

基于 DTW-OPTICS 分区的光伏集群输出方法的研究与预测

张家豪¹ 高庆忠²

1. 国家能源集团辽宁电力有限公司沈西热电厂, 辽宁 沈阳 110142

2. 沈阳工程学院, 辽宁 沈阳 110136

[摘要] 本文结合动态时间扭曲 (DTW) 和 OPTICS 算法, 提出了一种基于 DTW-OPTICS 分区簇的改进统计升级光伏簇输出预测建模方法。DTW 用于测量两个时间序列之间的相似性, 而 OPTICS 用于发现数据中的密度聚类结构。针对传统统计升级方法中子簇内光伏电站输出特性变化的问题, 通过皮尔逊相关系数和经验正交函数 (PCA) 优化特征空间, 使用 OPTICS 将该地区的光伏电站划分为簇, 提高一致性。本文提出了一种基于动态时间扭曲 (DTW) 的类似日期选择算法解决提取和预测待预测日期的权重系数时间序列的动态特征的问题, 并使用递归神经网络 (GRU) 模型预测光伏电站的输出。实验结果表明聚类预测方法的平均绝对百分比误差 (MAPE) 约为 6.35%。均方根误差 (RMSE) 约为 13.96kW, 均方误差 (MSE) 为 194.25kW。实际光伏电站数据证明了所提出方法的准确性和有效性。通过聚类分析和深度学习技术提高了光伏电站输出预测的准确性, 对于优化能源管理和提高可再生能源的利用效率具有重要意义。

[关键词] DTW-OPTICS; 光伏集群; 发电预测

DOI: 10.33142/sca.v7i12.14753

中图分类号: TM615

文献标识码: A

Research and Prediction of Photovoltaic Cluster Output Method Based on DTW-OPTICS Partition

ZHANG Jiahao¹, GAO Qingzhong²

1. Shenxi Thermal Power Plant of CHN Energy Group Liaoning Electric Power Co., Ltd., Shenyang, Liaoning, 110142, China

2. Shenyang Institute of Engineering, Shenyang, Liaoning, 110136, China

Abstract: This paper proposes an improved statistical upgrade photovoltaic cluster output prediction modeling method based on DTW-OPTICS partition clustering, combining dynamic time warping (DTW) and OPTICS algorithm. DTW is used to measure the similarity between two time series, while OPTICS is used to discover density clustering structures in the data. In response to the problem of changes in the output characteristics of photovoltaic power plants within a cluster in traditional statistical upgrading methods, the feature space is optimized using Pearson correlation coefficient and empirical orthogonal function (PCA), and OPTICS is used to divide the photovoltaic power plants in the region into clusters to improve consistency. This article proposes a date selection algorithm based on dynamic time warping (DTW) to solve the problem of extracting and predicting the dynamic features of weight coefficient time series of the predicted date, and uses a recursive neural network (GRU) model to predict the output of photovoltaic power plants. The experimental results show that the average absolute percentage error (MAPE) of the clustering prediction method is about 6.35%. The root mean square error (RMSE) is approximately 13.96kW, and the mean square error (MSE) is 194.25kW. The actual photovoltaic power plant data has demonstrated the accuracy and effectiveness of the proposed method. The accuracy of photovoltaic power plant output prediction has been improved through clustering analysis and deep learning techniques, which is of great significance for optimizing energy management and improving the utilization efficiency of renewable energy.

Keywords: DTW-OPTICS; photovoltaic cluster; power generation prediction

引言

准确预测分布式光伏发电集群的输出对于电网的稳定运行和能源管理至关重要。它可以帮助电网运营商优化电力调度, 减少对化石燃料的依赖, 并提高可再生能源的利用率。研究通常涉及集成多种预测模型和技术, 如时间序列分析、神经网络、支持向量机、随机森林等, 以提高预测的准确性和鲁棒性。预测结果可以用于智能控制策略, 如储能系统的充放电管理、需求响应等, 以进一步提高能源系统的效率和可靠性。

