

基于改进高斯混合模型的光伏短时波动游程聚类<sup>\*</sup>彭文静<sup>1</sup> 郑迪<sup>1</sup> 蔡慧<sup>1</sup> 邵海明<sup>2</sup> 王家福<sup>2</sup>

(1. 中国计量大学机电工程学院 杭州 310018; 2. 中国计量科学研究院 北京 100013)

**摘 要:** 针对大规模光伏发电短时波动性对电能准确计量的挑战, 本文提出一种基于改进高斯混合模型的光伏短时波动信号游程聚类分析方法。首先, 从游程理论出发分析了光伏输出的短时波动信号特征; 其次, 针对光伏短时波动信号分解得到游程过多、难以提取典型波动特征的问题, 采用基于改进高斯混合模型聚类方法对海量游程进行聚类; 进一步提出了主客观融合的聚类结果评价方法。最后, 对光伏电站现场录波数据的仿真结果表明, 相较于其他方法, 所提方法聚类结果评分在各方面有 1.1%~61.4% 的提升; 在不同噪声及异常值水平下所提方法也可以维持较好的聚类效果, 复合指标评分下降程度小于其他算法 0.92%~18.24%。所提方法通过深度学习技术和贝叶斯信息准则实现了高斯混合模型的自适应聚类, 提高了对含噪声和异常值数据的适应能力和稳定性, 能够实现光伏电站时波动信号游程的合理聚类。

**关键词:** 光伏短时波动信号; 游程分析; 改进高斯混合模型; 游程聚类; 贝叶斯信息准则

**中图分类号:** TB971; TN98 **文献标识码:** A **国家标准学科分类代码:** 470.4017

## Run clustering for short-term fluctuation of photovoltaic based on improved Gaussian mixture model

Peng Wenjing<sup>1</sup> Zheng Di<sup>1</sup> Cai Hui<sup>1</sup> Shao Haiming<sup>2</sup> Wang Jiafu<sup>2</sup>

(1. Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China;

2. National Institute of Metrology, Beijing 100013, China)

**Abstract:** To address the challenge of short-term fluctuation of large-scale photovoltaic power generations pose a challenge to accurate energy measurement, this paper proposes a new method for run clustering for short-term fluctuation of photovoltaic based on improved Gaussian mixture model. Firstly, the characteristics of short-term fluctuation signals of photovoltaic output are analyzed based on the run theory. Secondly, to address the issue of excessive run and difficulty in extracting typical features in the power generation of photovoltaic, the clustering method based on the improved Gaussian mixture model is adopted to cluster the massive run. Furthermore, a subjective-objective fusion evaluation method for clustering results is proposed. Finally, the simulation results of on-site recorded waveforms from photovoltaic power stations show that, compared with other methods, the proposed method has an improvement in clustering result scores ranging from 1.1% to 61.4% in different aspects. The proposed method can maintain good clustering effects under different noise and outlier levels, with a decrease in the composite index score that is less than that of other algorithms by 0.92% to 18.24%. The proposed method achieves adaptive clustering of the Gaussian mixture model through deep learning technology and the Bayesian information criterion, enhancing its adaptability and stability to noisy and outlier data, and enabling reasonable clustering of run-lengths of photovoltaic power station short-term fluctuation signals.

**Keywords:** short-term fluctuation signal of photovoltaic; run analysis; improved Gaussian mixture model; run clustering; Bayesian information criteria

## 0 引 言

随着“十四五”规划推进, 新能源发展继续保持快速增

长趋势, 推动新型电力系统建设不断深化<sup>[1]</sup>。其中, 光伏作为新能源领域中发展最为迅猛的电源类型, 其装机容量和发电量均呈现出爆发式增长, 成为新型电力系统中不可或

缺的重要组成部分。然而,不同于常规电源,光伏电源具有波动性大、高谐波分量等特征,大量接入电网时对电能计量的影响不断凸显,尤其在出力波动性方面的影响吸引了大量学者的关注<sup>[2]</sup>。

光伏受到光照等因素影响,其输出具有很强的波动性和随机性,对动态电能计量算法也提出了更高的要求。为准确分析光伏电站短时波动性对电能计量的影响,需要对其输出信号进行特征提取和建模。

光伏电站出力的季节性、日级、小时级等长时间波动通常具有规律性<sup>[3]</sup>,具有夏强冬弱、昼强夜弱的特点,呈现单峰曲线<sup>[4]</sup>。这些较长时间尺度的波动通常采用概率统计的方法来进行描述和表征<sup>[5-6]</sup>,现有算法和设备也可以较好地满足这种长时间尺度的电能计量需求<sup>[7]</sup>。部分文献采用概率密度函数描述光伏的短时波动特性,进行超短期的出力预测<sup>[8-9]</sup>。但电能计量误差的分析通常依赖于确定性的测试信号<sup>[10]</sup>,概率性的波动信号建模结果难以直接适用。为此,部分文献采用傅里叶、小波变换等时频域分析方法来提取光伏短时波动的典型特征,进而进行重构获得典型的测试信号,如文献[11]利用连续小波变换方法提取不同尺度上的光伏波动特性;文献[12]采用变尺度时间窗口提取时序变化特征以描述波动特性。这些时频域分析方法虽然能够描述信号非平稳特性,但是难以直接反映随机波动信号时域状态的持续性及异常波动信息,同时对于极短时间尺度的波动仍存在频率分辨率及尺度时间窗口切换速度不足等问题。

