DOI:10.20033/j.1003-7241.(2025)06-0018-05

基于灰色关联分析的分布式光伏发电出力短期预测算法

朱 齐，吕 正，王 沁，张 麟，沈 健，陈志樑

(国网上海市电力公司浦东供电公司，上海 200120)

摘 要：分布式光伏电站具备装机容量小和空间分布范围广泛的特点，但由于成本和技术条件等各种因素的约束，很多分布式光伏电站并缺乏单独而完备的电站监测体系，从而无法收集相应的数据信息对分布式电站光伏出力作出精准预测，研究基于灰色关联分析的分布式光伏发电出力短期预测算法。准备所需的分布式光伏发电出力历史相关数据，包括历史出力数据、气象数据与电力需求数据等；通过灰色关联分析方法整合全部数据，分别对历史出力数据与气象数据以及电力需求进行关联，确定影响光伏发电出力的关键性因素；根据灰色关联分析情况，在因素中提取与分布式光伏发电密切相关的特征；通过神经网络构建简化预测模型，使用相关特征与关键因素，模型可以针对待测日期或待测时间段，完成分布式光伏发电出力短期预测。结果表明，所研究方法可以实现不同地区分布式光伏发电出力短期预测，也可以实现不同天气类型下的分布式光伏发电出力短期预测，具有应用价值。

关键词：灰色关联；分布式光伏发电；光伏出力；短期预测

中图分类号：TP18 文献标识码：A 文章编号：1003−7241(2025)06−0018−05

Short Term Prediction Algorithm for Distributed Photovoltaic Power Generation Output Based on Grey Correlation Analysis

ZHU Qi, LYU Zheng, WANG Qin, ZHANG Lin, SHEN Jian, CHEN Zhi-liang

(State Grid Shanghai Pudong District Power Supply Co., Ltd., Shanghai 200120, China)

Abstract: Distributed photovoltaic power stations have the characteristics of small installed capacity and wide spatial distribution range. However, due to various constraints such as cost and technical conditions, many distributed photovoltaic power stations lack a separate and complete power station monitoring system, which makes it impossible to collect corresponding data information to make accurate predictions of the photovoltaic output of distributed power stations. A short-term prediction algorithm for distributed photovoltaic power generation output based on grey correlation analysis is studied. It prepares the necessary historical data related to distributed photovoltaic power generation output, including historical output data, meteorological data, and electricity demand data, integrates all data through grey relational analysis method, and correlates historical output data with meteorological data and electricity demand to determine the key factors affecting photovoltaic power generation output. It extracts features closely related to distributed photovoltaic power generation from the factors based on grey correlation analysis. By constructing a simplified prediction model using neural networks and utilizing the proposed relevant features and key factors, the model can achieve short-term prediction of distributed photovoltaic power generation output based on the test date or time period. The results indicate that the research method can achieve short-term prediction of distributed photovoltaic power generation output in different regions, as well as short-term prediction of distributed photovoltaic power generation output under different weather types, which has practical value.

Keywords: grey correlation; distributed photovoltaic power generation; photovoltaic output; short term prediction

0 引言

光伏发电出力短期预测的研究背景源于对可再生能源的大规模应用和智能电网建设的需求。随着全球对清洁能源的需求增加以及可再生能源发电装机容量的迅速增长，光伏发电成为了一个重要的能源来源。然而，由于天气、季节、时间和地理等因素的影响，光伏发电的出力存在着显著的波动性和不确定性。这使得系统运营和电力市场管理者面临着预测和调度的挑战。为了实现光伏发电的有效集成和最大化利用，光伏发电出力的短期预测成为了一个关键问题。光伏发电出力短期预测的研究致力于开发准确、可靠的预测模型和算法，以预测未来几小时至几天内的光伏发电出力。这些预测模型通常基于历史观测数据、气象数据、天文数据以及其他相关因素，并结合机器学习、统计分析和数学建模等技术手段进行分析和预测。

