

doi: 10.7690/bgzdh.2022.11.008

# 基于相似时段和 PCA-ELM 的超短期风电功率预测

王磊<sup>1,2</sup>, 马磊娟<sup>2</sup>

(1. 西北大学经济管理学院, 西安 710000; 2. 河南工业职业技术学院基础教学部, 河南 南阳 473000)

**摘要:** 为提高风电场输出功率的预测精度, 提出一种采用相似时段选取原则和基于主成分分析 (principal component analysis, PCA) 与多层自编码极限学习机 (multi-layer auto encoder extreme learning machine, ML-AE-ELM) 组合算法 (PCA-ELM) 的预测模型。通过关联度分析明确待测时段的相似时段范围, 结合天气数据、机组状态和历史功率构建训练和测试样本, 利用预测算法完成样本的训练和测试, 得到输出功率预测结果并验证。实验结果表明: 与常见的算法模型相比, 该预测模型在不同装机容量和不同工作状态的风电场中均具有较高的预测精度, 表现出良好的预测稳定性和泛化能力。

**关键词:** 风电功率预测; 相似时段; 主成分分析; 多层自编码极限学习机

**中图分类号:** TP301.6 **文献标志码:** A

## Ultra-Short-Term Wind Power Forecasting Based on Similar Time Interval and PCA-ELM

Wang Lei<sup>1,2</sup>, Ma Leijuan<sup>2</sup>

(1. School of Economics and Management, Northwest University, Xi'an 710000, China;

2. Basic Teaching Department, He'nan Polytechnic Institute, Nanyang 473000, China)

**Abstract:** In order to improve the forecasting accuracy of wind farm output power, a method based on the principle of similar time period selection and the principle of principal component analysis (PCA) is proposed. The forecasting model is based on the combination of PCA and multi-layer auto encoder extreme learning machine (ML-AE-ELM). Through the correlation analysis, the range of similar periods of time to be tested is determined, and the training and test samples are constructed by combining weather data, unit status and historical power, and the forecasting algorithm is used to complete the training and test of the samples, so as to obtain the forecasting results of output power and verify them. The experimental results show that compared with the common algorithm model, the proposed model has higher forecasting accuracy in different installed capacity and different working conditions of wind farms, and shows good forecasting stability and generalization ability.

**Keywords:** wind power forecasting; similar time period; PCA; ML-AE-ELM

## 0 引言

截至 2020 年, 风力发电系统在全国范围内已实现 2.81 亿千瓦的累计装机容量, 风力发电受环境天气因素和发电机组状态的影响, 风电功率表现出一定的随机性和波动性。随着生活和生产对电能质量要求的日益提高, 电力系统的调度和管理正面临着大规模风电并网所带来的巨大挑战, 而风电功率预测是电力系统调度和管理的前提, 准确预测风电功率有助于调度部门制定运行和控制方式, 确保电力系统运行的稳定性<sup>[1]</sup>。

风电功率的预测需要完成 4 个主要步骤: 1) 确定输出预测范围, 根据《风电功率预测功能规范》相关要求, 笔者设定预测时长为 4 h, 时间间隔为 15 min; 2) 确定待测时段的相似时段范围, 文献中

出现的样本选取范围主要是待测时段的相邻时段<sup>[2]</sup>或者前几日同时段<sup>[3]</sup>, 而忽略了季节规律对风电场的影响; 3) 构建训练和测试样本, 大多数文献中对样本参数的确定局限于依靠数值天气预报 (numerical weather prediction, NWP) 数据和风电场历史功率数据, 没有考虑风电机组信息, 造成预测模型输入信息的缺失<sup>[4]</sup>; 4) 建立预测算法模型, 支持向量机 (SVM) 及其优化算法的泛化能力强, 但是容易陷入局部最优, 且工作量较大<sup>[5]</sup>。优化的反向传播神经 (BP) 网络结构简单, 但对样本相似性要求较高, 泛化能力差<sup>[6]</sup>。采用极限学习机 (extreme learning machine, ELM) 算法可得到较高的预测精度, 但网络结构需要进一步优化<sup>[7]</sup>。