游程是一个序列中某种状态持续出现的历程,适合于对数字信号序列的随机性和连续性的分析。文献[13-14]提出采用游程理论对负荷的幅度域短随机特征进行分析,但直接分解会得到大量游程分量,湮没了短时波动的主要特征。此外,利用这些游程分量重构测试信号并分析对电能计量的影响时组合方式非常多样,给分析带来巨大的工作量和困难。为此,有必要对得到的海量游程分量进行聚类,获得能够反映短时波动典型特征的游程分量,进而用于后续对电能计量的影响分析中。特征聚类主要包括划分聚类、层次聚类、密度聚类、谱聚类和模型聚类等方法<sup>[15-16]</sup>。高斯混合模型(Gaussian mixture model, GMM)作为一种典型的模型聚类方法,通过期望最大化优化混合权重和高斯分布参数实现聚类,能够处理具有不同大小、形状和方向的集群,具有较强的灵活性和复杂数据处理能力,但也存在聚类数难以确定且对于初始数据敏感的问题。

为此,本文提出了基于改进 GMM 的光伏短时波动信号游程聚类分析方法,通过贝叶斯信息准则和深度学习实现了聚类数的高效计算并提高了对异常数据的鲁棒性,可对光伏短时波动信号的游程进行更合理的聚类。首先,分析了光伏出力短时波动信号的特点,采用幅度域游程分解获得其游程分量;其次,利用贝叶斯信息准则和深度学习对 GMM 进行改进,提出了光伏短时波动信号游程聚类方法;

然后,提出了基于主客观融合的聚类结果评价方法;最后,通过仿真算例验证了所提方法的有效性。

## 1 光伏短时波动信号的游程特征分析

### 1.1 光伏短时波动信号分析

按照不同时间尺度,光伏出力的波动可分为季节性、日级、小时级、分钟级和秒级等。当前动态电能计量算法对于季节性、日级、小时级、分钟级等较长时间尺度的波动能实现较为准确的计量<sup>[17]</sup>,因此本文主要针对秒级及以下的短时波动信号进行分析。

光伏发电系统利用光伏板将光照转化为电能,并通过逆变器实现交直流变换和能量传输<sup>[18-19]</sup>,其系统结构如图1所示。

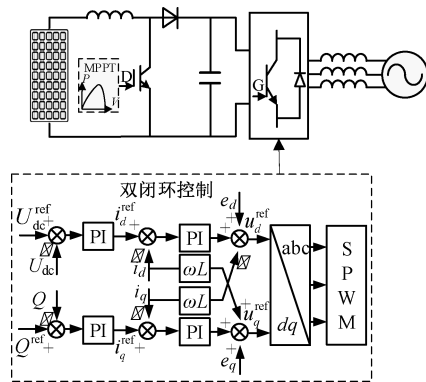


图1 光伏并网发电系统结构图

Fig. 1 Structure of grid-connected photovoltaic power generation system

通常采用双闭环控制实现有功和无功的解耦控制,并采用最大功率点跟踪(maximum power point tracking, MPPT)实现太阳能的充分利用。双闭环控制中,内环电流控制时间尺度为10 ms级,而外环功率控制时间尺度大约为百毫秒,MPPT通常采用电导增量法等方法实现,其响应时间尺度为几十毫秒到几秒不等,传统MPPT算法的响应时间通常在100 ms~3 s之间,而最新的改进算法可以将响应时间缩短到15~30 ms<sup>[20-21]</sup>。因此在秒级以上的光照变化输入下,可认为发电系统已完全响应,输出可以理想跟踪输入量的变化,输出波动主要来自光照强度的变化。

### 1.2 光伏短时波动信号游程分解及数字特征特征

光伏输出电流信号可表示为包含幅值、频率和相位的正弦函数。为简化分析光伏输出电流信号的波动性,将电流信号表征为非稳态随机信号,并考虑其时变性。首先,采用希尔伯特-黄变换(Hilbert-Huang transform, HHT)获取电流信号的瞬时幅值,从而反映电流幅值随时间的波动行为。其次,计算电流瞬时幅值的上、下包络线的均值作为准稳态项,以表示电流幅值的缓慢变化。最后,利用准稳态项对电流瞬时幅值波形进行分割,准稳态项以上为上升游程,以下为下降游程。这些游程代表动态分量,捕捉围绕缓

慢变化的准稳态分量的快速随机波动<sup>[13-14]</sup>。

由于直接通过幅度域电信号动态提取所得到的游程数量多且形态各异,难以通过其形态逐个分析其波动特性。为准确描述游程信息,计算所提取电流游程波形数字特征进行量化分析。分别计算游程幅值均方根值、游程调制深度、游程冲击强度、游程峰值系数以获取其数字特征并构建特征矩阵:

$$\mathbf{F} = [\text{RMS}, \Delta H, Q_f, C_f] \quad (1)$$

式中:  $\text{RMS} = \sqrt{\frac{1}{L} \sum_{j=0}^{L-1} A^2(n_j)}$  为游程幅值均方根值,其中,  $L$  表示游程波形的长度,  $A$  表示电流的瞬时幅值,  $n_j$  为第  $n$  个游程的第  $j$  个离散时间点;  $\Delta H = \frac{\max[A(n_j)] - \min[A(n_j)]}{\max[A(n_j)] + \min[A(n_j)]} \times 100\%$  为游程调制深度,  $\max[A(n_j)]$  和  $\min[A(n_j)]$  为第  $n_j$  个游程电流幅值的最大值和最小值;  $Q_f = \frac{\max[A(n_j)]}{\bar{A}(n_j)}$  为游程冲击强度,  $\bar{A}(n_j)$  为上升游程起始点或下降游程最小值点前 5 个电流幅值的平均值;  $C_f = \frac{\max[A(n_j)]}{\text{RMS}[A(n_j)]}$  为游程的峰值系数。

## 2 基于改进 GMM 的光伏短时波动信号游程聚类方法

### 2.1 深度学习与 GMM 融合的聚类算法

深度学习与 GMM 融合的聚类算法结合了深度学习的非线性特征提取和 GMM 的概率聚类,能够克服传统 GMM<sup>[22]</sup> 在处理复杂游程波形数据敏感和抗干扰能力不足。针对游程波形数字特征矩阵的复杂结构,通过改进的 GMM 聚类提升聚类的抗干扰能力,显著提升了聚类性能。

