电性能、识别组件故障和光伏功率预测，但异常数据不仅可能引发频繁的误报、漏报等问题，增加维护成本，还可能引起电压波动，降低电能质量，同时会干扰功率预测模型的准确性，影响系统调度。此外，潜在的异常可能逐渐累积恶化为严重故障，影响系统可靠性。因此，亟须开展光伏异常数据辨识工作，以提高数据质量，保障系统可靠性。

光伏异常数据的异常检测概念可划分为 2 类：相似性和差异性。光伏数据具有随时间规律变化的时序特性，即同时段输出数据具有一定的相似性。该特性可用于判断系统是否正常运行或准确定位异常，但若忽视数据的长期变化趋势则容易错误识别微小异常。同时，光伏组件故障的发生具有随机性，由此产生的异常数据与正常运行数据存在显著差异。基于这种差异可实现异常检测，但异常检测往往只用于识别异常点，可能无法实现异常点的有效定位。

相似性工作主要通过分析光伏正常工作模式下的历史数据建立正常工作状态的数学模型，以捕捉到的相似模式和特征来识别异常情况。文献[4]建立光伏发电量数学模型，并采用扩展卡尔曼滤波 EKF (extended Kalman filter)算法对系统采集到的发电量数据进行偏差值校核，实现了异常数据的有效检测和修复；文献[5]对光伏数据聚类并设定离群阈值，通过自编码网络筛选离群点，再对异常数据进行重构，实现了异常数据的高准确率筛查；文献[6]提出晴朗日筛选和分位数回归循环神经网络 QRRNN (quantile regression recurrent neural network)模型拟合出力区间的异常检测方法，通过分析晴天辐照度特征排除气象干扰，并结合同一区域电站输出相关性和同一电站不同晴天输出稳定性，利用 QRRNN 模型得到正常功率输出范围，显著提高了检测可靠性；文献[7]提出基于深度自编码器的无监督异常检测模型，仅利用光伏功率数据，通过重构误差均值和标准差推导最优阈值，并结合动态时间规整 DTW (dynamic time warping)误差分析，实现异常片段定位，在意大利热那亚光伏电站验证中，测试集上的 F1 Score 达 0.953 5，优于其他无监督模型。以上文献主要以光伏正常数据建立的模型为标准来检测异常，但光伏运行状况复杂，单一基准的静态数据训练无法完全捕捉这些复杂的非线性特征，且无法实现异常数据的准确定位，还需结合时序数据关联性来识别光伏数据的动态特征和变化趋势。文献[8]基于光伏逆变器监测数据构建了 Transformer 驱动的通信侧时序数据分析框架，并融合生成对抗网络 GAN (generative adversarial network，)与 DR-Score 评分机制以优化异常检测精度，进而采用多模态强化学习实现物理侧与通信侧数据的协同分析，为光伏系统实际运维提供了有效的解决方案；文献[9]使用正常光伏数据训练模型，通过构造特征矩阵将光伏多变量时间序列编码到二维图像中，实现异常数据的高准确率检测和定位；文献[10]针对光伏组件异常导致的逆变器数据异常，提出基于改进残差网络(SCCB-ResNet50)的检测方法。该方法利用马尔可夫转移场将时序数据转换为二维图像，并融合通道-空间注意力机制及优化训练策略，显著提升了检测性能。实验表明，其在曲线下面积 AUC(area under curve)、召回率与准确率等方面分别较卷积长短期记忆 ConvLSTM (convolutional long short-term memory)网络方法提升 4.6%、4.5%和 5.5%，显著优于现有光伏逆变器数据异常检测技术；文献[11]将光伏发电数据分解为多个模态分量以获取其周期和非周期特征参量，分析光伏数据的周期规律及总体趋势，建立了预测模型，并以预测值和量测值的动态阈值判别异常，该模型相比于单一模型取得超过 20%的性能提升；文献[12]对光伏功率数据进行了相似日划分，依据同时段均值对比算法识别连续异常数据，再使用四分位法识别离散型异常数据，识别效果较优；文献[13]使用时间序列聚类和早期随机分类算法从电力负荷历史数据中挖掘隐藏的深度信息，有效提高了数

据驱动模型的预测性能。对光伏数据相似聚类分析可更好地捕捉时序数据的变化模式，有助于提取数据中隐藏的规律和模式，但上述文献大多忽视了时序数据的长期趋势变化，故本文以时序聚类的方式作为相似性概念的切入点，并以时序交叉验证的方式来验证光伏数据的长期趋势。

差异性检测则更关注异常数据的差异特征。文献[14]通过分析光伏异常数据的特性建立了高斯混合模型，实现了高比例异常数据的检测和剔除；文献[15]以光伏阵列异常数据分布特征为依据使用滑动标准差实现了异常数据的高效清洗；文献[16]提出多站光伏支路功率的联合学习，通过提取数据的相似性和差异性特征进行异常检测，实现了大批量数据的快速准确识别；文献[17]针对光伏逆变器异常检测问题，提出一种基于长短期记忆深度自编码高斯混合模型 LSTM-DAGMM (long short-term memory-deep autoencoding Gaussian mixture model)的半监督检测方法。利用带噪声的基于密度的聚类算法 DBSCAN (density-based spatial clustering of applications with noise)预处理正常工况数据，进而集成特征增强技术与深度自编码高斯混合模型 DAGMM (deep autoencoding Gaussian mixture model)，实现无监督异常识别。验证表明，该方法能有效处理多维时序数据，利于数据的可视化处理。文献[18]采用长短期记忆自编码器 LSTM-AE (long short-term memory-autoencoder)模型，利用监控与数据采集 SCADA(supervisory control and data acqusition)系统采集的辐照度、环境温度等数据训练模型，通过重构正常运行模式，基于残差分析检测异常，同时设置 3σ 和 5σ 阈值区分异常等级，并结合三角移动平均过滤噪声生成预测性警报。在印尼努沙佩尼达 3.5 MWp 光伏电站验证中，测试集均方误差 MSE (mean squared error)、均方根误差 RMSE (root mean square error)、平均绝对值误差 MAE (mean absolute error)分别低至 10.95、3.30、2.76，优于 RNN-AE 等模型。但差异性检测大多只关注数据点相对于群体的差异，忽略了数据生成的上下文信息和时序动态变化，可能导致某些异常被忽略，且大多数检测算法依赖于参数选择，适用场景有限。文献[19]基于局部离群因子 LOF (local outlier factor)识别点异常，并基于 DTW 识别群体异常，有效构建了电力能耗数据的动态异常检测机制；文献[20]将投票的概念引入孤立森林 iForest (isolation forest)，将异常分数计算转换为考虑大多数树的集体决策，实现了模型性能的有效提升；文献[21]组合客观权重赋权 CRITIC (criteria importance through intercriteria correlation)和优劣解距离 TOPSIS (technique for order preference by similarity to ideal solution)法对目标系统进行分析，以提高决策者评估的准确性。本文考虑使用前述相似聚类概念形成子集，以捕捉时序数据之间的细微差异，但多个子集模型会加剧计算资源消耗，且聚类可能导致异常样本进一步减少，故需要选择优化计算策略和上述异常检测算法以应对这些挑战。

针对上述问题，本文建立了聚类-分类-异常分数决策的异常检测模型框架：首先，针对光伏数据结构与时序特性，基于 DTW 计算光伏电站时序数据的相似度构成相似度矩阵，并应用谱聚类 SC (spectral clustering)方法将时序周期数据聚类成相关特征集合，以作为异常分类器训练阶段的前置部分；接着，考虑光伏数据可能存在局部异常信息，引入 LOF 异常分数为异常分数特征，继而分析各个特征的数据关联性进行 CRITIC，同时建立子孤立森林联合检测模型并将上述特征作为输入；然后，输出各集合的光伏数据异常分数矩阵并进行选择重构，经过 TOPSIS 决策分析判别并定位光伏数据的异常，提升模型整体召回率；最后，通过仿真实验验证所提方法的有效性。

1　理论原理

1.1　光伏数据

选择辽宁某地 2023 年下半年的真实电力监测数据，包括交流电压电流频率、直流电压及其对应的光伏发电功率数据。一个周期样本的数据采集频率为 5 分钟一次，1 天共计 288 个数据点，选取连续 100 天的完整数据，形成包含 28 800 个数据点