在此背景下，许多研究人员针对这一课题展开深入

收稿日期：2024-02-28

探讨，在文献[1]中设计了一种基于改进鸡群算法优化支持向量机(improved chicken swarm optimization–support vector machine，ICSO–SVM)的预测模型，主要通过动态惯性权重改进鸡群算法建立预测模型，在模型中首先分析辐射强度、温度对光伏出力的影响，然后获取待预报的时间与历史日期相关程度，最后通过预测模型中进行样本训练，针对所在区域光伏阵列预估出力，有较好的预估准确度，但仍有改善空间。在文献[2]中提出一种时空相关性出力预测算法，以注意力网络融入时空图网络技术建立预测模型，该算法主要对光伏出力的影响因素进行量化，并在注意力机制赋予新的权重，根据权重计算与历史数据进行一系列训练，实现了某地区分布式光伏系统的出力预测。为进一步实现在不同场景下或不同天气类型下的出力预测，且保证出力预测的准确率得到显著提高，此次选择灰色关联分析方式，设计分布式光伏发电出力短期预测算法，为实现可再生能源的有效利用和电力系统稳定运行提供理论支持。

1 获取分布式光伏发电出力相关历史数据

对分布式光伏发电出力的预测，需要依靠大量数据作为依托，既包含电力数据又包含气象数据，其中，电力数据主要为电压、电流和实时发电功率以及发电量等，气象数据包括温度、太阳辐射强度、风速和风向、温室气体排放等。大量数据在应用时会存在差异性，因此，采用时间序列方式获取相应光伏电力数据：

$$\left\{\dddot{q}\right\}=\left(\dddot{w},\dddot{e},\dddot{r},\dddot{t},\dddot{t}_{\max},\dddot{t}_{\min},\dddot{t}_{d},\dddot{u}\right) \tag{1}$$

$$\{y\}=\left(y_1,y_2,\cdots,y_w\right) \tag{2}$$

式中，w为给定的时间长度；$\{y\}$为非零的数据实值时间序列；$\left\{\dddot{q}\right\}$为序列内的数据矩阵，包含有$\dddot{w}$、$\dddot{e}$、$\dddot{r}$、$\dddot{t}$、$\dddot{t}_{\max}$、$\dddot{t}_{\min}$、$\dddot{t}_{d}$、$\dddot{u}$共8组数据变量[3]。在给定的时间序列长度下，对数据矩阵多维映射操作，整理对矩阵中相关光伏发电数据：

$$\begin{bmatrix}\dddot{w}\\ \dddot{e}\\ \dddot{r}\\ \dddot{u}\end{bmatrix}=\begin{bmatrix}\dddot{w}_1 & \dddot{w}_2 & \cdots & \dddot{w}_i\\ \dddot{e}_1 & \dddot{e}_2 & \cdots & \dddot{e}_i\\ \dddot{r}_1 & \dddot{r}_2 & \cdots & \dddot{r}_i\\ \dddot{u}_1 & \dddot{u}_2 & \cdots & \dddot{u}_i\end{bmatrix} \tag{3}$$

$$\begin{bmatrix}\dddot{t}\\ \dddot{t}_{\max}\\ \dddot{t}_{\min}\\ \dddot{t}_{d}\end{bmatrix}=\begin{bmatrix}\dddot{t}_1 & \dddot{t}_2 & \cdots & \dddot{t}_i\\ \dddot{t}_{\max,1} & \dddot{t}_{\max,2} & \cdots & \dddot{t}_{\max,i}\\ \dddot{t}_{\min,1} & \dddot{t}_{\min,2} & \cdots & \dddot{t}_{\min,i}\\ \dddot{t}_{d,1} & \dddot{t}_{d,2} & \cdots & \dddot{t}_{d,i}\end{bmatrix} \tag{4}$$

$$\left\{\dddot{q}_i\right\}=\left(\dddot{w}_i,\dddot{e}_i,\dddot{r}_i,\dddot{t}_i,\dddot{t}_{\max,i},\dddot{t}_{\min},\dddot{t}_{d},\dddot{u}_i\right) \tag{5}$$