基于改进 GMM 聚类算法主要包括游程波形数字特征矩阵训练和 GMM 聚类两个方面,具体步骤包括:

1) 对于每一个全连接层,使用前向传播方法,公式为  $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$ , 其中  $\mathbf{x}$  是输入向量,  $\mathbf{W}$  是权重矩阵,  $\mathbf{b}$  是偏置向量,  $\mathbf{y}$  是输出向量。

2) 使用 RELU 激活函数增加网络非线性,其中  $\text{RELU}(\mathbf{x}) = \max(0, \mathbf{x})$ , 这里所有的负值都设为 0, 所有的非负值保持不变。

3) 解码器从编码后的低维表示中重构出原始数据。

4) 计算均方误差(mean-square error, MSE)作为损失函数:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2)$$

式中:  $y_i$  是原始游程数字特征矩阵的第  $i$  个元素,  $\hat{y}_i$  是解码器重构游程数字特征矩阵的第  $i$  个元素,  $N$  是样本数。

5) 通过反向传播算法计算损失函数关于每个权重的梯度,并使用 Adam 优化器更新权重。

6) 初始化高斯分布的参数(均值向量  $\boldsymbol{\mu}$  和协方差矩阵

$\boldsymbol{\Sigma}$ )和混合权重  $\pi$ 。

7) 计算每个数据点属于每个聚类的概率,即后验概率,公式为:

$$\gamma_{ik} = \frac{\pi_k \cdot p(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \cdot p(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \quad (3)$$

式中:  $\gamma_{ik}$  表示数据点  $\mathbf{x}_i$  来自聚类  $k$  的概率,  $p(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  是第  $k$  个高斯分布的概率密度函数。

8) 更新高斯分布的参数和混合权重,使得整体数据的对数似然函数最大化。

9) 重复步骤 7) 和 8) 直至收敛。

10) 根据每个游程波形所属的高斯分布概率确定聚类归属。

### 2.2 基于贝叶斯信息准则的聚类数计算

深度学习与 GMM 融合的聚类算法中聚类数量通常基于数据的先验知识确定,如肘部系数法则<sup>[23]</sup>等,但肘部法则在确定聚类数量时存在主观性、缺乏统计基础且对数据分布敏感,而贝叶斯信息准则(Bayesian information criterion, BIC)是将先验知识与观测数据结合,从而更新聚类数量,提供了一个更客观方法,能更好地平衡模型复杂度和拟合程度,且适用于处理复杂数据集。BIC 与 GMM 相结合方法能够提高处理大规模数据处理能力以及模型的准确性和稳定性。

BIC 是主观贝叶斯派中归纳理论的重要组成部分,采用主观概率估计方法处理部分未知分量,对事件是否发生的概率利用贝叶斯公式加以修正,最后综合利用修正概率和数据期望值二者评判出最优决策。它可用于评估给定数据对统计模型的拟合好坏,同时考虑模型的复杂度。BIC 鼓励数据拟合度较高的模型,但对包含更多参数的模型施加惩罚,从而防止过拟合。其中, BIC 定义为:

$$\text{BIC} = \ln(n) \cdot g - 2 \cdot \ln(\hat{L}) \quad (4)$$

式中:  $n$  是游程的数量,  $g$  是模型中参数的数量,  $\hat{L}$  是模型的最大似然估计值。

式(4)中的第 1 部分表示模型与数据的拟合优良状态,第 2 部分为复杂度的惩罚项,最终目的是通过逐步逼近获得最佳聚类数目。伴随着最大似然函数中残差和未知差的引入, BIC 值会增大,所以在选择合适的聚类数目时应使用较小的 BIC 值。

### 2.3 基于改进 GMM 的游程聚类方法

根据上述分析,综合深度学习、BIC 和 GMM 的优势,提出基于改进 GMM 的光伏短时波动信号游程聚类分析方法,其流程图如图 2 所示。具体步骤包括:

1) 采集光伏电站输出电流数据并进行预处理;

2) 对光伏电站输出电流信号进行 HHT 分解获取幅值信息,并提取上升游程波形和下降游程波形;

3) 计算幅度域游程波形数字特征,构建游程波形特征

矩阵;

4) 使用 BIC 确定最优聚类数;

5) 基于改进 GMM 聚类算法对海量游程波形进行聚类,并输出游程聚类结果及典型游程。

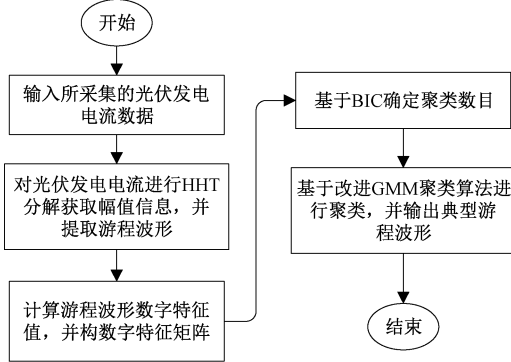


图2 基于改进 GMM 的光伏短时波动游程聚类方法流程图

Fig.2 Flow chart of short-time fluctuating run clustering method for PV based on improved GMM

### 3 基于主客观融合的聚类结果评价

#### 3.1 基于层次分析法的评价指标构建

为确定聚类结果的优劣,检验本文所提方法的有效性,通过层次分析法从聚类质量、物理意义、计算效率和可解释性4个层面构建评价指标。

轮廓系数和 Davies-Bouldin(DB)指标从不同的角度评估聚类质量,轮廓系数聚焦于个体聚类情况,DB 指标则关注整体簇的分布情况,二者结合可以更为全面的评估聚类质量。

轮廓系数结合了内聚度和分离度两个因素,用于衡量同一簇内与其他样本的相似度,以及不同簇之间的不相似度。其计算公式为:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