针对上述问题, 笔者首先提出按照待测时段的

收稿日期: 2022-07-19; 修回日期: 2022-08-23

基金项目: 河南省 2020 年科技发展计划 (202102210134); 河南省高等学校青年骨干教师培养计划 (2019GZGG098、2018GGJS229)

作者简介: 王磊 (1982—), 男, 河南人, 博士, 副教授, 从事应用数学、新能源技术及仿真研究。E-mail: 27452154@qq.com。

相邻时段和前一年同日期临近时段的相似时段选择方法，然后在训练和测试样本中综合考虑 NWP 数据、机组状态和历史功率，最后提出一种基于主成分分析(PCA)与 ELM 组合算法的风电功率预测算法，并通过与常见的预测模型进行对比，验证笔者所提预测模型的有效性，提高风电功率预测的精度，降低电能波动对电力系统的冲击。

## 1 影响风力发电的主要因素

单台风力发电机的输出功率可表示为：

$$P = \begin{cases} 0 & (v \leq v_{in}) \\ \pi \rho C_p R^2 v^3 / 2 & (v_{in} \leq v \leq v_E) \\ P_E & (v_E \leq v \leq v_{out}) \\ 0 & (v \geq v_{out}) \end{cases} \quad (1)$$

式中： $P$  为风力发电机的输出功率，W； $\rho$  为空气密度， $\text{kg}/\text{m}^3$ ； $C_p$  为风能利用系数； $R$  为叶片尺寸，m； $v$  为实际风速，m/s； $P_E$  为风力发电机额定输出功率，W； $v_{in}$  为风力发电机的切入风速，m/s； $v_E$  为风力发电机的额定风速，m/s； $v_{out}$  为风力发电机的切出风速，m/s。

由上式可知：1) 结合风电机组功率曲线， $v$  是影响  $P$  值大小的关键因素，尤其当  $v$  值大小在  $v_{in}$  和  $v_E$  之间时，较小的风速变化会带来较大的输出功率变化。2)  $\rho$  主要受温度  $T(^{\circ}\text{C})$ 、湿度  $H(\%)$  和气压  $P(\text{Pa})$  的影响。3)  $C_p$  与叶尖速比  $\varphi$  和桨距角  $\beta$  有关，当  $v \leq v_E$  时， $\beta=0$ ，通过变速装置改变转子转速，使  $\varphi$  固定在最佳叶尖速比，从而  $C_p$  恒定在最佳值；当  $v \geq v_E$  时，调节  $\beta$  从而减少发电机输出功率，使  $P$  稳定在  $P_E$  值。4)  $R$  对于单台风力发电机为定值，通常情况下风力发电机的额定功率越大，配备的叶片尺寸也越大。

风力发电场通常由多台同型号的风力发电机按照一定规则排成机群供电。受尾流效应的影响，处于同一风场中的多台发电机，下风向电机的输出功率弱于上风向电机，风向也是影响  $P$  值大小的重要因素。

笔者选取风速、风向、温度、湿度、气压、叶尖速比和桨距角作为影响风电功率的主要因素，并选择采用不同型号风力发电机的 3 个风电场作为研究对象。

## 2 预测模型

依据 NWP 提供的天气数据选取相似时段，在此基础上结合天气数据、机组状态和历史功率构成

训练和测试样本，通过主成分分析与多层自编码极限学习机组合算法(PCA-ELM)进行预测模型的训练和测试，得到风电功率预测结果。

### 2.1 相似时段的选取

相似时段是指从历史时段样本中选取的与待测时段的各参数数据最接近的样本，根据风电功率的主要影响因素和 NWP 数据的特点，选取风速—最大值、最小值和平均值，风向—正弦值和余弦值，温度—平均值，湿度—平均值，气压—平均值作为相似时段的判定因素。