的周期性样本数据集，所得多通道周期性电力监测时序数据可视化图如图 1 所示。

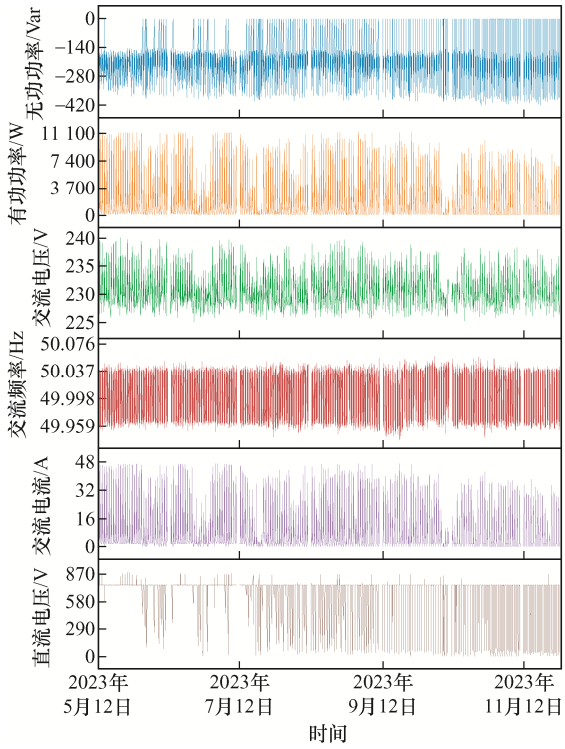


图 1 多通道周期性电力监测时序数据可视化图

Fig. 1 Multi-channel periodic power monitoring time series data visualization diagram

图 1 呈现了分布式光伏系统中 6 类关键电气量的真实监测时间序列特征。不同电气量采用不同颜色，反映其在整个采样周期内的动态变化规律。相比于正常运行状态，光伏阵列最大功率点的功率、电压及电流在异常状态下表现出显著偏离。有功功率与无功功率序列表现出明显的高频振荡及短时剧烈跳变，易受气象扰动或设备瞬态负载波动影响。交流电压信号呈现较低频率的周期性波动，反映日照强度变化或净输出功率的平滑趋势。交流频率表现为典型的二值跳变，与控制指令或运行状态转换相关，如光伏阵列的启停机制。交流电流及直流电压信号则存在间断、低幅值运行及突变现象，可能由数据采集异常、传感器故障或通信丢失所致。综合分析表明，监测数据既包含稳定且可预测的趋势成分，也掺杂多样的高频噪声、状态跳变及异常中断，为后续数据预处理、特征提取及异常检测提供了关键基础。

此外，对于光伏系统在无光照条件下的运行数据，本文进行了单独处理。鉴于夜间有功功率接近 0，但直流电压、交流电压等运行参数仍有记录，本研究将每日辐照度低于 50 W/m²的时段定义为夜间时段，并采取差异化处理策略：一方面，保留其有效电气量数据，作为低信号背景下的基准参考，帮助模型识别静态停机状态下的正常特征，避免将夜间低功率误判为异常；另一方面，剔除无有效电气量且功率始终为 0 的冗余样本，以减少对模型训练与检测性能的干扰。该策略既保留了低信号状态的参考意义，又确保了数据集在不同运行工况下的代表性和检测结果的可靠性，为后续异常类型的精准判别奠定了基础。

光伏发电系统可发生不同类型的故障[22]，包括电气故障、物理故障及环境异常等故障。借鉴文献[12]将光伏异常数据划分为突发性的离散型故障和持续存在的连续性故障 2 大类，每个大类中又可能包含各种小类故障。若某一小类别故障在短时间内被修复，则视为离散型故障。少量离散故障数据是由通信短时异常导致光伏功率记录数据发生突变产生的。若该故障长时间未被修复则会持续存在，此时则将其视为连续性故障。功率数据的连续故障是由设备故障、通信故障、光伏限电放弃及机组停机等事件引起的。

通过表 1 中的多种小类别故障模块构成多类型的故障数据输出，以 5 个包含典型故障数据点的周期功率数据做示例，选取 15 min 为一个时间点，如图 2 所示。连续性故障表现为功率曲线明显偏离正常水平，通常由设备或线路的持续性损坏引起，导致光伏功率在一段时间内持续下降，造成显著发电损失，需及时检测与修复；离散型故障则多为突发或间歇性事件，受随机噪声或环境扰动影响，故障点分散且短暂，修复后系统快速恢复，虽对整体趋势影响有限，但增加了数据处理和预测的复杂度。遮挡故障与组串间故障表现为功率点的轻微下移，遮挡效应具有随机性且持续时间短，难以形成连续故障，而组串间故障则影响多个组件的输出。组合型故障指多种故障同时或交互发生，更贴

近实际运行的复杂工况，其叠加效应放大功率损失，导致数据表现出杂乱无序的分布特征。

表 1　不同数据类型故障分布设置

Tab. 1 Distribution of different data types

类别设置	故障分布
大类	离散型故障、连续性故障
小类	环境异常(遮挡)、电气故障(接地故障、短路、断路等)、物理故障(组件故障、组串间故障)

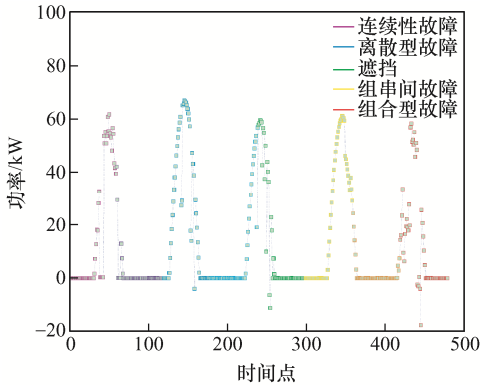


图 2　各类别故障功率分布示例

Fig. 2 Example diagram of fault power distribution of each category

参考相关文献与实际数据验证可知光伏输出功率与一些相关量的关系，因此本文选择光伏有功功率、无功功率、交流电压电流频率和直流电压作为参数数据。

采集的数据为 $\boldsymbol{X}=[x_0\ \ x_1\ \ x_2\ \ \cdots\ \ x_n]$，$n$ 为数据点的数量，其中包括有功功率、无功功率、交流电压电流频率和直流电压的数据集合。在原始光伏监测数据导入模型前，进行系统性的数据清洗。新型电力系统的主要特征之一是高比例的电力电子设备，运行状况复杂[23]，可能造成传感器数据缺失，故需先对数据进行缺失值查找。对缺失值部分，采集的数据为等时间间距数据点，利用牛顿插值多项式进行插补，表示为

$$N(x)=f[x_0]+\sum_{k=1}^{n}f[x_0,x_1,x_2,\cdots,x_k](x-x_0)\cdot (x-x_1)\cdots(x-x_k) \tag{1}$$

式中：x 为插值点；$x_0,x_1,\cdots,x_k$ 为给定数据点；$f[\cdot]$ 为差商函数值。

构造牛顿插值多项式后即可对指定数据进行插值。为消除不同物理量之间的量纲差异，光伏输出数据完成数据填补后需进行归一化处理，采用 Z-Score 归一化，其公式为

$$\boldsymbol{X}_{\text{norm}}=\frac{\boldsymbol{X}-\boldsymbol{X}_{\text{mean}}}{\boldsymbol{X}_{\text{std}}} \tag{2}$$

式中：$\boldsymbol{X}_{\text{norm}}$ 为归一化数据；$\boldsymbol{X}$ 为原始数据；$\boldsymbol{X}_{\text{mean}}$ 为原始数据的均值；$\boldsymbol{X}_{\text{std}}$ 为原始数据集的标准差。

1.2　时间序列聚类

通过对光伏时序数据集的模式识别与特征提取，实现不同气象条件及输出特性下的典型日型分类。该处理策略可显著提升数据表征的物理意义解析度，为后续异常检测提供更具针对性的分析基础，从而增强方法在实际应用场景中的解释性与工程适用性。已采集 1 个周期的数据点，样本数量为 288，光伏时序数据通常具有非线性、季节性和周期性等特点。传统的聚类方法(如 K-means)可能难以捕捉这些复杂的关系，而谱聚类能够通过图谱方法处理数据之间的相似度和关系，适合处理这类复杂的数据。查询光伏各时序周期输出数据的实值区间，并将其作为新时序序列，而后采用 DTW 算法求解周期时序的相似度并进行谱聚类。

