

基于ANN架构的新能源发电预测模型的研究

杨 强¹, 颜宗辉¹, 杜秀举¹, 吴胜安¹, 牟景艳¹, 李培霞²

(1. 贵州电网有限责任公司安顺供电局, 贵州 安顺 561000; 2. 东方电子股份有限公司, 山东 烟台 264000)

摘要: 以风能和太阳能为代表的新能源具有随机性、间歇性和波动性, 对新能源发电功率进行预测是有效解决以上问题的途径。为解决可再生能源 (Renewable Energies, REs) 在电力网络中不断整合对电力供应系统去中心化带来的问题, 提出了一种基于神经计算的预测模型。首先, 获取不同的数据集, 并介绍了数据使用方法; 其次, 对不同的人工神经网络 (Artificial Neural Network, ANN) 架构进行了深入研究, 同时探讨了不同的输入特征, 以及对预测结果不确定性的评估方法; 最后, 通过模拟实验表明在光伏发电方面, 本模型的预测相对平均绝对误差 (Relative Mean Absolute Error, RMAE) 可以在夏季达到 2.5%, 在冬季达到 0.5%。而在整个年度内, 通过使用输入类别特征的减小均值, 光伏和风力发电的 RMAE 分别可以达到 1.7% 和 4.9%。这一方法的实施为地区可再生能源的集成提供了有力支持, 同时也为应对可再生能源的波动性和不可控性提供了新的解决思路。

关键词: 可再生能源; 人工神经网络; 预测模型; 平均绝对误差

Research on the prediction model of new energy power generation based on ANN architecture

YANG Qiang¹, YAN Zonghui¹, DU Xiuju¹, WU Sheng'an¹, MU Jingyan¹, LI Peixia²

(1. Anshun Power Supply Bureau of Guizhou Power Grid Co., Ltd., Anshun 561000, China;

2. Dongfang Electronics Co., Ltd., Yantai 264000, China)

Abstract: The new energy represented by wind energy and solar energy is random, intermittent and fluctuating. The prediction of new energy power generation is an effective way to solve the above problems. In order to solve the problem of decentralization of power supply system caused by the continuous integration of renewable energies (REs) in power network, this paper proposes a prediction model based on neural computing. Firstly, different data sets are obtained and the data usage method is introduced. Secondly, different artificial neural network (ANN) architectures are studied in depth, and different input characteristics and evaluation methods for uncertainty of prediction results are discussed. Finally, through simulation experiments, it is shown that in terms of photovoltaic power generation, the relative mean absolute error (RMAE) of the prediction of this model can reach 2.5 % in summer and 0.5 % in winter. In the whole year, the RMAE of photovoltaic and wind power can reach 1.7 % and 4.9 % respectively by using the reduced mean of input category features. The implementation of this method provides strong support for the integration of regional renewable energy, and also provides a new solution to deal with the volatility and uncontrollability of renewable energy.

Key words: renewable energys; artificial neural network; prediction model; relative mean absolute error

0 引言

全球能源危机和环境问题的日益加剧，新能源技术得到快速发展。太阳能光伏以及风力发电等新能源发电技术具有无污染、地域灵活、安全可靠以及维护简单等优点而得到广泛应用。但受到光照强度、风速、温度等气象因素及光伏阵列、风机安装位置、组件转换效率等地理因素的影响，发电功率具有随机性、间断性及不稳定性等特点^[1-3]，导致系统大规模并网运行后对电网产生冲击，无法安全稳定运行^[4-6]。提升新能源风光发电技术水平，保证风光发电能发挥出应有的作用价值，对完善我国能源供给结构具有十分重要的意义^[7]。但是，风光发电具有一定的不确定性和波动性，通过对新能源发电功率进行预测，有效降低新能源并网对电网产生的不利影响，实现风光发电最大化效益，对保证电力系统安全稳定运行具有重要意义^[8-10]。