1) 选定历史时段样本范围。

影响风电功率的各因素在预测时长范围内既表现出连续性，又伴随有突变现象。笔者对历史时段样本的选取范围为：假设待测时段为某日的  $t$  时~ $t+4$  时，则选择  $t-4$  时~ $t$  时、 $t-5$  时~ $t-1$  时、 $t-6$  时~ $t-2$  时、……、 $t-38$  时~ $t-34$  时、 $t-39$  时~ $t-35$  时，以及前一年与待测时段同日期的当日和前后各 30 日的每天  $t$  时~ $t+4$  时，作为历史时段样本，共计 133 个时段。

2) 定义特征向量。

由相似时段的判定因素构建特征向量，即：

$$\mathbf{x} = [v_{\max} \ v_{\min} \ v_{\text{ave}} \ \sin\theta \ \cos\theta \ T_{\text{ave}} \ H_{\text{ave}} \ p_{\text{ave}}]^T \quad (2)$$

式中： $\mathbf{x}$  为判定特征向量； $v_{\max}$ 、 $v_{\min}$  和  $v_{\text{ave}}$  分别为历史时段内风速的最大值、最小值和平均值，m/s； $\sin\theta$  和  $\cos\theta$  分别为历史时段内风向的正弦值和余弦值； $T_{\text{ave}}$  为历史时段的平均温度， $^{\circ}\text{C}$ ； $H_{\text{ave}}$  为历史时段的平均湿度，%； $p_{\text{ave}}$  为历史时段的平均气压，Pa。

则待测时段和第  $i$  个历史时段的判定特征向量分别为：

$$\mathbf{x}_0 = [x_0(1) \cdots x_0(j) \cdots x_0(8)]^T \quad (3)$$

$$\mathbf{x}_i = [x_i(1) \cdots x_i(j) \cdots x_i(8)]^T \quad (4)$$

式(3)中： $\mathbf{x}_0$  为待测时段的判定特征向量； $x_0(j)$  为待测时段的第  $j$  个判定因素( $j=1, 2, \dots, 8$ )。式(4)中： $\mathbf{x}_i$  为第  $i$  个历史时段的判定特征向量( $i=1, 2, \dots, 133$ )； $x_i(j)$  为第  $i$  个历史时段的第  $j$  个判定因素( $i=1, 2, \dots, 133$ ； $j=1, 2, \dots, 8$ )。

3) 采用极差法对  $\mathbf{x}_0$  和  $\mathbf{x}_i$  进行归量纲的一化处理，降低计算难度和误差。

4) 计算各历史时段与待测时段的关联系数及关联度。

5) 确定相似时段范围。

根据关联度计算结果，对 133 个历史时段按照

降序排序:  $a_0, a_1, \dots, a_{132}$ , 选取  $a_1, a_2, \dots, a_{60}$  为相似时段, 定义  $a_0$  为最佳相似时段。

### 2.2 训练样本和测试样本

如表 1 所示, 笔者选择 60 个相似时段作为训练样本, 每个样本包括该时段的天气参数、机组参数和历史功率作为训练样本。

表 1 训练样本参数

分类	参数内容
$a_i$ 输入参数: $x_1 \sim x_{14}$	风速最大值、风速最小值、风速平均值、风向正弦值、风向余弦值、平均温度、平均湿度、平均气压、叶尖速比最大值、叶尖速比最小值、叶尖速比平均值、桨距角最大值、桨距角最小值、桨距角平均值
$a_i$ 输出参数: $y_1 \sim y_{16}$	$t$ 时 $\sim t+4$ 时的输出功率值, 间隔 15 min (不包括 $t+4$ 时刻)