式中: $a(i)$ 是点  $i$  到其所在簇内其他点的平均距离, $b(i)$ 是点  $i$  到最近聚类的所有点的平均距离。

DB 指标衡量每个聚类与其最相似的聚类之间的平均相似度。其计算公式为:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{s_i + s_j}{d(c_i, c_j)} \right) \quad (6)$$

式中: $k$  为聚类数目, $s_i$  和  $s_j$  是聚类  $i$  和聚类  $j$  中所有点到聚类中心的平均距离, $d(c_i, c_j)$  是聚类  $i$  和聚类  $j$  的聚类中心距离。

聚类的物理意义可以体现为游程波形特征的捕捉能力,可通过聚类间的方差来衡量。这反映了不同簇之间差异程度,方差越大,表示聚类结果越能有效区分不同类型的游程波形。聚类间方差公式为:

$$V_B = \frac{1}{n} \sum_{i=1}^k n_i \|m_i - m\|^2 \quad (7)$$

式中: $n_i$  是第  $i$  个聚类的游程波形数, $m_i$  是第  $i$  个聚类中心数据序列, $m$  是所有游程波形的中心, $\|\cdot\|$  表示欧几里得距离。

聚类算法计算效率使用算法运行时间  $T_c$  表示,其中  $T_c = t_{end} - t_{start}$ 。需要说明的是,如果算法需要多次计算才能选到合适的聚类数,则  $T_c$  为包含所有运算次数的总时间。

聚类的可解释性主要由聚类中心代表性来表示,通过观察游程波形数据序列到聚类中心的距离,可以直观地了解游程波形数据序列与聚类中心的关系,从而更好地解释聚类的结果。其计算方法如下:

1) 计算游程波形到最近聚类中心的距离:

$$d_{ij} = \|x_i - c_j\| \quad (8)$$

式中: $x_i$  表示第  $i$  个游程的数据序列, $c_j$  表示游程波形第  $j$  个聚类中心的数据序列, $\|\cdot\|$  表示欧几里得距离。

2) 计算聚类中心可代表性:

$$REP = \frac{1}{k} \sum_{j=1}^k \left( \frac{1}{n_j} \sum_{i \in C_j} d_{ij} \right) \quad (9)$$

式中: $n_j$  是聚类类别中的游程波形数据序列数目, $C_j$  是聚类类别  $j$  中所有游程波形数据序列集合。

#### 3.2 基于主客观融合的聚类结果评价

熵权法是一种客观赋权方法<sup>[24]</sup>,基于信息熵理论,通过计算各指标数据的信息熵值来反映其变异程度,进而确定其在整体评价中的重要性。为更全面地评估聚类结果,可通过熵权法计评价指标的权重,以构建复合指标,从而得到更加全面和准确的评价结果。其中,熵值越小表示该指标提供的信息量越大,其在评价中的权重也就越大。具体步骤如下:

1) 对聚类结果的各评价指标进行数据归一化处理;

2) 计算各评价指标比重:

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^m r_{ij}} \quad (10)$$

式中: $r_{ij}$  是归一化后各指标值, $p_{ij}$  是第  $i$  个指标的比重, $m$  是指标数量;

3) 计算每个评价指标的信息熵:

$$e_j = -f \sum_{i=1}^m p_{ij} \ln(p_{ij}) \quad (11)$$

式中: $f = \frac{1}{\ln(m)}$ , $e_j$  是第  $j$  个指标的信息熵, $f$  为常数,用于保证  $0 \leq e_j \leq 1$ 。

4) 计算评价指标权重:

$$w_j = \frac{1 - e_j}{\sum_{j=1}^m e_j} \quad (12)$$

式中: $w_j$  是第  $j$  个指标的权重。

5) 计算聚类结果复合指标评分:



$$Y = \sum_{j=1}^m w_j r_{ij} \quad (13)$$

式中:  $Y$  是聚类结果复合指标评分。

通过层次分析法构建评价指标,进一步利用熵权法计算数据本身特性,从而得到更为全面和可靠的聚类结果复合指标  $Y$ ,其中较高的  $Y$  表示聚类效果更好。

4 算例分析

为对光伏电站短时波动幅度域游程分析,采集了广东某地光伏电站 2023 年 3 月某日的实际电流数据,如图 3 中实线所示,采样频率为 5 kHz。对采集的数据进行 HHT 变换得到电流幅值信息,计算幅值上下包络均值作为准稳态项,进而提取得到的上升游程和下降游程,如图 3 所示。

对实测信号进行分解得到了上升游程 5 063 个和下降游程 5 064 个,计算了游程波形的数字特征,各特征的概率分布情况如图 4 所示。

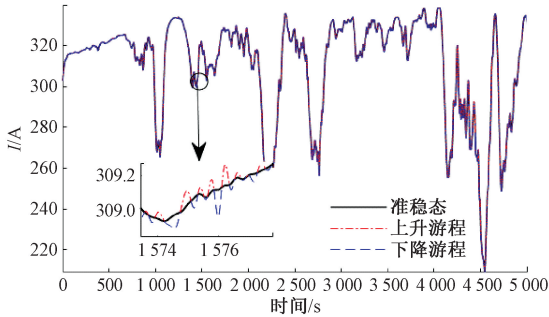


图 3 光伏电站原始输出信号及游程分解结果  
Fig. 3 Runs and original signal of photovoltaic station

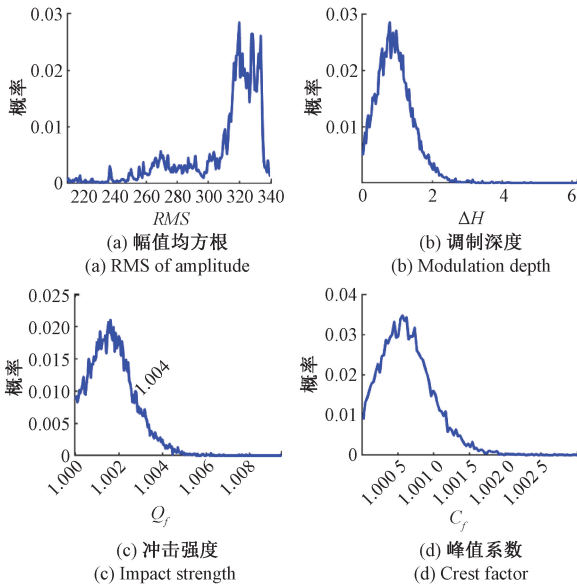


图 4 光伏电站游程数字特征概率密度分布

Fig. 4 Digital characteristics probability density distribution of runs of photovoltaic station