目前，新能源发电功率预测技术越来越受到国内外专家和学者的重视，针对此方法大体可分为物理法和统计法^[11]。物理法是基于光伏发电系统建立精确的物理模型，然后将气象预测数据输入到模型中，通过方程映射完成预测；统计法是分析历史功率和气象数据，发现其内在联系，完成预测。近年来，基于统计方法的传统预测模型应用较为广泛，例如灰色神经网络、小波神经网络、支持向量机等^[12]。

传统神经网络训练方式复杂，收敛速度慢，容易陷入局部最小，本文提出了一种基于 ANN 架构的新能源发电预测模型，通过将时间一致性的天气数据作为功率预测的输入特征，考虑装机容量随观测时段的发展，可以独立于装机容量预测区域归一化功率输出。探讨了不同的输入特征，以及对预测结果不确定性的评估方法，以平均绝对误差作为判据，通过模拟实验验证了所提模型预测的准确性。

1 气象数据

风速对光伏发电的主要影响体现在散热条

件上，影响光伏电池板的通风散热效果。此外，风速决定了云团的移动，而云团的移动会影响太阳辐照度和光伏板上的阴影。风电出力与风速之间的关系通常较为复杂，且具有非线性特点。风电功率与空气温度、太阳辐照度的线性相关性不大，但间接影响风速、空气密度等。因此，可以将这两个天气参数扩展作为风电功率估计的输入特征。

2 方法论

2.1 人工神经网络

一般来说，人工神经网络(ANN)分为三个部分：输入层、输出层和隐藏层。ANN通过结合大量被称为神经元的计算单元来估计输入特征 x 和目标输出 y 之间的关系。这种关系的描述体现在权重 ω 和偏置 b 因子（即向量）上。利用反向传播算法训练可训练参数（权重 ω 和偏置 b ）。

基于迭代获得的代价函数对ANN参数进行微调，以最小化预测值与真实目标值之间的误差率，并使模型更加可靠。一个过程层 i 的输出可以描述为

$$y_i = \text{AF}(w_i x_i + b_i) \quad (1)$$

式中， x_i 、 y_i 分别为第 i 层的输入和输出；AF是一种激活函数。使用激活函数的目的是使ANN能够处理复杂的非线性问题。使用最多的非线性AF是ReLU和Sigmoid函数。

2.2 支持向量机

支持向量机（Support Vector Machine，SVM）是用于分类和回归的监督学习方法之一。其主要思想是从一个或多个特征向量中确定一个函数或一个超平面（即曲面），并在训练数据集（向表面偏移）上最小化误差。事实上，大多数问题都不是线性的，满足这样条件的超平面是根本不存在的。对于非线性情况，将相关核函数引入到SVM方法中，通过将数据映射到高维空间解决原始空间中线性不可分的问题，在这项研究中，高斯径向基函数的默认核函数被用于

回归。

2.3 多元线性回归

多元线性回归 (Multiple Linear Regression, MLR) 方法旨在建立自变量 (输入特征) 和响应变量 (输出目标) 之间的线性关系。MLR 试图拟合带有系数的线性模型, 以最小化预测值与真实观测目标值之间的残差平方和。时间序列数据的 MLR 模型可表示为

$$y(t) = \beta_0 + \beta_1 x_1(t) + \dots + \beta_k x_k(t) + \epsilon(t) \quad (2)$$

式中, $y(t)$ 为预测值; $x_k(t)$ 为输入特征; β_k 为每个输入特征的斜率系数; β_0 为常数; $\epsilon(t)$ 为残差。

2.4 k-最近邻

k-最近邻 (k-Nearest Neighbors, kNN) 方法被用于解决基于 n 维空间的分类和回归任务, 代表了样本的 n 个特征。为了预测未知样本的标签, kNN 算法基于距离计算找到一组 k 个最近的实例, 例如欧氏距离。具有相似性质的实例在 n 维空间中根据特征向量相互靠近。新实例将被分配到其最近实例的多数类中。在本研究中, 将 k 值设定为 3。