采用上述训练样本参数, 可以有效地参考相似时段的天气特征和风电场的机组状态, 通过训练不断校正模型参数, 从而得到更适于测试的预测模型。

笔者选择待测时段的天气参数和最佳相似时段的机组参数作为测试样本的输入参数, 待测时段的输出功率值作为输出参数, 如表 2 所示。

表 2 测试样本参数

分类	参数内容
待测时段天气参数: $x_1 \sim x_8$	风速最大值、风速最小值、风速平均值、风向正弦值、风向余弦值、平均温度、平均湿度、平均气压
$a_0$ 机组参数: $x_9 \sim x_{14}$	叶尖速比最大值、叶尖速比最小值、叶尖速比平均值、桨距角最大值、桨距角最小值、桨距角平均值
待测时段输出 功率: $y_1 \sim y_{16}$	$t$ 时 $\sim t+4$ 时的输出功率值, 间隔 15 min (不包括 $t+4$ 时刻)

采用待测时段和  $a_0$  的数据作为输入参数, 可以找到与待测时段各参数值最接近的历史时段, 同时避免由于天气突变造成的预测误差, 提高预测精度。

### 2.3 预测算法

预测模型样本的输入参数包括天气参数和机组参数共计 14 个, 且存在一定的相关性; 训练样本利用的相似时段共计 60 个。输入参数较多且信息冗余, 会对运算速率和精度造成影响; 为了充分利用训练样本完成预测算法的学习, 提高预测算法的精度和泛化能力, 需要对算法的网络结构进行优化。

#### 2.3.1 主成分分析

样本的各输入参数之间存在一定的相关性, 例如, 温度和气压之间存在相互影响、风速和温度都会引起空气密度变化等现象, 采用 PCA 可以降低输入参数之间的信息冗余度, 在保留原始输入参数绝大部分信息的同时减少输入参数的维度, 降低问题的复杂性<sup>[8]</sup>。

1) 假设样本数量为  $n$ ,  $x_1, x_2, \dots, x_m$  为每个样本的指标变量 ( $m$  个), 则全部样本构成的指标矩阵为:

$$\mathbf{x} = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1n} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & & \vdots & & \vdots \\ x_{m1} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix} \quad (5)$$

式中:  $x_{ij}$  为第  $j$  个样本的第  $i$  个指标变量 ( $i=1, 2, \dots, m; j=1, 2, \dots, n$ )。

2) 计算原始样本指标矩阵  $\mathbf{x}$  的标准化矩阵  $\mathbf{x}^*$ 。

3) 根据标准化矩阵  $\mathbf{x}^*$  计算相关系数矩阵  $\mathbf{R}$ 。

4) 计算  $\mathbf{R}$  的特征值并排序:  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ , 以及  $\lambda_i$  对应的特征向量  $\mathbf{u}_i$  ( $i=1, 2, \dots, m$ )。

5) 计算每个主成分的贡献率, 以及前  $t$  个主成分的累计贡献率。

当累计贡献率超过一定数值 (本文中取 0.9) 后, 则认为可以用这  $t$  个主成分表示原始输入参数, 即:

$$\left. \begin{aligned} y_1 &= u_{11}x_1 + u_{12}x_2 + \cdots + u_{1m}x_m \\ y_2 &= u_{21}x_1 + u_{22}x_2 + \cdots + u_{2m}x_m \\ &\quad \dots \\ y_t &= u_{t1}x_1 + u_{t2}x_2 + \cdots + u_{tm}x_m \end{aligned} \right\} \quad (6)$$

式中:  $y_1, y_2, \dots, y_t$  为新的指标变量;  $x_1, x_2, \dots, x_m$  为原始指标变量, 通常情况下  $t < m$ 。