截取 $[x_0,x_1,x_2,\cdots,x_n]$ 中各数据周期的实值序列，任选 2 个 x_i 和 x_j 为典型示例，并计算皮尔逊相关系数 ω_{ij}，表达式为

$$\omega_{ij}=\frac{\operatorname{cov}(x_i,x_j)}{\sigma_{x_i}\sigma_{x_j}} \tag{3}$$

式中：i、j 为任意 2 个时序周期索引；x_i 和 x_j 分别为有功功率、无功功率、交流电压电流频率和直流电压中任意 2 个时序周期实值序列；cov(·) 为集的协方差；σ_{x_i} 和 σ_{x_j} 为有功功率、无功功率、交流电压电流频率和直流电压中任意 2 个时序周期实值序列的标准差。

采用欧式距离构建 $[\boldsymbol{x}_{is}\ \ \boldsymbol{x}_{js}]$ 各时序周期之间的距离矩阵并求解规整路径，继而求解 2 个序列的累积距离，最终找到一条最优规整路径。

将 $\boldsymbol{x}_{is}$ 分解为 $\boldsymbol{x}_{is}=[x_{is,1}\ \ x_{is,2}\ \ \cdots\ \ x_{is,m}]^{\mathrm{T}}$，$m$ 为 $\boldsymbol{x}_{is}$

的时间序列长度，其子序列之间扭曲拉伸并求解路径，使得累积距离最小，具体为

$$\mathrm{Dist}(p,q)=d(p,q)+\min\begin{cases}\mathrm{Dist}(p-1,q)\\ \mathrm{Dist}(p,q-1)\\ \mathrm{Dist}(p-1,q-1)\end{cases} \tag{4}$$

式中：p、q 为 $\boldsymbol{x}_{is}=[x_{is,1}\ \ x_{is,2}\ \ \cdots\ \ x_{is,m}]^{\mathrm{T}}$ 中任意 2 个子序列中不同索引对应的数据点；$d(p,q)$ 为当前索引对应数据点的欧式距离；min 为各项取值的最小值；$\mathrm{Dist}(p,q)$ 为累积距离。

同理，可计算出 $\boldsymbol{x}_{js}$ 各子序列的累积距离。更新累积距离为 $[\mathrm{Dist}_i\ \ \mathrm{Dist}_j]$，引入前述 ω_{ij} 进行自适应加权平均计算，求解最终的最优距离 Dist_s 为

$$\mathrm{Dist}_\mathrm{s}=\omega_{ij}\times\mathrm{Dist}_i+(1-\omega_{ij})\times\mathrm{Dist}_j \tag{5}$$

式中：Dist_i 为 $\boldsymbol{x}_{is}$ 子序列计算出的最优距离；Dist_j 为 $\boldsymbol{x}_{js}$ 子序列计算出的累积距离。

遍历计算所有周期时序之间的最优距离 Dist_s，基于

$$D_\mathrm{s}=\frac{1}{\mathrm{Dist}_\mathrm{s}+\alpha} \tag{6}$$

引入逆距离概念，并将所有计算值转换为相似度矩阵 $\boldsymbol{D}_\mathrm{s}$。式中，$\alpha$ 为极小常数，取 0.1 以避免分母为 0。

基于高斯核函数构造邻接矩阵 $\boldsymbol{K}$，进行谱聚类，高斯核函数构造为

$$\boldsymbol{K}(D_{\omega,i},D_{\omega,j})=\exp\left(\frac{\left\|D_{\omega,i}-D_{\omega,j}\right\|^2}{2\sigma^2}\right) \tag{7}$$

式中：$D_{\omega,i}$、$D_{\omega,j}$ 为 $\boldsymbol{D}_\mathrm{s}$ 中的元素；σ 为高斯核距，为调优参数。

获取度矩阵 $\boldsymbol{D}_\mathrm{k}$，构建拉普拉斯矩阵为

$$\boldsymbol{L}=\boldsymbol{D}_\mathrm{k}^{-1/2}(\boldsymbol{D}_\mathrm{k}-\boldsymbol{K})\boldsymbol{D}_\mathrm{k}^{-1/2} \tag{8}$$

将矩阵 $\boldsymbol{L}$ 特征分解后对特征值进行降序，提取特征值最小的前 n_l 个特征向量构成矩阵 $\boldsymbol{D}_\mathrm{l}$。根据 K-means++算法进行聚类，并计算特征向量矩阵 $\boldsymbol{D}_\mathrm{l}$ 的误差平方和 SSE (sum of squared errors)为

$$S_\mathrm{SSE}=\sum d[\boldsymbol{D}_\mathrm{l}(u,v),c_{k_\mathrm{m}}] \tag{9}$$

式中：u、v 为 $\boldsymbol{D}_\mathrm{l}$ 矩阵各数据点索引；k_m 为聚类数目；$\boldsymbol{D}_\mathrm{l}(u,v)$ 为矩阵样本数据点；c_{k_m} 为聚类中心；$d(\cdot)$ 为 $\boldsymbol{D}_\mathrm{l}(u,v)$ 与 c_{k_m} 间的欧氏距离。

计算 SSE 以确定 k_m 个最优聚类数目，聚类簇 $D_g\,(g=1,2,\cdots,k_\mathrm{m})$ 对应不同气象因素下的光伏输出数据。为保证后续决策矩阵计算的一致性，本文在多次交叉验证的基础上，选择最优聚类数目出现频次最高的值作为最终设定。结合 SSE 肘部判据与交叉验证结果，最优聚类数最终确定为 k_m=4。在所采用的 2023 年 5~11 月光伏运行数据(覆盖夏季、秋季和初冬的典型工况)中，实验结果显示，最优 k_m 值在 3~5 范围内存在波动。综合考虑类内紧凑性与类间分离度，固定 k_m=4 能够在多数工况下保证聚类结果的稳定性。需要说明的是，本研究数据主要集中在下半年，因此尚未覆盖春季工况。后续研究将进一步引入全年及不同地区的典型光伏场站数据，以验证聚类数选择的稳健性，并探索自适应聚类数确定方法。聚类肘部判据如图 3 所示，此时在拐点处 SSE 下降速率明显减缓。

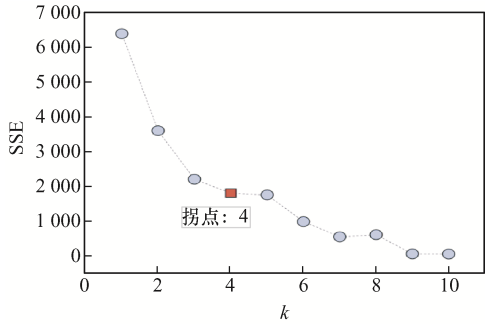


图 3 聚类肘部判据图

Fig. 3 Elbow criterion diagram for clustering

1.3 异常检测框架构建

聚类有助于发现数据模式和结构，且如果数据量庞大，先聚类形成子集有助于在一定程度上克服大规模数据集带来的挑战[24]，虽会导致计算量加大，异常样本进一步减少，但可以选取合适的算法避免这些弊端。iForest 算法是一种适用于异常检测的无监督学习算法，无需计算距离、密度等指标，其模型以疏离程度刻画异常[25]，故选 iForest 算法为主体算法构建异常检测框架。光伏输出数据中功率数据为主体关注数据，故以异常功率数据作为主

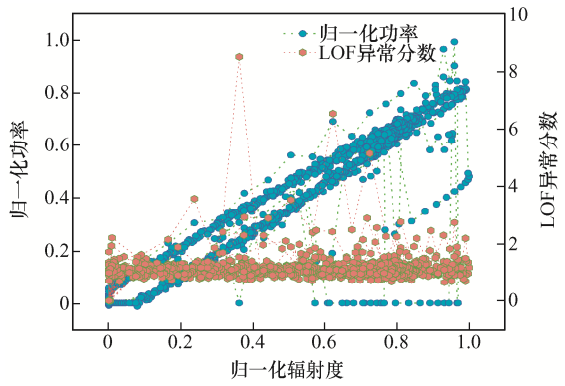
要研究对象。第 1.2 节已将多个周期的光伏输出数据聚类成集合，数据呈现密度特性，引入 LOF 异常分数作为其第 n_2+1 个数据特征，进一步识别局部区域内的异常点，以增强异常检测性能。在 LOF 特征构造过程中，关键参数为 k_n 近邻数。本文通过在训练数据集上进行交叉验证来确定 k 近邻数的取值范围，并结合不同工况下的检测性能进行评估。实验结果表明，当 k_n 取值在 15~25 之间时，异常检测的稳定性与精度较优，通过后续实验固定 k_n=20。该设定在保证检测敏感度的同时，避免了过小 k_n 值导致噪声敏感，过大 k_n 值导致局部结构过度平滑。