1 分布式光伏电站集群划分

本文按照分布式光伏单元对于每个单元光伏装机容量

量分别进行统计, 然后就近选取集中式光伏功率预测数据进行比对, 得到该区域内分布式光伏的功率预测数据, 即:

$$P_{DG} = [P_{DG, 1}, P_{DG, 2}, \dots, P_{DG, i}, \dots, P_{DG, n}] \quad (1)$$

P_{DG} 为次日 24h 全地区分布式光伏功率预测数据序列; $P_{DG, i}$ 为次日时刻 i 的分布式光伏功率预测值; n 为采样点个数, 本文取值 100, 即每间隔 15min 一个点。时刻 i 的分布式光伏功率预测值为:

$$P_{DG, i} = \sum_{j=1}^m P_{DG, ij} \quad \forall i \quad (2)$$

式中, $P_{DG, ij}$, 为时刻 i 第 j 个单元域内所有分布式

光伏的功率预测值； m 为单元区域的个数。时刻 i 第 j 个单元内所有分布式光伏的功率预测值为：

$$P_{DG,ij} = P_{CP,ij} \frac{R_{DG,j}}{R_{CP,j}} \quad \forall j \quad (3)$$

式中， $P_{CP,ij}$ ，为时刻 i 第 j 个区域内选定的某集中式光伏功率预测数据，该数据可从集中式光伏场站直接获取； $R_{CP,j}$ 为小区域 j 内选定的集中式光伏装机容量； $R_{DG,j}$ 为小区域 j 内分布式光伏装机总容量。

2 统调负荷预测

日前负荷预测一直是调度运行中的一项重要工作，通常需在当日对次日的负荷进行功率预测，此时可供参考的数据主要有过去几日的实际负荷、分布式光伏近似实际出力等。本文通过对比前日分布式光伏近似实际功率及次日预测功率曲线，在前日实际负荷功率曲线的基础上，计算得到次日的负荷功率预测曲线：

$$\text{即：} P_{IL} = P_{OL} + P_{ODG} - P_{IDG} \quad (4)$$

式中， P_{IL} 为次日负荷功率预测数据序列； P_{OL} 为前日负荷实际功率序列； P_{ODG} 为前日分布式光伏近似实际功率序列，可通过式 (1) ~ (3) 计算得到，但在计算时需将式 (3) 中的集中式光伏预测功率替换为前日该集中式光伏的实际功率； P_{IDG} 为次日分布式光伏功率预测数据，可通过式 (1) ~ (3) 计算得到。

3 基于 DTW 的光伏相似性指数

DTW 算法来计算不同光伏电站之间的时间序列相似性。DTW 是一种测量两个时间序列相似度的方法，它允许时间序列在时间轴上进行局部伸缩，以找到最佳匹配。计算每个光伏电站之间的皮尔逊相关系数，以评估它们之间的线性关系。使用 PAC 算法来重建相关系数，可能涉及到数据的预处理或特征提取步骤。结合发电站的地理位置、接入电网的母线等级等信息，在 OPTICS 聚类模型中输入优化后的特征并做类别区分从而提高特征的质量。OPTICS 是一种基于密度的聚类算法，能够识别出数据中的高密度区域。使用 DTW 算法来计算历史日期和待预测日期之间气象数据时间序列的相似性，选择相似性最高的序列作为待预测日期的权重时间序列。采用门控循环单元 (GRU) 神经网络模型来预测基准电站的出力。

3.1 基于 DTW 的类似日期选择

DTW 可用于计算等长和不等长时间序列的相似性。^[5] 并且对异常值不敏感。DTW 测量时间序列之间距离的方法算法采用自动识别时间序列中的最佳对齐点。对于长度为 M 和 N 的时间序列 X 和 Y ，生成时间节点矩阵，表达式为：

$$\{(X_1, Y_1), \dots, (X_m, Y_n)\}, m \in [1, M], n \in [1, N] \quad (5)$$

计算距离矩阵中时间节点对的距离，并将其距离之和最小化。在算法实现过程中，它通常被写为时间对应节点之间的基距离与到匹配对应时间节点的最小距离之和。公式为：

$$D_{dtw}(X_M, Y_N) = \min \left(\sum_{m,n=1}^{M,N} |X_m, Y_n| \right) \quad (6)$$