通过 PCA 降维, 可以保留输入参数的绝大部分信息, 减少输入参数数量, 提高预测模型运算效率。

#### 2.3.2 多层自编码极限学习机

ELM 由输入层、隐层和输出层组成, 各层之间采用随机生成的权值矩阵和偏置向量, 网络结构中容易引入非最优的随机模型参数, 导致部分网络节点无效。笔者采用自编码极限学习机 (auto encoder ELM, AE-ELM), 避免了随机参数的引入, 克服了 ELM 节点无效的问题<sup>[9]</sup>。

假设样本空间为  $\{[\mathbf{x}_j \quad \mathbf{y}_j]^T\}$ , 样本数量为  $N$ , 即样本输入矩阵和输出矩阵分别为  $\mathbf{X}=[\mathbf{x}_1 \cdots \mathbf{x}_j \cdots \mathbf{x}_N]$  和  $\mathbf{Y}=[\mathbf{y}_1 \cdots \mathbf{y}_j \cdots \mathbf{y}_N]$ , 其中  $\mathbf{x}_j$  和  $\mathbf{y}_j$  分别为第  $j$  组样本的输入向量和输出向量。

首先, 令  $\mathbf{X}=\mathbf{Y}=[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ , 如图 1 所示, 在输入层与隐层之间引入正交权值矩阵  $\mathbf{W}$  和偏置向量  $\mathbf{b}$ , 其中  $\mathbf{W}=[\mathbf{w}_1 \cdots \mathbf{w}_s \cdots \mathbf{w}_l]^T$ ,  $\mathbf{w}_s$  表示所有输入层节点与隐层第  $s$  个节点之间的正交随机权值向量,  $\mathbf{b}=[b_1, b_2, \dots, b_l]$ 。

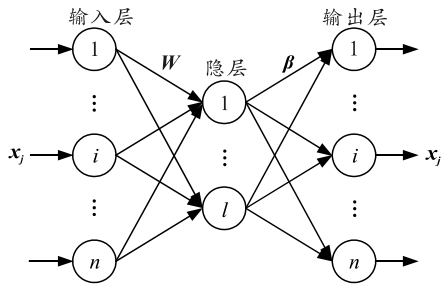


图 1 AE-ELM 算法

可得：

$$H = \begin{bmatrix} f(w_1 \cdot x_1 + b_1) & \cdots & f(w_1 \cdot x_N + b_1) \\ \vdots & & \vdots \\ f(w_l \cdot x_1 + b_l) & \cdots & f(w_l \cdot x_N + b_l) \end{bmatrix} \quad (7)$$

式中： $H$  为隐层的输出矩阵； $f(x)$  为激活函数。

其次，通过  $X = \beta H$  将  $H$  重构为  $X$ ， $\beta$  可由下式求出：

$$\beta = X(I/C + H^T H)^{-1} H^T \quad (8)$$

式中： $I$  为单位矩阵； $C$  为正则项系数。

利用最小二乘法，并用  $Y$  代替式(8)中的  $X$  得到  $\beta^*$ ，即：

$$\beta^* = Y(I/C + H^T H)^{-1} H^T \quad (9)$$

最后，将  $\beta^T$  ( $\beta$  的转秩矩阵) 作为网络结构中输入层与隐层的权值矩阵， $\beta^*$  作为网络结构中隐层与输出层的权值矩阵，如图 2 所示。