设 $D_g(g=1,2,3,\cdots,k_n)$ 中计算的数据点为 P，点 O 为点 P 的 k_n 近邻对象，计算局部异常因子[26]为

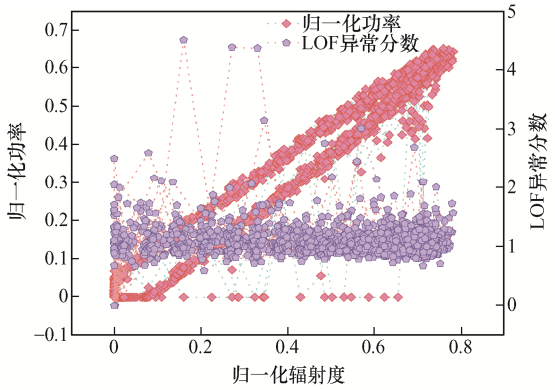
$$\mathrm{LOF}(P)=\sum_{O\in N_{k_n,\mathrm{dist}}(P)}\frac{\mathrm{Lrd}(O)}{\mathrm{Lrd}(P)}\Big/\left|N_{k_n,\mathrm{dist}}(P)\right| \tag{10}$$

式中：$\mathrm{Lrd}(P)$ 为数据点 P 的局部可达密度；$\mathrm{Lrd}(O)$ 为数据点 O 的局部可达密度；k_n,dist 为数据点 P 的 k_n 最近邻的最大距离；$N_{k_n,\mathrm{dist}}(P)$ 为 P 的 k_n 邻域中的对象集合。

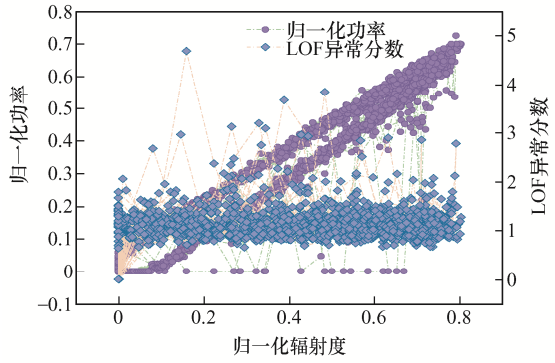
LOF 异常分数特征示例如图 4 所示，图中不同聚类簇内的各数据点均对应其特定的 LOF 异常分数，反映了样本点在局部邻域中的异常程度。正常样本的 LOF 异常分数主要集中分布于某一稳定的数值区间，表现出较为一致的局部密度特征；而异常样本则表现出显著偏离该区间的高异常分数，显示出其在局部邻域中密度明显低于周围点的特征。



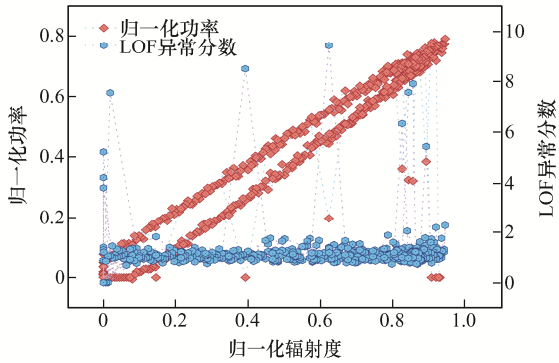
(a) 聚类簇1 LOF异常分数特征



(b) 聚类簇2 LOF异常分数特征



(c) 聚类簇3 LOF异常分数特征



(d) 聚类簇4 LOF异常分数特征

图 4　不同聚类簇 LOF 异常分数特征示例

Fig. 4 The feature example diagramafter adding LOF anomaly scores to different clustering clusters

计算出 LOF 特征后更新原有特征为

$$\boldsymbol{X}_{D_g}=[x_{\mathrm{r},D_g}\quad x_{p,D_g}\quad x_{\mathrm{LOF},D_g}] \tag{11}$$

式中：x_{r,D_g} 为辐射度数据聚类集合；x_{p,D_g} 为功率数据聚类集合；x_{LOF,D_g} 为异常分数特征聚类集合。

考虑光伏输出数据的波动性及彼此之间的相关性，对特征进行 CRITIC，基于

$$C=\sigma_e\times\sum_{f=1}^{n_2}(1-\varsigma_{ef}) \tag{12}$$

计算对比强度和冲突性的指标，并进行标准化，以获取数据信息[27]，同时将加权后的特征作为 iForest 模型的输入。式中：C 为计算出的指标权重；σ_e 为特征标准差；ς_{ef} 为各个特征之间的相关系数。

初始化 iForest 模型默认参数，通过构建不同子样本的二叉树 iTree，组合训练出 iForest，再以测试数据遍历分支的平均深度来计算异常分数，具体为

$$s(x,n_3)=2^{-E(h(x))/c(n_3)} \tag{13}$$

式中：$h(x)$ 为样本在树中的路径长度；$E(h(x))$ 为单个数据点 x 在所有树中达到深度的平均值；$c(n_3)$ 为修正值，其计算式为

$$c(n_3)=2H(n_3-1)-[2(n_3-1)/n_3] \tag{14}$$

式中：$H(\cdot)$ 为谐波数，由欧拉常数估计；n 为构造树的点数。

借鉴文献[20]所述投票机制构建基于 TOPSIS 的分数决策方法，以减少计算资源消耗[28]。TOPSIS 决策为多属性决策分析，本质上是一种统计方法，可通过调整权重来优化决策过程。鉴于主体算法为 iForest，去掉测试集数据与已建立模型的数据簇之间的相似计算，使用测试集数据遍历异常检测模型，获取异常分数矩阵 $\boldsymbol{s}_1=[s_{11}\ \ s_{12}\ \ s_{13}\ \ s_{14}]$，重构异常分数矩阵后基于 TOPSIS 方法进行多属性决策分析[29]，获取决策输出。

使用异常检测模型获取 $\boldsymbol{s}_1$ 并获取行中位数选择矩阵与行最大值选择矩阵，更新 $\boldsymbol{s}_1=[s_{11}\ \ s_{12}\ \ s_{13}\ \ s_{14}]$ 为 $\boldsymbol{s}_{\mathrm{m}}=[\boldsymbol{s}_{\mathrm{mid}}\ \ \boldsymbol{s}_{\max}]$，基于

$$\boldsymbol{s}_{\mathrm{ml}}=(a\times\boldsymbol{s}_{\mathrm{mid}}+b\times\boldsymbol{s}_{\max})/(a+b+\alpha) \tag{15}$$

完成最终的异常分数矩阵计算。式中：$\boldsymbol{s}_{\mathrm{ml}}$ 为基于统计信息选择的异常分数矩阵；a、b 为常数，本文中取 $a=3$、$b=2$；α 为经验性调优参数，本文中取 $\alpha=0.05$；$\boldsymbol{s}_{\mathrm{mid}}$ 为异常分数矩阵的行中位数选择矩阵；$\boldsymbol{s}_{\max}$ 为异常分数矩阵的行最大值选择矩阵。

2 模型训练与异常检测流程

2.1 光伏数据异常检测流程

在光伏运行数据中，既存在由设备故障、通信异常、传感器失灵等引发的真实异常，也存在因天气变化、光照遮挡或调度指令造成的正常波动。为避免将正常波动误判为异常，本文引入了多重判别逻辑。首先，根据辐照度与功率的历史对应关系构建正常运行包络线，并在不同气象条件下分别建立参考曲线；其次，将功率变化与辐照度变化的相关性低于 0.6 且偏离参考包络超过±3σ 的样本标记为疑似异常；最后，对疑似异常进行 LOF 分析与多周期对比，若在同一工况下连续周期均出现类似偏离，且无对应气象变化或调控事件，则判定为真实异常，否则归类为正常波动。该逻辑在数据进入异常检测模型前可有效剔除由环境或调控因素引起的波动，从而提高检测的准确性与可解释性。在此基础上，图 5 为本文的总体技术路线，架构聚类-分类-决策的异常检测模型框架。对光伏并网模型设置不同类型故障并获取异常数据，划分数据周期时序组别。基于 DTW 算法测量不同特征周期序列组别的相似度，基于谱聚类算法将时序周期数据聚类成集合。将 LOF 异常分数特征引入聚类数据集合，基于 CRITIC 加权并作为子孤立森林联合检测模型的数据输入，以获取异常分数矩阵。提取异常分数矩阵后进行重构选择，基于多属性决策分析方法对模型异常分数矩阵进行分数决策，获取异常分数输出。