2.5 决策树

决策树回归方法 (Decision Tree Regression, DTR) 的原理: 通过学习从样本特征推断的简单决策规则对目标值进行建模, 从而创建一个流图式的树结构模型。该方法可以看成是 if-then 规则的集合, 也可以根据样本特征看成是条件概率分布。决策树生成算法递归地生成一棵树结构, 直到无法继续。然而, 这种方式生成的复杂且深层的树结构可能会导致过拟合。因此, 本文将树的最大深度设置为 5。

3 ANN架构设置

本文的 ANN 结构包含 3 个隐藏层, 每个隐藏层有 10 个神经元单元。输入层神经元的大小取决于相应的输入特征的数量。

对于时间序列数据, 本文使用时间一致模型作为输入特征和输出目标值。整个可用的 x 和 y 数据集用时间戳连接, 分为训练和验证两部分。

3.1 LA算法的计算

为独立获取输入特征, 使用一种小型地震检波器缩放器。输入参数 (特征) 在 $[0,1]$ 的区间范围内取值, 因为不同的输入数据分布和规模会导致较大的权重值和较差的训练结果。小范围的缩放值更适合在神经网络模型中训练。输入特征的变换过程为

$$x_{\text{scaled}}(t) = \frac{x(t) - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

将归一化功率 (如 $P_{\text{norm}}^{\text{PV}}(t)$) 作为输出参数, 计算式为

$$P_{\text{norm}}^{\text{PV}}(t) = \frac{P^{\text{PV}}(t)}{P_{\text{inst}}^{\text{PV}}(t)} \quad (4)$$

同样地, 将风电功率转化为归一化功率。将发电功率的时间序列按其装机容量进行归一化处理, 以比较独立于各分布式能源各自规模的功率输出曲线。最后, 将预测的功率输出与区域装机容量反变换为基于 E_q 的功率值。

3.2 模型集

天气参数对 RES 发电量的影响不同, 根据输入特征和输入层神经元可以识别来自不同环境的信息, 这里是测量天气数据的时间序列。输入特征分为 3 个不同的类别 (Category, CT)。设置不同的天气数据作为 ANN 模型的输入特征。CT-1 覆盖了输入特征, 属于高度相关的天气数据类别。对于 PV 预测模型和风速预测模型, 分别引入区域太阳辐照度和风速。CT-2 和 CT-3 用另外两个天气参数对输入特征进行了扩展。

为减少气象站测量对整个区域功率预测的影响, 本文设计了 3 个额外的 ANN 模型, 分别接收 CT-1、CT-2 和 CT-3 的输入特征的平均值。它们的输入类别分别称为 CT-1-mean、CT-2-mean 和 CT-3-mean。同时, 平均类别的模型具有较少的输入特征和后续计算。

3.3 评价指标

平均绝对误差 (Mean Absolute Error, MAE) 和平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 是预测最常用的指

标。为了更好地比较预测性能，计算了归一化预测误差 $\varepsilon(t)$ 。此外，还引入了装机容量的相对平均绝对误差（RMAE）。其数学表达式为

$$\varepsilon(t) = \frac{P_{\text{pred}}(t) - P_{\text{true}}(t)}{P_{\text{inst}}(t)} \quad (5)$$

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |P_{\text{pred}}(t) - P_{\text{true}}(t)| \quad (6)$$

$$\text{RMAE} = \frac{1}{N} \sum_{t=1}^N \left| \frac{P_{\text{pred}}(t) - P_{\text{true}}(t)}{P_{\text{inst}}(t)} \right| \quad (7)$$

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N \left| \frac{P_{\text{pred}}(t) - P_{\text{true}}(t)}{P_{\text{true}}(t)} \right| \quad (8)$$

式中， N 为预测数据集的样本量； $P_{\text{pred}}(t)$ 为预测功率输出； $P_{\text{true}}(t)$ 为发电量的真实测量值； $P_{\text{inst}}(t)$ 为观测RES的区域装机容量。

4 结果与讨论

在实际应用中，将有效的历史数据分为两个数据集：70%作为训练数据集，30%作为验证数据集。在TensorFlow中对ANN进行了仿真。首先，研究了用于功率预测的不同输入类别。在每个训练阶段使用不同的随机数据进行仿真。