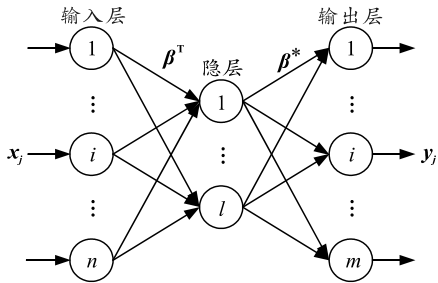


图 2 AE-ELM 结构

如图 3 所示，多层自编码极限学习机 ML-AE-ELM 以 AE-ELM 算法和结构为基础，网络结构包含多层隐层。

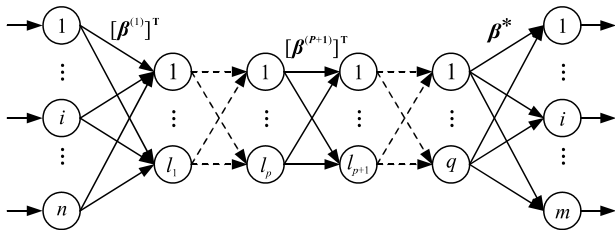


图 3 ML-AE-ELM 网络结构

### 2.4 预测步骤

笔者提出的基于相似时段原则，以及 PCA-ELM

的风电功率预测模型的工作流程如图 4 所示。

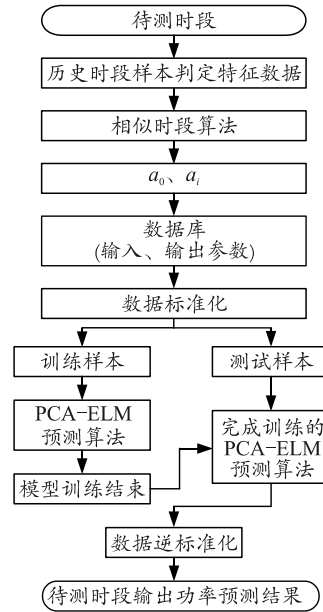


图 4 预测模型流程

## 3 实验结果与分析

笔者采用 Matlab 编写风电功率预测模型的各项算法步骤并进行仿真验证，分别对装机总容量为 30、50、70 MW 的 3 个风电场进行功率预测。

### 3.1 评价指标

为了评估预测效果，笔者采用平均绝对百分比误差 (mean absolute percentage error, MAPE) 表征算法模型的预测能力，采用均方根误差 (root mean square error, RMSE) 表征算法模型的误差，即：

$$MAPE = \frac{1}{N} \sum_{i=1}^N |(y_{ai} - y_{fi}) / y_{ai}| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{ai} - y_{fi})^2} \quad (11)$$

式中： $N$  为待测时段采样点个数； $y_{ai}$  和  $y_{fi}$  分别为  $i$  时刻风电功率的实际值和预测值。

### 3.2 网络结构设置

笔者通过实验确定 ML-AE-ELM 的最优网络结构。首先设置隐层层数为 1，隐层节点数从 50 开始，并以 50 为变化步长，得到每个节点数对应的 MAPE 值和 RMSE 值，当评价指标达到极小值点时，对应的节点数为第 1 层的最优节点数，然后以此为基础增加一层新的隐层并采用相同的方法确定该层的最佳节点数，以此类推可以得到最佳的网络结构参数。实验中，评价指标数值取多次实验结果的平均值。