在这个公式中： D_{dtw} 是长度 M 和长度 N 之间的 DTW 距离； $|X_m, Y_n|$ 是 X 系列中第 m 个和 Y 系列中第 n 个之间的距离。

3.2 基于 PCA-OPTICS 的集群划分

特征降维与重构预测区域内共有 m 个分布式光伏电站，其空间位置相当于 m 个空间点。每个空间点有 n 个时间序列观测值。矩阵形式为：

$$X_{m \times n} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \quad (7)$$

对特征筛选后的分布式光伏电站数据进行集中处理，得到新的数据矩阵，通过线性变换计算协方差矩阵 R ：

$$R = \frac{1}{n} (Y)^T Y \quad (8)$$

计算矩阵 R 的特征值和特征向量，并根据特征值的降序进一步计算第 k 个主成分的贡献率和累积贡献率：

$$\beta_k = \frac{\lambda_k}{\sum_{n=1}^{\infty} \lambda_k} \quad (9)$$

$$\beta(t) = \sum_{k=1}^t \beta_k \quad (10)$$

3.3 OPTICS 集群划分

PCA-OPTICS 聚类划分方法可分为两部分：一是保留原始信息，DTM 相关系数特征筛选优化。二是降维后的数据优化，采用 OPTICS 聚类算法划分光伏电站簇别。

主要流程：(1) 进行主成分分析 PCA，用于减少数据集的维度，同时尽可能保留原始数据的变异性。(2) 应用基于密度的空间聚类 OPTICS 聚类算法，用于将数据点分组，形成簇。(3) 使用动态时间扭曲 (DTW) 算法处理数据矩阵。(4) 根据相关系数筛选，选择一组代表性的样本数据。(5) 核心矩阵的计算，处理矩阵以获得距离矩阵 X 和中心化处理后的点 Y 。(6) 协方差矩阵 R 的计算，计算样本点的协方差矩阵 R ，了解数据点之间的相关性。(7) 样本中心的计算，计算样本点的中心，进一步地分析或聚类。(8) 特征值和特征向量的计算，计算协方差矩阵 R 的特征值和特征向量，确定数据的主成分。(9) 累计贡献率的计算，计算特征值的累计贡献率，确定保留多少主成分以最大化数据的变异性。(10) 主成分的选择，根据累计贡献率选择主成分，以便在后续分析中使用。(11) 聚类划分，完成聚类划分，将样本数据分配到不同的簇中。

4 实例分析

4.1 单元参数特性

光伏集群输出预测的测试数据已辽宁某地区屋顶分

布式光伏电站的数据为依据,包括:光伏组件温度、日发电量、功率、统计电量、交直流电压、各回路组串数量等。本文选取 15 分钟为单元记录数据(时间为 2023 年 6 月至 2024 年 7 月),外在因素含环境温度、风速、风向和相对湿度。

4.2 实例结果的比较分析

以 2023 年 10 月 12 日为例,优化统计升级方法与传统统计升级方法的权重时间序列比较如图 1 所示。结果表明,与传统方法相比,权重时间序列的方差的显示数据明显有下降趋势,平稳地提供时间序列。集群内站点日输出数据得到优化。

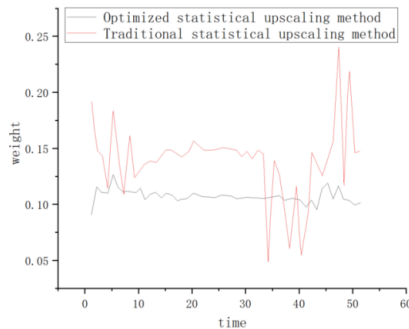


图 1 优化统计放大方法与传统统计放大方法的权重时间序列比较

在参考电站的光伏输出预测实验中,输入包括温度、天气类型、天气数据、板温、交流/直流电压/电流光伏器件数据,采样频率为 15 分钟,输出为预测日电站的功率。GRU 和其他模型的聚类预测结果如图 1 所示。