采用所提改进 GMM 聚类算法得到的游程聚类结果如图 5,其中黑色加粗实线表示该类游程波形的聚类中心,蓝色不加粗虚线为该类型所有游程波形。其中上升游程波形图 5(a),下降游程波形图 5(f)为单峰尖顶游程波形呈现正/负半周正弦幅度变化;上升游程波形图 5(b)、(c),下降游程波形图 5(g)为双峰游程波形,波形前沿波动幅度较大;上升游程图 5(d)、(e),下降游程波形图 5(h)为多峰游程波形,波形前沿波动较快且逐渐趋于平缓。

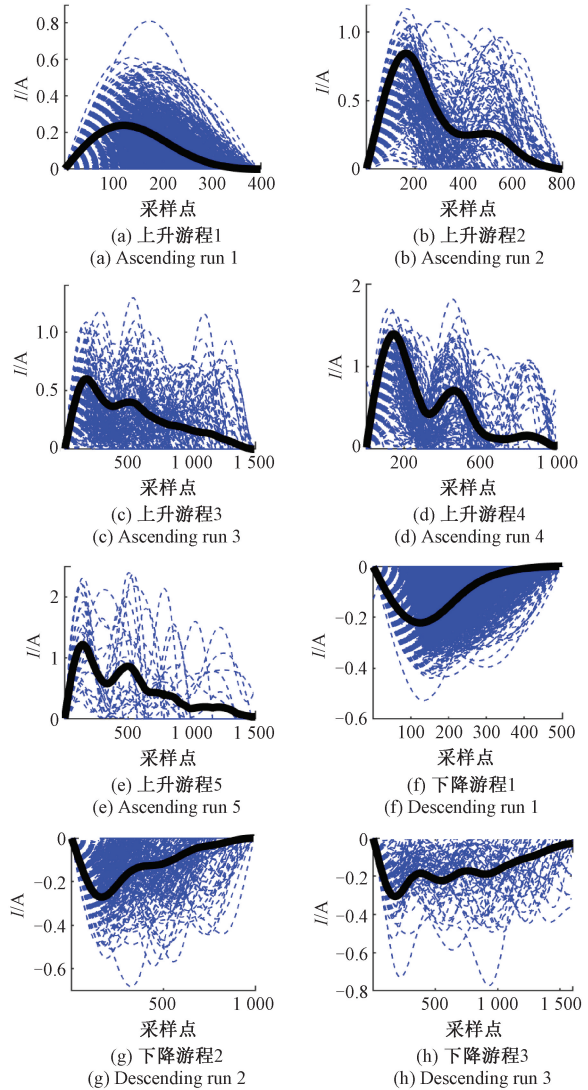


图 5 基于改进 GMM 的光伏电站游程聚类结果  
Fig. 5 Clustering result of runs of photovoltaic station based on improved GMM

为评估聚类结果的有效性,采用所提聚类结果评价方法进行评价,将传统 GMM 聚类算法<sup>[25]</sup>、K-means++ 聚类算法<sup>[26]</sup>、基于密度的聚类算法(density-based spatial clustering of applications with noise, DBSCAN)<sup>[26-27]</sup>和所提改进 GMM 聚类算法评分进行对比。其中,传统 GMM 是本文所提改进方法的基础;K-means++ 是一种改进的 K-means 方法,通

过优化初始中心点选择提升数据的聚类效果;DBSCAN 通过密度识别簇结构,可识别游程中任意形状的簇且能够自动

确定簇的数量。不同方法对上升游程聚类评分结果如表 1 所示,对下降游程聚类评分结果如表 2 所示。

表 1 光伏输出电流上升游程聚类评价结果

Table 1 Evaluation on ascending run clustering results of output current of photovoltaic station

聚类算法	k 值	评分指标评分				复合指标评分
		聚类质量	物理意义	计算效率	可解释性	
改进 GMM	5	2.84	2.68	1.98	2.96	2.82
	4	2.41	2.53	1.84	2.71	2.41
传统 GMM	5	2.49	2.58	1.92	2.76	2.52
	6	2.39	2.41	1.72	2.45	2.32
K-means++	3	1.87	1.81	1.87	1.97	1.88
	4	2.24	1.96	1.82	2.14	2.07
	5	1.98	2.13	1.76	1.83	1.97
DBSCAN	3	2.14	2.32	1.91	1.81	2.13

表 2 光伏输出电流下降游程聚类评价结果

Table 2 Evaluation on descending run clustering results of output current of photovoltaic station

聚类算法	k 值	评分指标评分				复合指标评分
		聚类质量	物理意义	计算效率	可解释性	
改进 GMM	3	2.97	1.88	1.83	2.13	2.33
	4	2.11	1.66	1.71	1.82	1.81
传统 GMM	5	2.22	1.71	1.79	1.86	1.91
	6	2.15	1.74	1.51	1.81	1.78
K-means++	3	1.57	1.61	1.78	1.87	1.63
	4	1.84	1.86	1.81	1.91	1.82
	5	1.68	1.81	1.76	1.79	1.77
DBSCAN	4	2.31	1.83	1.80	2.01	2.04