式中，i为所在日期；$\dddot{q}_i$为第i日获取的相关数据矩阵；$\dddot{w}_i$为第i日发电功率；$\dddot{e}_i$为第i日太阳辐照度；$\dddot{r}_i$为第i日日照时数；$\dddot{t}_i$、$\dddot{t}_{\max,i}$、$\dddot{t}_{\min,i}$、$\dddot{t}_{d,i}$分别为平均环境温度、最高气温、最低气温、气温日差；$\dddot{u}_i$为第i日平均风速[4]。在获取和整理分布式光伏发电出力相关数据后，需要在数据中提取出与光伏发电短期出力有关的要素和密切特征，为更加贴合光伏发电特性，采用灰色关联方式进行分析。

2 确定影响短期出力关键因素

对于光伏电站系统来讲，其发电出力与多种因素有关，这些因素或多或少会影响发电出力趋势，在获取的相关数据中已经包含了相关因素，为确定哪一种因素对发电出力的影响较大或是较小，选择灰色关联方法对因素的重要程度排序。表示为

$$\left[\dddot{q}_i\right]=\left[\dddot{q}_i(1)\quad \dddot{q}_i(2)\quad \cdots\quad \dddot{q}_i(a)\right] \tag{6}$$

$$\left[\dddot{p}_i\right]=\left[\dddot{p}_i(1)\quad \dddot{p}_i(2)\quad \cdots\quad \dddot{p}_i(a)\right] \tag{7}$$

式中，$\dddot{q}_i$作为获取的相关数据矩阵，作为归一化表现方式，$\dddot{q}_i(a)$可以表示上述所有相关因素的数列，a为总列数；$\dddot{p}_i$为$\dddot{q}_i$对应的灰色关联矩阵；$\dddot{p}_i(a)$为灰色关联数列，同样也可以表示数据中的变量因素。根据设定的矩阵，以灰色关联分析方法确定因素之间的主次关系：

$$s(a)=\frac{\min\limits_{a}\min\limits_{i}\left|\dddot{p}_i(a)-\dddot{q}_i(a)\right|+g\max\limits_{a}\max\limits_{i}\left|\dddot{p}_i(a)-\dddot{q}_i(a)\right|}{\left|\dddot{p}_i(a)-\dddot{q}_i(a)\right|+g\max\limits_{a}\max\limits_{i}\left|\dddot{p}_i(a)-\dddot{q}_i(a)\right|} \tag{8}$$

式中，$s(a)$为关联函数；$\left|\dddot{p}_i(a)-\dddot{q}_i(a)\right|$为$\dddot{p}_i(a)$与$\dddot{q}_i(a)$在$a$列的绝对误差；$\min\limits_{a}\min\limits_{i}\left|\dddot{p}_i(a)-\dddot{q}_i(a)\right|$为两个相关因素在$a$列最小差；$\max\limits_{a}\max\limits_{i}\left|\dddot{p}_i(a)-\dddot{q}_i(a)\right|$为$\dddot{p}_i(a)$与$\dddot{q}_i(a)$在$a$列的最大差；$g$为关联系数[5–7]。

将所有影响因素通过灰色关联分析方式对相关数据进行灰色关联，在取值过程中需要注意，每一次在a列的关联时，需要保证$\dddot{p}_i(a)\neq\dddot{q}_i(a)$，才能确定某一个数据变量的重要程度，相比较之下数值越大的重要程度越高。根据不同数据之间的关联分析结果，在关键因素中提取与分布式光伏发电短期出力最为密切的特征。

3 提取分布式光伏发电短期出力密切特征

不同影响因素在不同时段内对光伏电站出力的影响同样具有差异性，为此，为减少发电出力的预测随机性，需要在不同因素中提取出与分布式光伏发电短期出力的密切特征，这一过程应用马尔科夫链来描述，过程如下：

$$h\left(\overset{\cdots}{q_i}(a)\middle|\overset{\cdots}{p_1}(a),\overset{\cdots}{p_2}(a)\cdots,\overset{\cdots}{p_i}(a)\right)=h\left(\overset{\cdots}{q_i}(a)\middle|\overset{\cdots}{p_i}(a)\right) \tag{9}$$

$$\begin{aligned}&h\left[\overset{\cdots}{q_i}(a)_{t+\Delta t}=\overset{\cdots}{q_i}\middle|\overset{\cdots}{q_i}(a)_{\Delta t}=\overset{\cdots}{p_i}\right]=\\&h\left[\overset{\cdots}{q_i}(a)_{\Delta t}=\overset{\cdots}{p_i}\middle|\overset{\cdots}{q_i}(1)=\overset{\cdots}{p_i}\right]=h_{\overset{\cdots}{q_i},\overset{\cdots}{p_i}}(\Delta t)\end{aligned} \tag{10}$$