图1比较了2021年6月和12月两个选定时段

的预测和实际测量光伏功率的分布。采用原始或均值分类数据的ANN模型可以跟踪夏季和冬季光伏发电量的变化。图2给出了风电功率预测。不同输入特征的ANN模型在大风期和平静期也有很好的效果。

对于每个实验，用验证数据集对建立的训练模型进行了测试。基于不同输入特征，训练和验证数据集的区域功率预测误差结果表明，光伏和风电预测的训练集和验证集的MAE和RMAE值相差不大。模型训练良好，没有出现拟合现象。光伏训练模型相对于区域装机容量的最小RMAE可以达到1.456%。所有原始实测气象数据（CT-3）作为输入的ANN风速模型表现最好，RMAE为3.465%。

通过使用所有原始实测天气数据（三类），CT-3输入特征的ANN模型效果优于CT-1和CT-2。

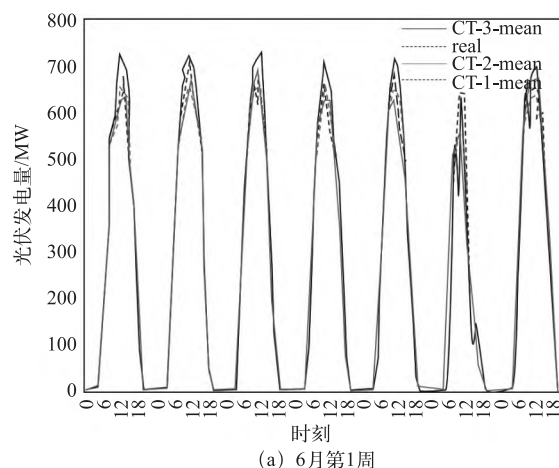
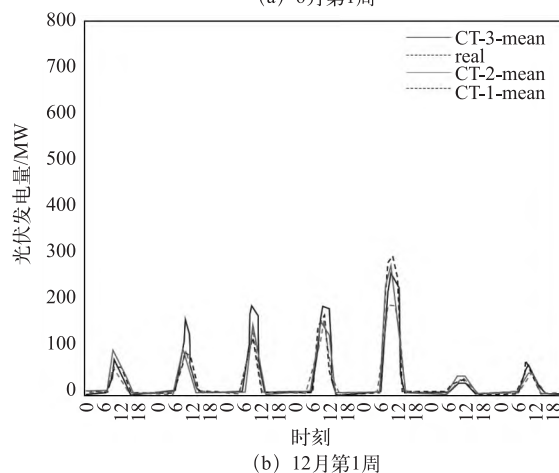


图1 光伏发电功率预测

Fig. 1 Photovoltaic power generation prediction

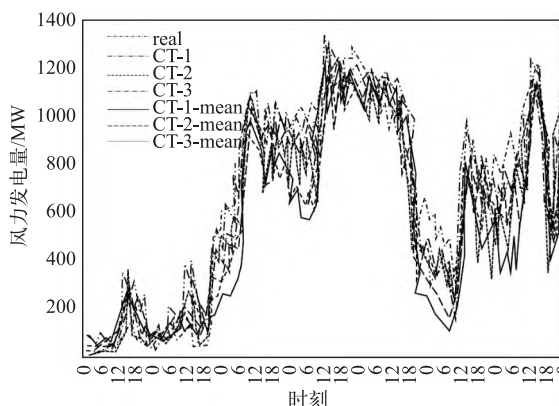


图2 风力发电功率预测

Fig. 2 Wind power prediction

输入特征越多，预测效果越好。ANN模型中的神经元可以从不同类型的天气数据中识别出更多的信息。扩展的天气参数可以提高预测性能。同时，根据输入特征，将输入数据减

少到平均值 (CT-1-mean、CT-2-mean、CT-3-mean)。与具有13个输入特征的CT-1相比,仅使用两个输入特征的CT-2-mean训练得到的模型具有更小的RMAE。使用平均太阳辐照度和空气温度作为输入特征的ANN光伏模型,无论是原始测量数据还是平均值,都比仅使用太阳辐照度作为输入数据的模型表现更好。