本文方法涉及若干关键超参数，包括谱聚类 σ、k_{m} 与子孤立森林的树数等。在当前本文方法中，k_{m} 依据 SSE 肘部判据并结合交叉验证选取；σ 通过在开发集上综合聚类可分性与第 2.2 节的评价指标(精准率 P、召回率 R、FI Score F_1、G Mean G)确定稳定区间。为降低人工依赖并提升跨场站适配性，后续将继续研究引入自动化调参：采用贝叶斯优化在交叉验证框架下统一搜索 $\{\sigma,k\}$ 及 iForest 超参数；对阈值类参数引入基

于数据分布的自适应设定，并通过早停机制与增量评估控制时间和算力开销。该策略不改变现有实验结论，且可在不同气象与工况下提升鲁棒性。

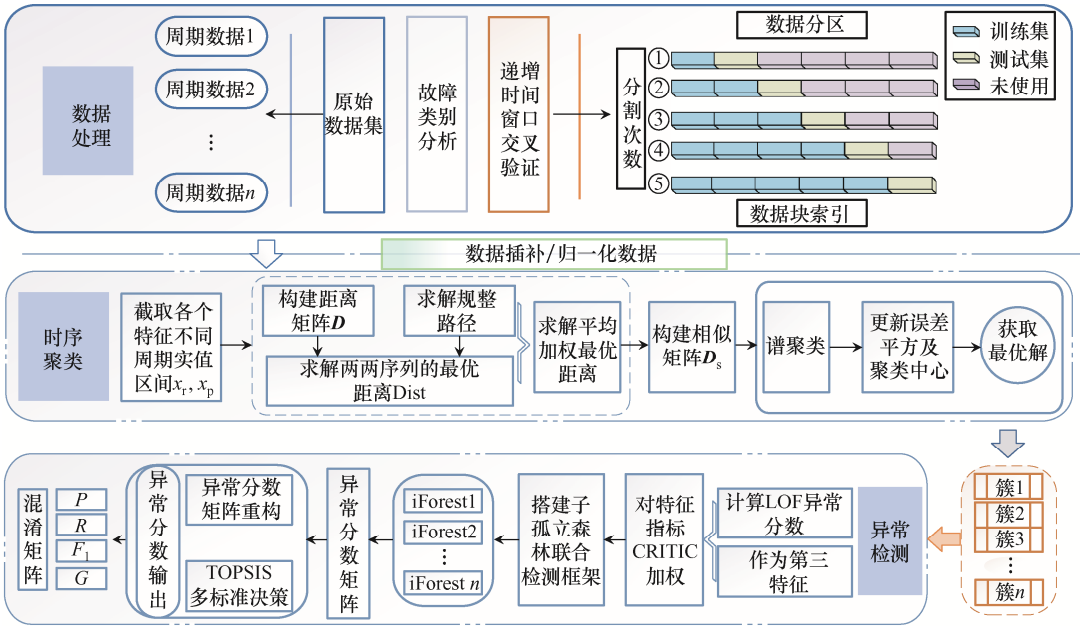


图 5　考虑时间序列聚类与分数决策的异常检测模型框架

Fig. 5 Framework diagram of anomaly detection model considering time series clustering with score decision making

2.2　评价指标

在光伏阵列故障诊断领域，模型需从海量数据中挖掘特征信息，实现不同故障类型的辨识与故障溯源[30]；而异常检测侧重于电网侧的运行状态监测，主要目标是基于光伏系统采集的数据快速识别偏离正常运行模式的异常状态。鉴于此，本文将异常检测问题简化为二分类任务，通过区分正常与异常两类状态来评估检测效果。为了全面衡量模型的性能，本文将混淆矩阵作为基础工具，对其结果进行检验，再选用 P、R、F_1 和 G 这 4 个指标来评估模型的测试性能。各指标的计算公式为

$$\begin{cases} P=\dfrac{\mathrm{TP}}{\mathrm{TP+FP}} \\ R=\dfrac{\mathrm{TP}}{\mathrm{TP+FN}} \\ F_1=\dfrac{2\times P\times R}{P+R} \\ G=\sqrt{P\times R} \end{cases} \tag{16}$$

式中：TP (真正例)为标签异常且被检测为异常的样本；FP (假正例)为标签正常但被检测为异常的样本；FN (假负例)为标签异常但被检测为正常的样本。

混淆矩阵是用于评估分类模型性能的一种工具，TN(真负例)为标签正常且被检测为正常，与 FP、FN、TP 构成混淆矩阵，TP 与 TN 数值越接近 1，模型越好；FP、FN 则相反。进一步，计算出模型指标：P 可以衡量模型识别出的异常占真正异常的比例；R 为真正异常被模型识别出异常的比例；F_1 为精准率和召回率的调和平均数；G 为召回率和精准率的几何平均值，两者可衡量类别不平衡数据集上模型异常检测的综合效能。上述 4 个模型评估指标的理想数值均为尽可能接近 1，其中，理想数值表明其在异常检测任务上优秀。

3　算例分析

3.1　算例结果分析

应用 MATLAB 建立基于异常的检测程序，对所有样本数据做递增时间窗口交叉验证：将前 50 个周期的数据作为初始窗口，记为训练集；将后续

5 个周期数据作为窗口步长，记为测试集。首先使用训练集数据建立异常检测模型，使用测试集数据遍历异常检测模型并投票选出最优结果输出；然后将该次窗口步长数据加入到初始窗口训练集中，再取新窗口后续 5 个周期数据作为新测试集；最后进行重复训练测试，获取多次模型输出，并取均值，进行综合性能评估。基于第 2.2 节所提及的评价指标验证异常检测模型的有效性及优越性，检测结果如图 6 所示。

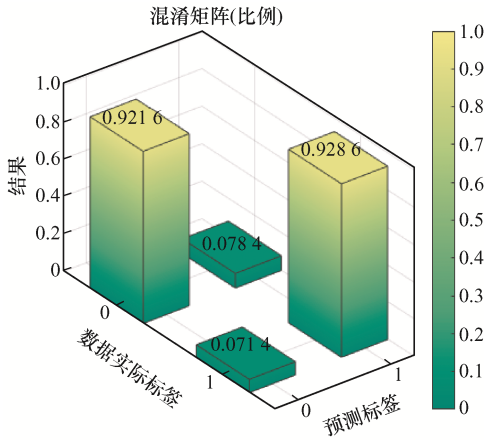


图 6 异常检测模型输出结果混淆矩阵示例

Fig. 6 Sample confusion matrix of anomaly detection model outputs

图 6 的混淆矩阵中，0 表示正常，1 表示异常。可知，其最优输出的混淆矩阵中 TP 和 TN 结果均可达到 0.920 0，所提方法在正常和异常样本上的分类能力较强。模型运行 3 次的检测指标结果见表 2。

表 2 异常检测模型交叉验证检测指标

Tab. 2 Anomaly detection model cross-validation detection metrics

指标	次数 1	次数 2	次数 3
P	0.922 1	0.926 1	0.896 2
R	0.928 6	0.900 0	0.923 1
F_1	0.925 3	0.912 9	0.909 5
G	0.925 3	0.913 0	0.909 6

表 2 中 P 与 R 可达到 0.920 0，借鉴文献[31]及文献[9]中有标签数据集的异常检测输出，本文中 R、F_1 等指标在采集的数据集中表现良好，表明其对异常的识别能力优越，不平衡样本数据下也可捕获大部分异常样本。为验证该模型各部分模块对计算结果的影响，选取设置不同模块的异常检测模型进行对比实验，见表 3。表 3 中“+”表示模型设置了该模块；“–”表示模型未设置该模块；“K-means+”、“灰色关联度+”表示模块算法被替换为当前所列算法。

表 3 不同模块设置下的异常检测模型

Tab. 3 Anomaly detection model with different module settings

模型	设置时序聚类	设置 LOF 特征和 CRITIC 权重	TOPSIS 决策输出
模型 1	K-means+	+	+
模型 2	灰色关联度+	+	+
模型 3	+	–	–
模型 4	+	+	–
模型 5	+	–	+