实验结果发现，当隐层层数为 5 时，误差值最小，ML-AE-ELM 最优网络结构为“输入层—隐层(400—1250—700—750—750)—输出层”。

### 3.3 算例分析

笔者以装机容量分别为 30、50、70 MW 的 3 个风电场为例，验证所提 PCA-ELM 预测模型的预测效果，并与 BP 算法、SVM 算法进行对比。

待测时段的实际风电功率与 3 种算法模型的预测结果如图 5 所示。可以发现，笔者预测模型对 3 个风电场的功率预测值与实际值表现出较高的拟合度。

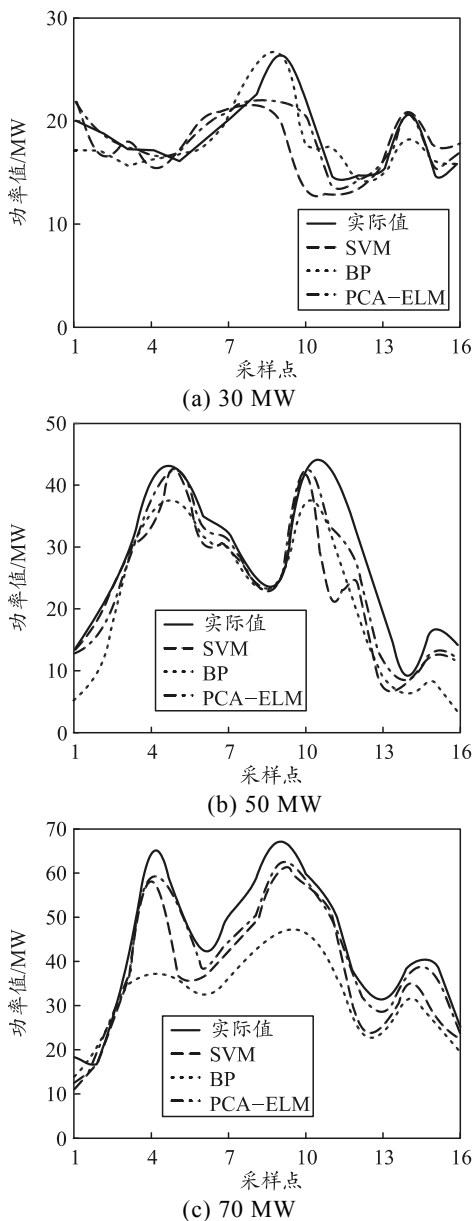


图 5 不同风电场的功率预测结果

BP 算法、SVM 算法与文中 PCA-ELM 算法模型在不同风电场的 MAPE 值和 RMSE 值如表 3

所示。

表 3 误差指标值

指标	风电场/MW	SVM	BP	PCA-ELM
MAPE/%	30	11.81	8.14	5.16
	50	13.95	25.71	8.56
	70	15.97	26.39	7.54
RMSE/MW	30	3.23	1.93	1.52
	50	6.49	6.75	3.21
	70	7.73	13.57	3.77

由上表可见，PCA-ELM 算法模型的预测误差值明显低于 SVM 算法和 BP 算法，并且在不同装机容量的风电场中均表现出良好的应用效果。

综合图 5 和表 3 可进一步发现：1) 30 MW 风电场在待测时段内的实际功率波动较小，3 种算法都表现出较高的预测精度，BP 算法与 PCA-ELM 算法的 RMSE 值比较接近，都小于 2 MW，但是 PCA-ELM 算法的 MAPE 值更低，为 5.16%，表现出更高的预测精度；2) 3 种算法对 50 和 70 MW 风电场的功率预测精度有所下降，是由于在待测时段内实际功率有较大的波动，BP 算法对变化明显的风电场表现出较差的适应性，PCA-ELM 算法受到环境波动的影响仍表现出较高的预测精度，MAPE 值在 8% 左右，RMSE 值不超过 4 MW，均优于 SVM 算法和 BP 算法。

由此可得，PCA-ELM 算法对不同装机容量的风电功率预测精度更高，对环境变化的适应性更强，预测效果更稳定且具有较高的泛化能力。

## 4 结论

笔者所提出的预测模型与 BP 算法和 SVM 算法相比，对不同装机容量以及不同工作状态的风电场均表现出更好的预测效果。

- 1) 相似时段的选取综合考虑了天气因素的连续性和突变性，以及季节变化对天气的影响，使样本选取范围更加合理；
- 2) 在预测和训练样本参数中加入机组状态信息，使预测模型输入信息更加完整；
- 3) 采用 PCA-ELM 算法，保留输入信息完整度的同时提高了运算速率，经过网络优化的 ELM 具有预测精度高、稳定性高和泛化能力强的特点。

## 参考文献：

[1] 宿凤明, 孙财新, 李端开. 基于神经网络的风电短期功率预测模型研究[J]. 节能技术, 2018, 36(6): 556-560.

[2] 杨正钊, 田勇, 张广涛, 等. 相似日短期负荷预测的非线性理论基础与改进[J]. 电网技术, 2006, 30(6): 63-66.