在初步分析的基础上,计算了全年预测偏差的分布情况。归一化误差的累积分布如图3所示。它显示了不同类别不断增加的输入特征如何影响预测的全局性能。S7给出了 $\varepsilon(t)>10\%$ 、 $\varepsilon(t)<10\%$ 的小时数和在误差范围内的百分比($\pm 10\%$)。在 $\pm 10\%$ 的归一化误差容许范围内的光伏预测分布均在96%以上。其中,输入特征为CT-3的模型预测效果最好(为97%),其同样可以通过CT-2-mean输入的ANN模型实现。

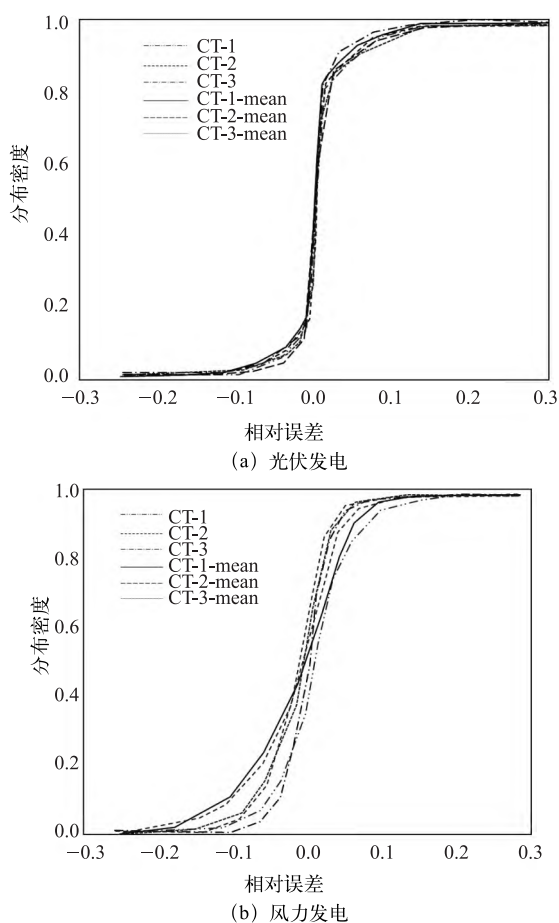


图3 归一化误差的累积分布

Fig. 3 Normalized error cumulative distribution power prediction

然而,风电功率预测与输入特征不同。与其他模型输入相比,使用CT-3输入的模型在风速预测中表现良好。超94%的预测样本处于许可范围内($\pm 10\%$)。同时,天气参数的平均值(CT-3-mean)相较于只有一个高相关参数的输入特征性能更优。

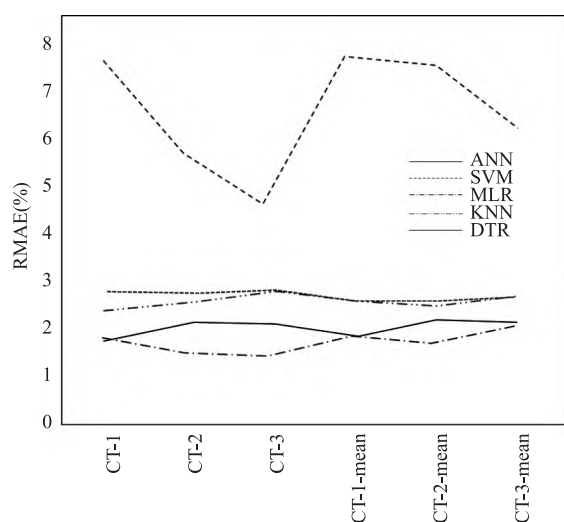
在光伏功率预测方面,夏季比冬季的RMAE值更高。这可能是由于夏季光伏发电量高于冬季,预测误差较大,导致RMAE较大。此外,夏季天气变化复杂,对于每个输入量,风电功率预测的整体性能在全年更加稳定。总体而言,三类天气参数的组合在不同季节的光伏和风电预测中预测效果更优。