由表 1 可知,改进 GMM 聚类算法在聚类质量、物理意义、计算效率和可解释性等各个方面均优于传统 GMM、K-means++和 DBSCAN。传统 GMM 方法需要对比不同聚类数下的评分才能确定最佳的聚类数,同时受数据中噪声的影响,聚类效果不佳。而所提改进方法不仅可以根 据 BIC 自适应确定聚类数目,还通过深度学习技术提高了游程特征提取和抗干扰能力,复合指标评分相比于传统 GMM 最佳聚类结果( $k=5$ )提高 11.9%,各分项指标提升 1.6%~14.1%。

进一步与其他方法比较,所提改进 GMM 算法的复合指标评分相比 K-means++最佳聚类结果( $k=4$ )和 DBSCAN 聚类结果( $k=3$ )的复合指标评分分别提高了 36.2%和 32.40%,各分项指标提升 1.1%~61.4%。此外,尽管所提方法需要对特征矩阵深度训练,增加了总体运算时间,但考虑到传统 GMM、K-means++算法主观地选择聚类数时需要多次计算,因此所提方法计算效率仍然更优。

光伏电站短时波动信号易受外界噪声等环境干扰,为评估不同噪声水平场景下所提算法的聚类性能,对所采集的光伏电站输出电流分别加入不同水平的噪声和异常值,

以模拟实际环境中的噪声干扰以及突发性异常情况。通过随机方式将噪声加入到原始录波数据中,噪声强度分别为录播电流额定值的 1%、5%、10%、15%。其次,随机添加 50 个异常值到原始的录波数据中,异常值的幅值设定为录波电流额定值的 20%、50%、80%、100%,在不同噪声和异常值场景下,各聚类方法在最优聚类数目下的上升游程复合指标评价结果如图 6、7 所示。

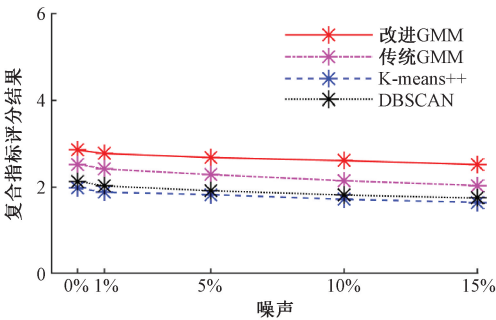


图 6 不同噪声水平光伏输出上升游程聚类结果评价  
Fig. 6 Evaluation on ascending run clustering results of photovoltaic output under different noise levels

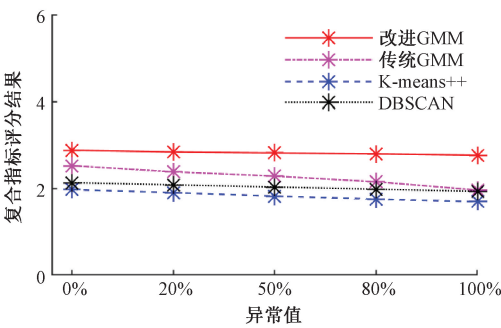


图 7 不同异常值水平下光伏输出上升游程聚类结果评价  
Fig. 7 Evaluation on ascending run clustering results of photovoltaic output under different outlier levels

分别计算不同噪声和异常值场景下的不同方法上升游程聚类结果复合指标评分相比于无噪声和异常值情况的下降程度,来反映噪声和异常值对算法的影响,结果如表 3、表 4 所示。其中传统 GMM 和 K-means++ 分别选择其最佳聚类结果进行对比。

表 3 噪声水平对光伏上升游程聚类结果的影响

Table 3 Influence of noise level on ascending run clustering results of photovoltaic station %				
噪声水平 / %	改进 GMM (k=5)	传统 GMM (k=5)	K-means++ (k=4)	DBSCAN (k=3)
1	2.85	4.37	4.84	5.16
5	6.09	9.52	7.61	10.33
10	8.59	15.08	13.20	15.02
15	11.75	19.44	16.67	18.31

表 4 异常值水平对光伏上升游程聚类结果的影响

Table 4 Influence of outlier level on ascending run clustering results of photovoltaic station %				
异常值水平 / %	改进 GMM (k=5)	传统 GMM (k=5)	K-means++ (k=4)	DBSCAN (k=3)
20	1.43	5.56	3.55	2.35
50	2.20	9.52	7.61	4.69
80	2.94	14.68	11.17	7.04
100	3.98	22.22	14.13	9.39

由图 6 和 7、表 3 和 4 结果可知,噪声水平从 1% 增加到 15% 时,改进 GMM 复合指标评分下降程度从 2.85% 下降到 11.75%,传统 GMM 则从 4.37% 下降到 19.44%,K-means++ 从 4.84% 下降到 16.67%,DBSCAN 从 5.16% 下降到 18.31%。尽管 4 种聚类算法复合指标评分都随着噪声强度增加而降低,但所提方法在不同噪声水平下评分

的下降程度小于其他算法 1.52%~7.69%,表明所提方法具有更强的抗噪声能力。

在异常值的影响方面,在异常值水平从 20% 增加到 100% 时,改进 GMM 复合指标评分下降程度从 1.43% 下降到 3.98%,传统 GMM 则从 5.56% 下降到 22.22%,K-means++ 从 3.55% 下降到 14.13%,DBSCAN 从 2.35% 下降到 9.39%,所提方法在不同异常值水平下评分的下降程度小于其他算法 0.92%~18.24%。由于所提方法通过深度学习训练去除了异常值,而 DBSCAN 对异常值的鲁棒性较强,因此二者聚类效果受异常值的影响较小,均小于 K-means++ 聚类算法和 GMM 聚类算法。同时所提改进 GMM 算法在异常值下的评分下降百分比低于 DBSCAN,表现出更强的抗干扰能力。