鉴于第 3.1 节中选择的阈值为最优阈值，为了直观表征模型不同模块在一定阈值区间上的变化性能，分别设置阈值为 0.90、0.85、0.80 进行检测。

3.2 考虑时序聚类的实用性验证

为验证时间序列聚类对计算结果的影响，对模型 1 与模型 2 进行对比实验，实验结果见表 4。

表 4 不同聚类算法设置下异常检测模型检测指标

Tab. 4 Detection metrics of anomaly detection models under different clustering algorithms settings

指标	模型 1			模型 2		
	次数 1	次数 2	次数 3	次数 1	次数 2	次数 3
P	0.989 8	0.980 9	0.986 8	0.979 3	0.980 7	0.983 1
R	0.053 6	0.272 7	0.442 9	0.051 3	0.216 2	0.466 7
F_1	0.101 6	0.426 8	0.611 3	0.097 5	0.354 3	0.632 9
G	0.230 3	0.517 2	0.661 1	0.224 1	0.460 5	0.677 3

由表 4 可知，求取相似性条件时，不同聚类方法对异常数据的检测性能影响颇大。虽然模型 1 与

模型 2 的 P 优秀，但 R 最高也只有 0.466 7，表明有大量的正常样本被预测为异常样本。这是因为异常样本相对于正常样本数量较少，聚类导致异常样本进一步减少，所以在模型训练的过程中学习了大量正常样本，捕捉异常样本的能力偏弱，导致召回率极低。而时间序列聚类比较可以有效捕捉不同周期时序数据的相似性和形状变化，有助于洞察后续异常检测中存在差异的数据模式，且在一定程度上可以抵抗噪声和异常值的干扰。与表 2 对比可验证时间序列聚类比较的有效性和实用性。

3.3 异常检测模型子模块效用验证

为验证权重参数的取值正确性及 LOF 特征对计算结果的影响，选取模型 3 进行实验，结果见表 5。

表 5　异常检测模型 3 检测指标

Tab. 5 Anomaly detection model 3 detection metrics

指标	模型 3		
	次数 1	次数 2	次数 3
P	0.721 8	0.718 3	0.724 9
R	0.984 6	0.977 8	0.984 8
F_1	0.833 0	0.828 2	0.835 1
G	0.843 0	0.838 1	0.844 9

将表 5 与表 4 比较可知，模型 3 的 P 有了小幅下降，但 R 提升了近 50 个百分点，F_1 及 G 指标也稳定在 0.800 0 左右，表明其可以有效捕获大部分异常样本且异常识别检测能力提升，再次验证了时间序列聚类比较的实用性。

模型 3 的检测指标虽然已有较大提升，但鉴于光伏数据的特殊结构及不均衡性质，还需结合其他异常检测算法以提高 P，增强鲁棒性。

进一步验证模型 4 与模型 5 的检测效果，结果见表 6。

由表 6 可知，模型 4 的 P 均稳定在 0.9，R 最高可接近 0.800 0。对比表 5 中模型 3 的输出，其 P 有了小幅度提高，但 R 有所降低。模型 4 的高 P 得益于在模块中引入了 CRITIC 权重和 LOF 特征设置，更好地优化了模型对高置信度异常样本的判别能力。而模型 5 中由于缺少模块，未引入 LOF 特征设置或 CRITIC 权重，因此它在特征表达上的区分度不足，但该模型加入了 TOPSIS 决策输出，更倾向于扩大检测范围，确保不会漏掉任何潜在的异常样本，因此召回率较高，可与模型 4 形成一定的互补关系。模型 4 和模型 5 的差异本质上是 P 和 R 之间的权衡，其性能受模块设置的影响显著。选择哪种模型，需结合具体的应用场景和对误报与漏报容忍度的权衡。

表 6　异常检测模型 4 与模型 5 检测指标

Tab. 6 Anomaly detection model 4 and model 5 detection metrics

指标	模型 4			模型 5		
	次数 1	次数 2	次数 3	次数 1	次数 2	次数 3
P	0.907 2	0.942 5	0.932 7	0.928 9	0.764 7	0.744 5
R	0.111 1	0.523 8	0.791 7	0.720 0	0.964 3	0.969 4
F_1	0.198 0	0.673 4	0.856 4	0.811 2	0.853 0	0.842 2
G	0.317 5	0.702 6	0.859 3	0.817 8	0.858 7	0.849 6

综上，设定不同的阈值梯次可表征模型在阈值区间上的检测性能。完整的时序聚类异常检测框架模型可有效增大模型输出指标值，在异常识别中能够成功地捕获大部分异常样本，识别异常能力优越。

3.4 异常检测模型优越性验证

为凸显本文方法与现有同类研究的差异性与创新性，本文选取了文献[19]中提出的“DTW+LOF”异常检测方法作为代表性对比对象。该方法利用 DTW 计算时序相似度，并通过 LOF 检测异常，在一定程度上考虑了光伏数据的时序特征。然而，该方法在数据处理与特征构建阶段未引入工况分层策略，且异常判别逻辑主要依赖局部密度差异，易受气象波动等外部因素影响。相比之下，本文方法针对分布式光伏运行数据的非平稳性与多工况特征，首先通过时序聚类对日运行曲线进行工况分层，再在每类工况下提取多源特征(功率−辐照度关系、统计阈值特征、LOF 分数等)，并结合多模型集成与多准则决策(TOPSIS)实现异常类型判别。这一设计在保留时序相似度信息的同时，显著增强了方法在多工况、多源扰动下的鲁棒性与可解释性。具体对比见表 7。性能评估结果表明，本文方法在 P、R、G 等指标上均优于 DTW+LOF，且运行时间

增长率呈线性，兼顾了检测精度与运行效率。

表 7 本文方法与 DTW+LOF 方法的技术路径差异对比

Tab. 7 Comparison of technical paths between the proposed method and DTW+LOF method

对比维度	DTW+LOF 方法	本文方法
数据特性处理	仅使用原始功率曲线，未区分运行工况；易受气象波动影响	基于 DTW 的时序聚类对日曲线进行工况分层，剔除夜间冗余数据，保留低信号基准
特征构建	单一 LOF 分数作为异常判别依据	多源特征(功率−辐照度关系、统计阈值特征、LOF 分数、气象相关性等)融合
异常判别逻辑	阈值基于局部密度差异，缺乏多重验证机制	历史包络+阈值检测+LOF+多周期对比，结合 TOPSIS 多准则决策
性能表现	对设备异常检测较好，但易误判气象波动；运行时间随数据规模增加而显著上升	多工况下 P、R 提升明显，在本研究数据规模下运行时间近似呈线性增长
可解释性与可扩展性	异常来源解释性较弱，适用性受限于单一特征	异常类型可解释性强，可扩展至多源数据及大规模监测场景

为进一步验证本文算法的优越性，搭建了基于三倍偏差算法的异常检测模型、基于 LOF 算法的异常检测模型、基于支持向量机 SVM(support vector machine)算法的异常检测模型。对不同模型做异常检测测试，计算不同模型下样本的异常检测指标。实验结果如图 7 所示，惩罚系数 c 表示软间隔支持向量机对错分样本的惩罚权重，可知三倍偏差法的 P 较高但 R 极低，因为三倍偏差法假设数据服从正态分布，光伏数据易受天气影响，导致该方法适用性受限；LOF 算法的 R 高但 P 不足，因为光伏功率异常点稀疏，过于关注少数类样本，从而误报正常数据为异常；SVM 算法的 P 极佳但 R 低，因为数据类别不平衡，算法可能过度偏向于学习正常数据的特征，导致对异常数据的学习不足。综合而言，本文算法具有更高的检测精度，对异常样本有较好的识别效果。除检测精度外，实际工程应用中运行效率同样是重要考量，因此需对所提方法的计算复杂度进行进一步分析。

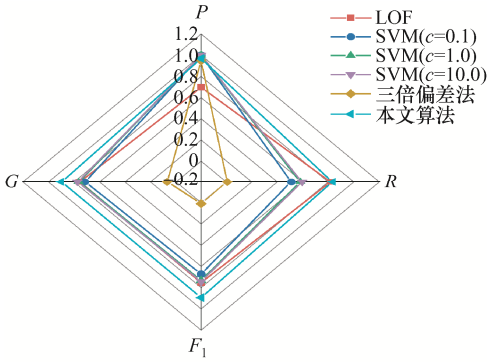


图 7 不同异常检测模型检测结果

Fig. 7 Detection results of different anomaly detection models

3.5 计算复杂度分析

为全面评估所提方法的工程适用性，除 P、R、F_1 及 G 指标等检测性能指标外，本文还对算法的计算复杂度进行了分析。设光伏时序数据集包含 N 个周期样本，每周期包含 M 个时间点，谱聚类步骤基于 DTW 的相似度计算，其时间复杂度约为 $O(N^2M^2)$，在本研究的数据规模下 ($N=100, M=288$) 可通过并行计算显著加速。LOF 特征计算的复杂度为 $O(N\log N)$，孤立森林模型的训练与检测复杂度约为 $O(t\cdot n_4\log n_4)$，其中 t 为构建树的数量，n_4 为样本数。TOPSIS 决策部分的复杂度较低，为 $O(pm)$，其中 p 为方案数，m 为属性数。