为进一步考察所提ANN方法的有效性,本文还引入了其他常用的机器学习方法,如支持向量机(SVM)回归、多元线性回归(MLR)、k-近邻(kNN)和决策树回归(DTR)等方法。这些方法使用与ANN模型相同的训练数据集来拟合输入特征与输出目标之间的关系。

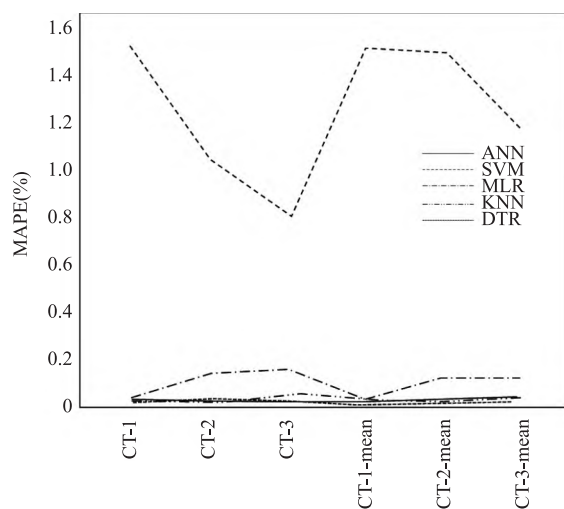
光伏预测指标如图4所示。图4表明,利用原始实测气象资料(CT-1、CT-2、CT-3),ANN和MLR的预报方法性能较优。相比之下,SVM在拟合光伏预测模型中表现不佳。由于光伏发电的固有特性,在夜间,太阳辐照度输入特征的数据值几乎有一半为零。这可能导致光伏功率回归拟合合适的SVM超平面较为困难。通过加入额外参数(CT-2、CT-3)的SVM模型可以减轻误差,相较于仅加入CT-1输入特征的SVM模型RMAE值更小。

此外,SVM的性能高度依赖于特征、样本和核函数。MAPE可以在与真实预测样本的比较中给出关于预测的性能优劣。光伏预测模型特别是SVM模型的MAPE值很大,这可能是因为式(8)中夜间时刻光伏功率($P_{true}^{PV}(t) \cong 0$)很小,稍有偏差就可能造成MAPE值偏大。一类太阳辐射数据作为输入参数便足以使kNN和DTR方法获得良好的预测结果。相比之下,输入参数的不同组合对预测结果影响不大。