5 结 论

针对光伏电站输出短时波动特征随机性强、难以建模、不利于分析对电能计量影响的问题,提出一种基于改进 GMM 的光伏短时波动信号游程聚类分析方法。首先,基于光伏电站短时波动特点进行幅度域游程波形分解;其次,计算游程波形数字特征,构建游程波形数字特征值,最后,使用改进 GMM 聚类算法提取典型游程波形。通过对光伏电站实测数据的分析结果表明,相较于其他方法,所提方法聚类结果评分在不同方面均有 1.1%~61.4% 的提升,在不同噪声及异常值水平下所提方法也可以维持较好的聚类效果,复合指标评分下降程度小于其他算法 0.92%~18.24%。实验结果验证了所提方法通过结合深度学习技术和 BIC 的优势,提高了对含噪声和异常值的信号的适应能力和稳定性,克服了传统聚类方法处理包含异常值和噪声数据的不稳定性以及确定聚类数量存在主观性的问题,能够实现对光伏短时波动信号游程的合理聚类。所得结果可进一步用于光伏输出信号的重构和对电测量设备误差影响的评估中。

参考文献

[1] 王佳,吴任博,肖健,等.新型电力系统能源调度与碳排放计价优化方法研究[J].电测与仪表,2024,61(10): 17-25.  
WANG J, WU R B, XIAO J, et al. Research on the optimization method of energy dispatching and carbon emission pricing for novel power system[J]. Electrical Measurement & Instrumentation, 2024, 61 (10): 17-25.

[2] 周文斌,靳阳,刘士峰,等.能源互联网下电能计量系统改进技术研究[J].自动化仪表,2023,44(2): 102-105,110.  
ZHOU W B, JIN Y, LIU SH F, et al. Research on improvement technology of electric energy metering system under energy internet[J]. Process Automation

- Instrumentation, 2023, 44(2):102-105, 110.
- [3] 孙家奇,梁国勇,王聰,等. 基于交流储能的光伏柔直系统多时间尺度调压控制[J]. 电网与清洁能源, 2024, 40(7): 11-17, 27.
- SUN J Q, LIANG G Y, WANG C, et al. Multi-time scale voltage regulation control of photovoltage flexible HVDC system based on AC energy storage[J]. Power System and Clean Energy, 2024, 40(7):11-17, 27.
- [4] 姜海洋,杜尔顺,马佳豪,等. 考虑长周期供需不平衡风险的新型电力系统规划方法[J]. 中国电机工程学报, 2024, 44(15):5845-5858.
- JANG H Y, DU ER SH, MA J H, et al. Power system optimal planning method considering long-term imbalance risk[J]. Proceedings of the CSEE, 2024, 44(15):5845-5858.
- [5] 李建宜,李鹏,徐晓春,等. 基于综合概率模型与深度学习的智能电网功率-电压映射方法[J]. 电力建设, 2022, 43(2): 37-44.
- LI J Y, LI P, XU X CH, et al. Power-voltage mapping method based on comprehensive probability model and deep learning for smart grid[J]. Electric Power Construction, 2022, 43(2):37-44.
- [6] 黄越辉,曲凯,李驰,等. 基于 K-means MCMC 算法的中长期风电时间序列建模方法研究[J]. 电网技术, 2019, 43(7): 2469-2476.
- HUANG Y H, QU K, LI CH, et al. Research on modeling method of medium-and long-term wind power time series based on K-means MCMC algorithm[J]. Power System Technology, 2019, 43(7): 2469-2476.
- [7] 邵方静,宋晓林,刘坚,等. 动态电能计量算法研究综述[J]. 电测与仪表, 2023, 60(4): 1-10.
- SHAO F J, SONG X L, LIU J, et al. A review on dynamic energy metering algorithms [J]. Electrical Measurement & Instrumentation, 2023, 60(4): 1-10.
- [8] 王梓齐,张书瑶,刘长良. 基于增量式相对熵的风电机组实时状态监测[J]. 电子测量与仪器学报, 2020, 34(12): 125-132.
- WANG Z Q, ZHANG SH Y, LIU CH L. Real-time condition monitoring of wind turbine based on incremental relative entropy[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(12): 125-132.
- [9] KOUSOUNADIS-KNOUSEN M A, BAZIONIS I K, SOUDRIS D, et al. A new co-optimized hybrid model based on multi-objective optimization for probabilistic wind power forecasting in a spatio-temporal framework[J]. IEEE Access, 2023, 11:84885-84899.
- [10] 王智,贺星,苏玉萍,等. 基于小波分析的智能电表动态测试信号模型与误差分析[J]. 电测与仪表, 2024, 61(8): 203-210.
- WANG ZH, HE X, SU Y P, et al. Dynamic test signal model and error analysis of smart electricity meter based on wavelet analysis [J]. Electrical Measurement & Instrumentation, 2024, 61(8): 203-210.
- [11] 王飞,宋士瞻,曹永吉,等. 基于连续小波变换的风光发电资源多尺度评估[J]. 山东大学学报(工学版), 2018, 48(5): 124-130.
- WANG F, SONG SH ZH, CAO Y J, et al. Multi-scale assessment of wind-solar generation resources based on continuous wavelet transform[J]. Journal of Shandong University(Engineering Science), 2018, 48(5): 124-130.
- [12] 赵向阳,王杏玄. 抑制光伏功率对电网影响的电池管理研究[J]. 电力科学与工程, 2015, 31(9):40-46.
- ZHAO X Y, WANG X X. Research on battery management to reduce the effect of photovoltaic power on grid[J]. Electric Power Science and Engineering, 2015, 31(9):40-46.
- [13] 王学伟,刘建平,袁瑞铭,等. 复杂动态负荷电流幅度域典型游程波形模态与特征提取[J]. 电网技术, 2023, 47(6): 2497-2503.
- WANG X W, LIU J P, YUAN R M, et al. Extraction of typical run length waveform modes and characteristics of complex dynamic load current amplitude domain [J]. Power System Technology, 2023, 47(6):2497-2503.
- [14] 王学伟,顾鹏婷,袁瑞铭,等. 复杂动态负荷幅度域波形模态聚类与电能表误差敏感特征[J]. 北京化工大学学报(自然科学版), 2024, 51(1):92-100.
- WANG X W, GU P T, YUAN R M, et al. Waveform mode clustering in the amplitude domain of complex dynamic loads and sensitive characteristics of the errors in electricity meters[J]. Journal of Beijing University of Chemical Technology (Natural Science Edition), 2024, 51(1):92-100.
- [15] 李佩佩,崔凤英. 基于二分 K-means 聚类的曲率分级点云数据精简优化算法研究[J]. 电子测量技术, 2022, 45(4): 66-71.
- LI P P, CUI F Y. The optimization algorithm for curvature graded point cloud data based on dichotomous K-means clustering [J]. Electronic Measurement Technology, 2022, 45(4):66-71.
- [16] 丁明,黄冯,邹佳芯,等. 改进谱聚类与遗传算法相结合的电力时序曲线聚类方法[J]. 电力自动化设备,