所有实验均在 Windows11 64 位系统的计算机上运行，硬件配置为 Intel(R)Core(TM)i7-12700F CPU(2.10 GHz)和 64 GB RAM，软件环境主要依赖 MATLAB 2022b。对比三倍偏差法、LOF、SVM 及本文方法的平均运行时间，结果见表 8。由表可知，本文方法在保证检测精度的同时，单次检测运行时间低于 2.10 s，虽高于三倍偏差法，但明显优于单纯的 LOF 及高维 SVM 模型。从整体流程来看，本文方法的运行时间随数据规模扩大呈多项式增长趋势，其中谱聚类阶段为主要计算开销，其复杂度约为 $O(N^2M^2)$，在数据量显著扩大时，耗时将呈平方级增长，而非严格的线性关系。然而，在本研究所采用的数据规模下，总体运行时间随数据增多仍近似表现为线性增长，与实验统计结果保持一致。由此可见，所提方法在大规模光伏监测

场景中具有良好的可扩展性。同时，该方法在检测性能与运行效率之间实现了较优的平衡，具备进一步应用于大规模监测系统的潜力。但在超大规模应用中，仍需引入稀疏矩阵分解、近似谱分解或并行化计算等优化策略，以有效缓解谱聚类阶段的复杂度瓶颈。

表 8　不同算法平均运行时间对比

Tab. 8 Comparison of average running time of different algorithms

算法模型	平均运行时间/s	数据规模扩大 2 倍运行时间/s	时间增长率	检测精度综合表现
三倍偏差法	0.32	0.64	2.00	中
LOF	3.46	6.92	2.00	高
SVM(c=10.0)	2.78	5.51	1.98	高
本文方法	2.08	4.15	1.98	高

分布式光伏场景中的边缘设备(如工控机、嵌入式 CPU 等)通常受限于算力、内存与能耗。为保证所提方法在资源受限环境下稳定运行，实践中可采用“边缘快速初筛+云端精细分析”的协同架构：边缘侧仅执行轻量级流程，复杂的时序聚类与多准则决策在中心服务器或云端离线完成；同时对边缘端计算引入增量更新与缓存，避免全量重算。部署时建议设置实时性与资源约束，并定期同步云端模型参数与簇划分，从而在实时性–精度–资源之间取得可工程化的平衡。

4　结论

为有效捕捉光伏异常数据的规律，提出考虑时序聚类的分布式光伏异常数据辨识方法，具体如下。

(1)基于时序交叉方法验证时序数据的长期趋势，划分时间序列组别并构建相似性矩阵。在保持时序数据的完整性和连续性的基础上，实现了时间序列聚类，使得样本数据相似性增强。

(2)架构了聚类–分类–决策的异常检测模型框架，基于该框架获取异常分数矩阵进行分数决策，兼顾降低计算复杂度的同时，基于样本的差异性特征完成了异常检测。通过模型仿真与算例验证表明了所提方法的有效性与可行性。同时，计算复杂度分析结果表明，该方法在保持高检测精度的同时，运行效率较高且具有良好的可扩展性，适合在光伏监测系统中推广应用。

此外，在系统集成方面，方法在数据接口层支持与现有电力监控系统(如 SCADA、EMS)的无缝对接，可通过采集标准化量测数据直接输入检测模块；在部署架构上，该方法可灵活运行于场站本地服务器，实现边缘计算与快速告警，也可部署于调度中心或云平台，实现多场站集中监测；在计算资源方面，对存储和显存要求较低，普通工作站或虚拟机即可运行，大规模场景可通过多线程或分布式计算进一步提升处理效率。

参考文献：

[1] 杨挺, 许哲铭, 赵英杰, 等. 数字化新型电力系统攻击与防御方法研究综述[J]. 电力系统自动化, 2024, 48(6): 112-126.

Yang Ting, Xu Zheming, Zhao Yingjie, et al. Review on research of attack and defense methods for digitalized new power system [J]. Automation of Electric Power Systems, 2024, 48(6): 112-126. (in Chinese)

[2] 王宇飞，吕晓宁，时海，等。基于 EGAT 的高比例分布式光伏配电网的脆弱性辨识[J]. 电源学报，2026, 24 (5): 219-228.

Wang Yufei, Lü Xiaoning, Shi Hai, et al. Vulnerability identification of high proportion distributed photovoltaic power distribution network based on EGAT [J]. Journal of Power Supply, 2026, 24 (5): 219-228. (in Chinese)

[3] 杨亚东, 耿丽清, 杨耿煌, 等.基于双通道特征融合并行优化的光伏阵列故障诊断[J/OL].电源学报,1-13. (2025-06-18) [2025-06-30]. https://link.cnki.net/urlid/12.1420.TM.20250617.1647.015.

Yang Yadong, Geng Liqing, Yang Genghuang, et al. Fault diagnosis of photovoltaic arrays based on dual-

channel feature fusion and parallel optimization [J/OL]. Journal of Power Sources, 1-13. (2025-06-18) [2025-06-30]. https://link.cnki.net/urlid/12.1420.TM.20250617.1647.015.

[4] 左松林, 陈伟, 付真斌, 等. 基于 EKF 算法的分布式光伏发电异常数据排查技术[J]. 电力工程技术, 2020, 39(5): 120-125.
Zuo Songlin, Chen Wei, Fu Zhenbin, et al. Abnormal data inspection technology of photovoltaic power generation based on EKF algorithm [J]. Electric Power Engineering Technology, 2020, 39(5): 120-125. (in Chinese)

[5] 彭勃, 李耀东, 龚贤夫. 基于自编码的改进 K-means 光伏能源数据清洗方法[J]. 计算机科学, 2024, 51(S1): 713-717.
Peng Bo, Li Yaodong, Gong Xianfu. Improved K-means photovoltaic energy data cleaning method based on autoencoder [J]. Computer Science, 2024, 51(S1): 713-717. (in Chinese)

[6] Yi Chengcheng, Peng Yu, Su Sheng, et al. Anomaly detection of photovoltaic power generation based on quantile regression recurrent neural network [J]. Electric Power Systems Research, 2025, 238: 111132.

[7] Miraftabzadeh S M, Longo M, Leva S, et al. Data anomaly detection in photovoltaic power time-series via unsupervised deep learning with insufficient information [J]. Sustainable Energy, Grids and Networks, 2025, 43: 101769.

[8] 郝佳欣. 基于深度生成模型的光伏逆变器异常检测方法研究[D]. 北京: 华北电力大学, 2024.
Hao Jiaxin. Research on anomaly detection method of photovoltaic inverter based on deep generation model [D]. Beijing: North China Electric Power University, 2024. (in Chinese)

[9] 周嘉琪, 毕利. 基于 GAN 的光伏逆变器数据异常检测技术[J]. 电力系统保护与控制, 2022, 50(1): 133-140.
Zhou Jiaqi, Bi Li. Abnormal detection technology of photovoltaic inverter data based on GAN [J]. Power System Protection and Control, 2022, 50(1): 133-140. (in Chinese)

[10] 姚森山, 庞成鑫. 基于改进残差网络的光伏逆变器数据异常检测方法[J]. 太阳能学报, 2025, 46(4): 322-330.
Yao Senshan, Pang Chengxin. Data anomaly detection method for photovoltaic inverter data based on improved residual network [J]. Acta Energiae Solaris Sinica, 2025, 46(4): 322-330. (in Chinese)

[11] 赵博超, 马嘉骏, 崔磊, 等. 基于改进 VMD-XGBoost-BiLSTM 组合模型的光伏发电异常检测[J]. 计算机工程, 2024, 50(3): 306-316.
Zhao Bochao, Ma Jiajun, Cui Lei, et al. Anomaly detection for photovoltaic based on improved VMD-XGBoost-BiLSTM combination model [J]. Computer Engineering, 2024, 50(3): 306-316. (in Chinese)

[12] 叶林, 崔宝丹, 李卓, 等. 光伏电站高比例异常运行数据组合识别方法[J]. 电力系统自动化, 2022, 46(20): 74-82.
Ye Lin, Cui Baodan, Li Zhuo, et al. Combined identification method for high proportion of abnormal operation data in photovoltaic power station [J]. Automation of Electric Power Systems, 2022, 46(20): 74-82. (in Chinese)

[13] Chen Zhe, Chen Yongbao, Xiao Tong, et al. A novel short-term load forecasting framework based on time-series clustering and early classification algorithm [J]. Energy and Buildings, 2021, 251: 111375.