式中，$t+\Delta t$为时间段；Δt为时间差；$h_{\overset{\cdots}{q_i},\overset{\cdots}{p_i}}(\Delta t)$为$\Delta t$内$\overset{\cdots}{p_i}$、同$\overset{\cdots}{q_i}$转移的概率函数[8]。设定在转移空间内涵盖有多种变化形式，即与光伏发电出力的密切特征会发生转变，则令每个截取时间段为等同状态，建立密切特征提取矩阵，同样用条件概率表示：

$$h=h_{\overset{\cdots}{q_i},\overset{\cdots}{p_i}}=\begin{bmatrix}h_{\overset{\cdots}{q_1},\overset{\cdots}{p_2}} & h_{\overset{\cdots}{q_1},\overset{\cdots}{p_3}} & \cdots & h_{\overset{\cdots}{q_1},\overset{\cdots}{p_i}}\\ h_{\overset{\cdots}{q_2},\overset{\cdots}{p_3}} & h_{\overset{\cdots}{q_2},\overset{\cdots}{p_3}} & \cdots & h_{\overset{\cdots}{q_2},\overset{\cdots}{p_i}}\\ \vdots & \vdots & \vdots & \vdots\\ h_{\overset{\cdots}{q_i},\overset{\cdots}{p_1}} & h_{\overset{\cdots}{q_i},\overset{\cdots}{p_2}} & \cdots & h_{\overset{\cdots}{q_i},\overset{\cdots}{p_{i-1}}}\end{bmatrix} \tag{11}$$

$$\sum_{p=8,q=8}^{i}h_{\overset{\cdots}{q_i},\overset{\cdots}{p_i}}=1,h_{\overset{\cdots}{q_i},\overset{\cdots}{p_i}}>0 \tag{12}$$

$$h_{\overset{\cdots}{q_i},\overset{\cdots}{p_i}}=h\left(\overset{\cdots}{w_i},\overset{\cdots}{e_i},\overset{\cdots}{r_i},\overset{\cdots}{t_i},\overset{\cdots}{t_{\max,i}},\overset{\cdots}{t_{\min}},\overset{\cdots}{t_d},\overset{\cdots}{u_i}\right)\to h(w',e',r',t',u') \tag{13}$$

式中，对$\overset{\cdots}{q_i}$、$\overset{\cdots}{p_i}$的选择为上文提出的数据相关因素，即$\overset{\cdots}{w_i}$、$\overset{\cdots}{e_i}$、$\overset{\cdots}{r_i}$、$\overset{\cdots}{t_i}$、$\overset{\cdots}{t_{\max,i}}$、$\overset{\cdots}{t_{\min,i}}$、$\overset{\cdots}{t_{d,i}}$、$\overset{\cdots}{u_i}$，最终提取出来5个密切特征，其中$w'$为发电功率；$e'$为辐照度、$r'$为日照时间、$t'$为环境温度、$u'$为风速[9−12]。根据提取的相关密切特征，为实现具体日期的光伏发电出力预测，需要通过特征与相关数据进行计算，最终实现预测目的光伏发电出力短期预测。

4 基于神经网络预测分布式光伏发电短期出力

利用灰色关联分析出的具体结果和情况，提取出与分布式光伏发电出力密切相关的特征，根据全部特征和历史出力数据，将其应用在待测日期或时间段的短期出力预测，这一过程中由于数据的量级关系，需要应用一个预测模型对数据进行训练，此次选择神经网络构建简易模型实现短期出力预测[13−14]。模型见图1。