- 2019, 39(2): 93-99,114.
- DING M, HUANG F, ZOU J X, et al. Power time series curve clustering method combining improved spectral clustering and genetic algorithm[J]. Electric Power Automation Equipment, 2019, 39(2): 93-99,114.
- [17] 徐超然,徐潇源,严正,等. 考虑风电统计特性挖掘的分布鲁棒优化调度方法[J]. 电力系统自动化,2022,46(2):33-42.
- XU CH R, XU X Y, YAN ZH, et al. Distributionally robust optimal dispatch method considering mining of wind power statistical characteristics[J]. Automation of Electric Power Systems, 2022,46(2):33-42.
- [18] 张兴,李明,郭梓暄,等. 新能源并网逆变器控制策略研究综述与展望[J]. 全球能源互联网, 2021, 4(5): 506-515.
- ZHANG X, LI M, GUO Z X, et al. Review and perspectives on control strategies for renewable energy grid-connected inverters[J]. Journal of Global Energy Interconnection, 2021, 4(5): 506-515.
- [19] 许津铭,凌子涵,程成,等. 并网逆变器非线性特性建模及稳定性研究综述[J]. 高电压技术, 2024, 50(1): 370-385.
- XU J M, LING Z H, CHENG CH, et al. Overview on nonlinear modeling and stability analysis of grid-tied inverters[J]. High Voltage Engineering, 2024, 50(1): 370-385.
- [20] 张东宁. 基于改进电导增量法的光伏最大功率点跟踪策略研究[J]. 太阳能学报, 2022, 43(8): 82-90.
- ZHANG D N. Research on photovoltaic maximum power point tracking strategy based on improved conductance increment method [J]. Acta Energiæ Solaris Sinica, 2022, 43(8): 82-90.
- [21] 王淑超,段胜朋,王健,等. 光伏电站面向快速频率响应的优化控制技术研究与实践[J]. 电力系统保护与控制, 2019, 47(14): 59-70.
- WANG SH CH, DUAN SH P, WANG J, et al. Research and practice of fast frequency response oriented control optimization technology of PV power stations[J]. Power System Protection and Control, 2019, 47(14): 59-70.
- [22] 舒珏淋,张力,胡建. 基于高斯混合模型的智能电表误差数据挖掘与分析方法[J]. 电子测量技术, 2021, 44(15):56-61.
- SHU J L, ZHANG L, HU J. Data mining and analysis method for smart meter error data based on Gaussian mixture model[J]. Electronic Measurement Technology, 2021,44(15):56-61.
- [23] SHI C M, WEI B T, WEI SH L, et al. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm[J]. Eurasip Journal on Wireless Communications and Networking, 2021, 1: 31.
- [24] BAI X Y, FAN Y F, HOU J J, et al. A comprehensive evaluation method for reliability confidence capacity of renewable energy based on improved DEMATEL-AHP-EWM [J]. Journal of Electrical Engineering & Technology, 2024, 19: 1205-1216.
- [25] 朱映秋,黄丹阳,张波. 基于高斯混合模型的分布因子聚类方法[J]. 统计研究,2024,41(6):147-160.
- ZHU Y Q, HUANG D Y, ZHANG B. Clustering with distributional factors based on gaussian mixture model[J]. Statistical Research, 2024,41(6):147-160.
- [26] 张玉琴,梁莉,张建亮,等. 基于改进 K-means++ 和 DBSCAN 的大数据聚类方法[J]. 国外电子测量技术, 2022, 41(9):40-46.
- ZHANG Y Q, LIANG L, ZHANG J L, et al. Big data clustering method based on improved K-means++ and DBSCAN[J]. Foreign Electronic Measurement Technology, 2022,41(9):40-46.
- [27] 李文杰,闫世强,蒋莹,等. 自适应确定 DBSCAN 算法参数的算法研究[J]. 计算机工程与应用,2019,55(5): 1-7,148.
- LI W J, YAN SH Q, JIANG Y, et al. Research on method of self-adaptive determination of DBSCAN algorithm parameters[J]. Computer Engineering and Applications, 2019,55(5):1-7,148.

## 作者简介

**彭文静**, 硕士研究生, 主要研究方向为新能源电力系统电能计量。

E-mail:19550272306@163.com

**郑迪**(通信作者), 讲师, 硕士生导师, 主要研究方向为新能源电力系统分析与控制。

E-mail:di\_zh@cjlu.edu.cn

**蔡慧**, 教授, 硕士生导师, 主要研究方向为电力计量与大数据分析。

E-mail: caihui@cjlu.edu.cn

**邵海明**, 研究员, 主要研究方向为高压电磁计量。

E-mail: shaozm@nim.ac.cn

**王家福**, 副研究员, 主要研究方向为高压电磁计量。

E-mail: jiafu\_wang@nim.ac.cn