[14] 刘岩, 李文文, 周丽霞, 等. 基于高斯混合模型的光伏发电出力中高比例异常数据检测方法研究[J]. 电测与仪表, 2021, 58(9): 14-21.
Liu Yan, Li Wenwen, Zhou Lixia, et al. Research of high proportion of outliers detection method in photovoltaic power output data based on Gaussian mixture model [J]. Electrical Measurement & Instrumentation, 2021, 58(9): 14-21. (in Chinese)

[15] 时珉, 尹瑞, 胡傲宇, 等. 基于滑动标准差计算的光伏阵列异常数据清洗办法[J]. 电力系统保护与控制, 2020, 48(6): 108-114.
Shi Min, Yin Rui, Hu Aoyu, et al. A novel photovoltaic array outlier cleaning algorithm based on moving standard deviation [J]. Power System Protection and Control, 2020, 48(6): 108-114. (in Chinese)

[16] 苏雍贺, 左颖, 靳健, 等. 基于多站支路功率联合学习的分布式光伏支路异常检测方法[J]. 计算机集成制造

系统, 2022, 28(7): 2149-2161.
Su Yonghe, Zuo Ying, Jin Jian, et al. Abnomaly detection of distributed photovoltaic array based on joint learning of multi-station array power [J]. Computer Integrated Manufacturing Systems, 2022, 28(7): 2149-2161. (in Chinese)

[17] 马乐. 基于多维时序数据的光伏逆变器异常检测系统的研究与实现[D]. 银川: 宁夏大学, 2022.
Ma Le. Research and implementation of photovoltaic inverter anomaly detection system based on multidimensional time series data [D]. Yinchuan: Ningxia University, 2022. (in Chinese)

[18] Syamsuddin A, Adhi A C, Kusumawardhani A, et al. Predictive maintenance based on anomaly detection in photovoltaic system using SCADA data and machine learning [J]. Results in Engineering, 2024, 24: 103589.

[19] Lei Lei, Wu Bing, Fang Xin, et al. A dynamic anomaly detection method of building energy consumption based on data mining technology [J]. Energy, 2023, 263: 125575.

[20] Chabchoub Y, Togbe M U, Boly A, et al. An in-depth study and improvement of isolation forest [J]. IEEE Access, 2022, 10: 10219-10237.

[21] 陈晨, 闫向阳, 齐桓若, 等. 基于 FAHP-改进 CRITIC 组合赋权的屋顶光伏接入配网评价方法[J]. 电力系统保护与控制, 2023, 51(15): 97-108.
Chen Chen, Yan Xiangyang, Qi Huanruo, et al. An evaluation method of a roof photovoltaic access distribution network based on the weight of FAHP-improvement CRITIC combination [J]. Power System Protection and Control, 2023, 51(15): 97-108. (in Chinese)

[22] 史洪岩, 彭云喆. 基于 Transformer-BiLSTM 并行网络的光伏阵列复合故障诊断[J/OL]. 电源学报, 1-13. (2025-06-18) [2025-06-30]. https://link.cnki.net/urlid/12.1420.tm.20250617.1557.011.
Shi Hongyan, Peng Yunzhe. Composite fault diagnosis of photovoltaic arrays based on a Transformer-BiLSTM parallel network [J/OL]. Journal of Power Sources, 1-13 (2025-06-18) [2025-06-30]. https://link.cnki.net/urlid/12.1420.tm.20250617.1557.011.

[23] 秦潘昊, 陈威宇, 胡秦然, 等. 新型电力系统设备状态监测与故障诊断传感芯片关键技术与展望[J]. 电力系统自动化, 2024, 48(6): 83-95.
Qin Panhao, Chen Weiyu, Hu Qinran, et al. Key technologies and prospect of equipment condition monitoring and diagnosis sensor chips for new power systems [J]. Automation of Electric Power Systems, 2024, 48(6): 83-95. (in Chinese)

[24] 李俊辉, 曾君, 樊淼嘉, 等. 基于改进 CRPS 和 QR-CNN-LSTM-SA 模型的短期光伏功率非参数概率预测[J/OL]. 电源学报, 1-11. (2025-05-18) [2025-06-30]. https://link.cnki.net/urlid/12.1420.tm.20250515.1454.006.
Li Junhui, Zeng Jun, Fan Miaojia, et al. Short-term nonparametric probability prediction of photovoltaic power based on improved CRPS and QR-CNN-LSTM-SA models [J/OL]. Journal of Power Sources, 1-11. (2025-05-18) [2025-06-30]. https://link.cnki.net/urlid/12.1420.tm.20250515.1454.006.

[25] Liu F T, Ting Kai ming, Zhou Zhihua. Isolation-based anomaly detection [J]. ACM Trans Knowl Discov Data, 2012, 6: 3: 1-3: 39.

[26] 黄子蒙, 余娟, 向明旭, 等. 基于改进动态时间弯曲的 PMU 频率异常检测及类型识别[J]. 电力系统自动化, 2022, 46(24): 104-112.
Huang Zimeng, Yu Juan, Xiang Mingxu, et al. Frequency anomaly detection and type identification of PMU based on improved dynamic time warping [J]. Automation of Electric Power Systems, 2022, 46(24): 104-112. (in Chinese)

[27] Abdel-Basset M, Mohamed R. A novel plithogenic TOPSIS-CRITIC model for sustainable supply chain risk management [J]. Journal of Cleaner Production, 2020, 247: 119586.

[28] Hashemi A, Dowlatshahi M B, Nezamabadi-pour H. MFS-MCDM: Multi-label feature selection using multi-criteria decision making [J]. Knowledge-Based Systems, 2020, 206: 106365.

[29] 李君卫, 余洋, 蒋衍君, 等. 面向电网调峰的负荷友好性评价[J]. 电力系统自动化, 2023, 47(20): 115-124.
Li Junwei, Yu Yang, Jiang Yanjun, et al. Load friendliness evaluation for peak regulation of power grid [J].

Automation of Electric Power Systems, 2023, 47(20): 115-124. (in Chinese)

[30] 孙鸿哲, 丁强, 曾启星, 等. 基于改进人工神经网络的太阳能光伏阵列故障诊断研究[J/OL]. 电源学报, 1-13. (2024-12-25) [2025-06-30]. https://link.cnki.net/urlid/12.1420.TM.20241225.1226.006.

Sun Hongzhe, Ding Qiang, Zeng Qixing, et al. Research on fault diagnosis of solar photovoltaic arrays based on an improved artificial neural network [J/OL]. Journal of Power Sources, 1-13. (2024-12-25) [2025-06-30]. https://link.cnki.net/urlid/12.1420.TM.20241225.1226.006.

[31] 杨茂, 孟玲建, 李大勇, 等. 基于类3σ准则的光伏功率异常数据识别[J]. 可再生能源, 2018, 36(10): 1443-1448.

Yang Mao, Meng Lingjian, Li Dayong, et al. Identification of abnormal data of photovoltaic power based on class 3σ [J]. Renewable Energy Resources, 2018, 36(10): 1443-1448. (in Chinese)



毕校然

作者简介：

毕校然(1999—)，男，硕士研究生。研究方向：电力设备在线监测与故障诊断。E-mail：2127358827@qq.com。

王亮(1983—)，男，通信作者，博士，副教授。研究方向：电力设备在线监测与故障诊断。E-mail：wl1983620@163.com。

梁宏杰(1997—)，男，硕士研究生。研究方向：电力设备在线监测与故障诊断。E-mail：18681146972@163.com。

左帅(2002—)，男，硕士研究生。研究方向：光伏阵列。E-mail：1871832558@qq.com。

徐彬红(2000—)，男，硕士研究生。研究方向：智能优化算法。E-mail：260271681@qq.com。

王梓涵(2000—)，男，硕士研究生。研究方向：智能优化算法。E-mail：773874595@qq.com。

(责任编辑：孙美菊)