在CT-1输入特征下,DTR和kNN方法的



(a) RMAE



(b) MAPE

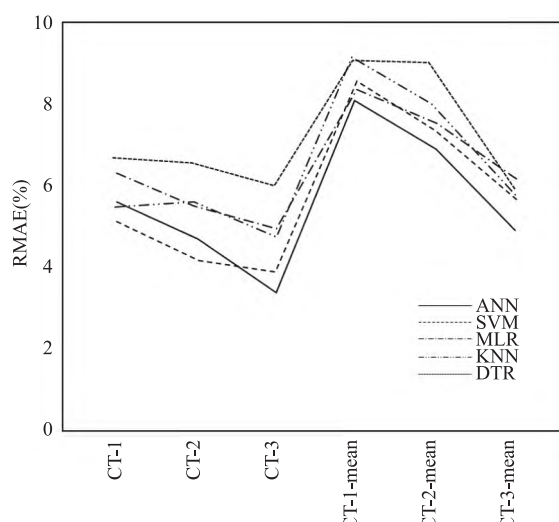
图 4 光伏预测指标

Fig. 4 Photovoltaic prediction indicators

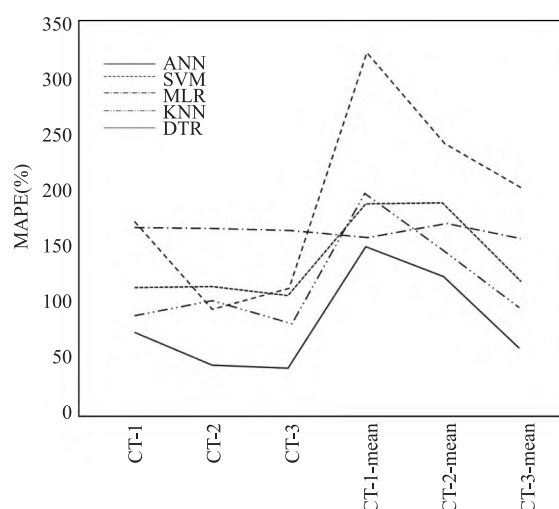
RMAE分别维持在2.7%和2.4%左右。额外的气象数据（气温和风速）未能带来显著的性能提升。相关系数较小的扩展特征对分类结果影响较小。同时，kNN中的 k 值和决策树结构的深度对预测结果影响较大。

图5对风力预测指标进行了比较。各类预测方法在风电功率预测中表现良好。

CT-1和CT-2输入的SVM均小于ANN模型。CT-3输入的ANN模型预测效果最好。考虑MAPE时，ANN模型较于支持向量机、高斯等不同的机器学习方法在风速预测中性能最优。



(a) RMAE



(b) MAPE

图 5 风力预测指标

Fig. 5 Wind prediction indicators

5 结束语

本文提出了一种基于ANN架构的新能源发电预测模型，将原始实测天气数据简化的平均天气数据作为模型输入特征进行风光发电功率预测，采用ANN架构从相关参数中提取信息。通过模拟实验，以平均绝对误差作为判据，光伏和风力发电的RMAE分别可以达到1.7%和4.9%，模型准确度较高；此外，ANN模型在全年不同季节均能提供稳定的风电功率预测准确度，在冬季可以获得较高的光伏功率估计准确度。

参考文献

- [1] 高贵刚. 新能源风光发电预测技术进展综述[J]. 电工技术, 2022(20): 49-52.
- [2] 李继峰, 张阿玲. 我国新能源和可再生能源发展预测方法研究——风能发电预测案例[J]. 可再生能源, 2004(3): 1-4, 28.
- [3] C Hailang, Z Haiqian. New energy generation forecasting and dispatching method based on big data[J]. Energy Reports, 2021(7): 1280-1288.
- [4] 金楠, 郭磊磊, 窦智峰. 清洁能源发电并网逆变器有限状态模型预测控制[J]. 电机与控制学报, 2018, 22(4): 89-97.
- [5] 卫鹏, 刘建坤, 周前. 基于随机潮流的新能源发电预测误差对电网影响研究[J]. 电器与能效管理技术, 2017(2): 60-65, 70.
- [6] 彭光虎, 马景超, 龚文杰. 基于ESN的光伏发电功率预测模型研究[J]. 青岛大学学报, 2015, 30(3): 12-15, 32.
- [7] LIU Kewen, WANG Jiang. A combination model for short term wind power forecasting[J]. Advanced Materials Research, 2013, 2649: 791-793.
- [8] 林琳, 邓国新, 樊浩. 基于天气模式识别与时空图神经网络的新能源发电功率预测[J]. 电气自动化, 2023, 45(3): 30-33.
- [9] CARPANETO E, CHICCO G, AKILIMALI J S. Characterization of the loss allocation techniques for radial systems with distributed generation[J]. Spain: IEEE Melecon, 2006: 1028-1031.
- [10] 张树华, 王继业, 王辰. 基于分层模糊神经网络的边缘侧光伏发电能量预测[J]. 现代电力, 2023: 1-10.
- [11] 张健. 基于卷积神经网络的新能源光伏电站发电功率预测方法[J]. 机械设计与制造工程, 2022, 51(10): 69-73.
- [12] 邹文进, 郝少飞, 马刚. 基于CEEMD-GA-BP神经网络的风光发电功率预测[J]. 电网与清洁能源, 2022, 38(03): 111-118.

作者简介

杨强（1977—），男，本科，高级技师，主要研究方向：电力调度运行和管理。

颜宗辉（1978—），男，工程师，主要研究方向：电力调度运行和管理。

（收稿日期：2023-12-12）