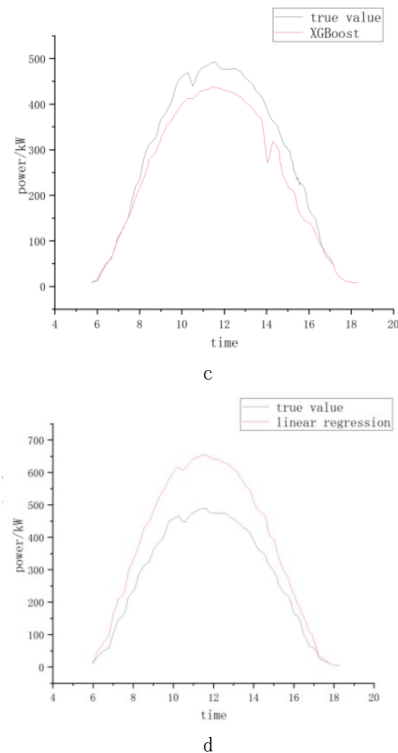
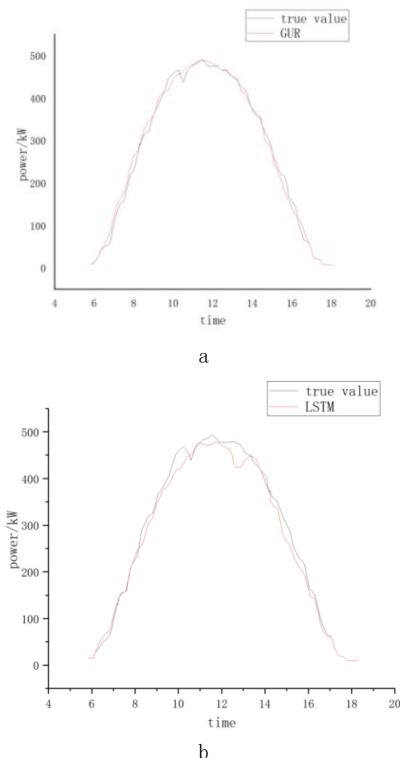


图 2 各模型预测结果比较

从图 2 中每个模型的预测结果可以看出,GRU 神经网络模型的预测效果优于其他预测模型。计算每个模型预测结果的 MAPE、RMSE 和 MSE,并进一步比较每个模型的预测精度。各模型的预测误差统计如表 1 所示。

表 1 GRU 和模型网络预测误差表

使用模型	/%	/kW	/kW
线性回归	28.98	102.13	10611.52
XGBoost	14.72	38.17	1632.12
LSTM	6.58	16.32	312.68
GRU	5.14	12.18	175.29

5 结论

对四种不同的预测模型进行了比较,包括 GRU(门控循环单元)网络模型、线性回归、XGBoost 和 LSTM(长短期记忆网络)。在比较中,所有模型使用的输入变量是一致的,以确保比较的公平性。与线性回归、XGBoost 和 LSTM 算法相比,GRU 网络模型显示出了改进,具有更高的预测精度。

预测结果表明,线性回归模型性能低,应用范围不广泛。相对比较 GRU 网络模型具有明细优势,预测精度更高。实验结果证明了结合了 DTM(可能是指 DTW,动态时间扭曲)、OPTICS 聚类 and GRU 的光伏集群输出预测方法的有效性。

基金项目:沈阳市科技计划项目(22-322-3-29)。

[参考文献]

[1]李振生,孙佳龙,卜晓刚,等.基于分布式光伏功率分区

拟合的统调负荷预测方法[J]. 电工技术, 2021(15): 18-20.

[2]肖传亮,赵波,周金辉,等.配电网中基于网络分区的高比例分布式光伏集群电压控制[J].电力系统自动化, 2017, 41(21): 147-155.

[3]王文倬,谢丁,谢醉冰,等.考虑集群划分的分布式光伏无功电压控制策略[J].浙江电力, 2024, 43(7): 64-75.

[4]陈霆,曾君,张轩,等.面向规模化分布式可再生能源并

网的群网协同互动优化方法[J].电力系统自动化, 2024(7): 1-12.

[5]张姍,冬雷,纪德洋,等.基于NWP相似性分析的超短期光伏发电功率预测[J].太阳能学报, 2022, 43(4): 142-147.

作者简介:张家豪(1999.3—),毕业于沈阳工程学院,本科,自动化专业,就职于国家能源集团辽宁电力有限公司沈西热电厂,集控巡检员,助理工程师。