如图1所示，此次选择3层人工神经网络，第1层为输入层、第2层为隐含层、第3层为输出层，输入层中为密切特征，隐含层为设定的标准历史数据样本，输出层则有具体的光伏出力结果[15]。具体计算过程如下：

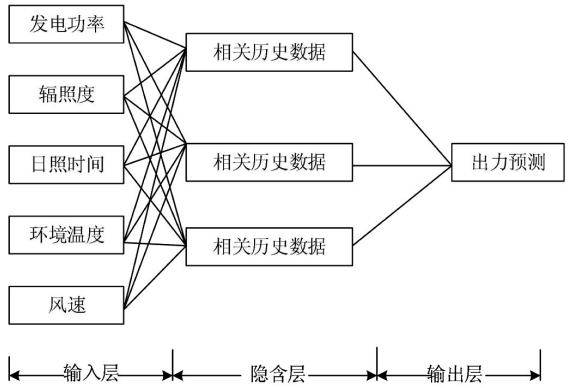


图1 人工神经网络简易预测模型

$$\{x\}=(w',e',r',t',u') \tag{14}$$

$$c_v=\frac{1}{2}b\left(x,\overset{\cdots}{q_i}\right)+m \tag{15}$$

式中，$\{x\}$为密切特征集合；v为待测日期；c_v为最终输出数值；$b(\cdot)$为距离函数；n为权值；m为误差。至此，本文在基于灰色关联分析基础上实现分布式光伏发电出力短期预测算法设计。

5 实验测试分析

上文基于灰色关联分析方法设计了一个新的预测方法，为验证所研究方法可以实现分布式光伏发电出力短期预测，应用MATLAB测试平台搭建测试环境，并设定测试算例。选择实际运行的分布式光伏电站作为参考实例，在获取光伏电站运行数据的基础上，完成所研究方法的验算和测试。此次选择两个地区的分布式光伏电站，这两个地区在地势和气候上具有差异性，在这两组电站中选择的数据来源情况，如下：

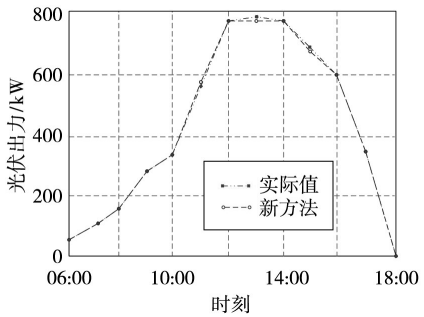
(1) A地区分布式光伏电站：一部分采用2022年8月1日到2023年9月30日数据，包括光伏出力数据和气象数据，其中，气象数据包括温度、辐照、相对湿度、降雨量、云量；另一部分数据选自2023年10月1日到11月30日，包括光伏出力、温度数据和辐照数据。

(2) Z地区分布式光伏电站：来源数据日期为2022年10月31日到2023年11月30日，分别为光伏出力数据和气象数据。

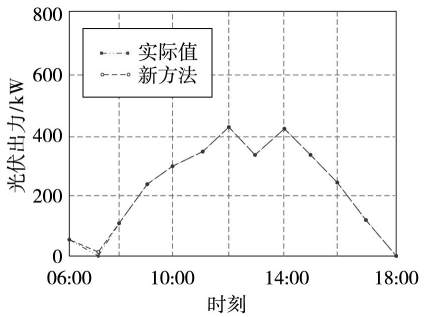
根据两组光伏电站选择的历史数据，此次实验算例共分为两个部分，第一部分验算所研究方法的应用效果，即其是否可以对两组光伏电站的出力进行短期预测；第二部分增加两组传统预测方法，针对不同天气类型的光伏出力进行短期预测，验证不同方法的预测精度。除此之外，考虑夜间不存在光照条件，即夜晚期间光伏电站不存在光伏出力输出，因此测算中仅展示6:00到18:00的

光伏出力情况。

在第一部分的验算过程中，A、Z地区分布式光伏电站出力预测日，分别选定为2023年12月5日和2023年12月8日，这两组日期在A、Z地区分布式光伏电站中已经存在真实数据，具有先验条件。经过气象局监控数据调取结果，在这两天该地区的天气均为晴天，将所研究方法连接在测试平台中，分别对A、Z地区分布式光伏电站的短期出力情况进行预测，结果见图2。



(a) A地区光伏电站2023年12月5日出力预测



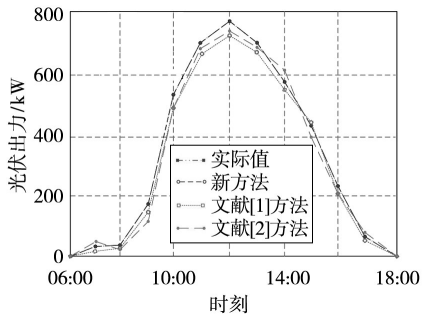
(b) Z地区光伏电站2023年12月8日出力预测

图2 不同地区分布式光伏电站出力短期预测结果

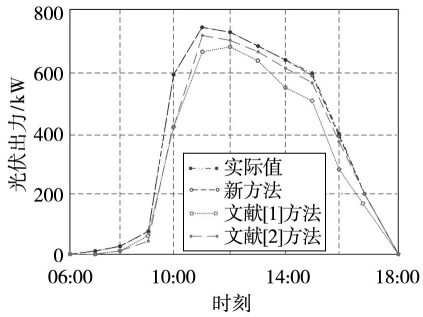
如图2所示，由于两个光伏电站所在地区具有差异性，选择的预测日期也不同，因此A、Z地区光伏电站出力情况具有差异性。在所研究方法应用下，能够对不同地区的光伏电站的不同日期的光伏出力情况进行预测，得出的预测结果与实际情况具有一致性，说明所研究方法可以实现分布式光伏电站发电短期预测。

为进一步验算所研究方法的应用效果，在第二部分针对不同天气类型的光伏出力完成短期预测，选择的对照组为文献[1]和[2]中提出的预测方法，分别为基于改进CSO算法的预测算法、基于时空图网络的预测算法。在这一阶段仅以A地区分布式光伏电站的历史数据作为参考，主要是其数据来源较为全面。不同天气类型的选择为晴天、阴天、雪天、多云类型，日期对应为12月2日、12月4日、12月9日、12月11日。具体结果见图3。

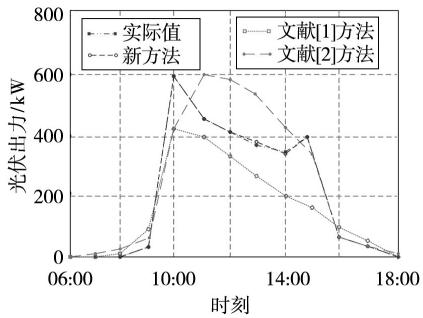
如3图所示，对A地区分布式光伏电站的出力情况预测，按照晴天、阴天。雪天、多云这4类天气划分，两组传统算法仅在晴天天气类型时可以完成较为精准的预测，当天气类型为阴天和雪天时，预测效果较差，而多云时预测误差最大。所研究方法的预测结果基本与实际值相一致，且不受天气类型的影响，说明通过灰色关联方法可以最大限度地贴近天气类型特征，保证光伏发电出力的预测精准度。



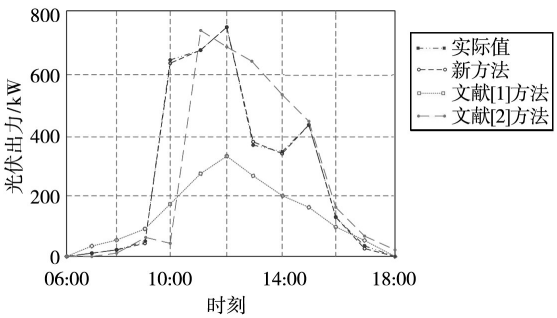
(a) 晴天类型



(b) 阴天类型



(c) 雪天类型



(d) 多云类型

图3 不同方法对A地区分布式光伏电站出力短期预测

6 结束语

为实现分布式光伏发电出力短期预测，采用灰色关

(下转第84页)