

# 基于 OS-ELM 的光伏发电中长期功率预测

钱子伟<sup>1,2</sup>, 孙毅超<sup>1,2</sup>, 王 琦<sup>1,2</sup>, 季顺祥<sup>1,2</sup>, 周 敏<sup>1,2</sup>, 曾柏琛<sup>1,2</sup>

(1.南京师范大学南瑞电气与自动化学院,江苏 南京 210023)

(2.南京师范大学江苏省电气互联综合能源工程实验室,江苏 南京 210023)

**[摘要]** 为了进一步提高光伏出力预测的精度,提出了一种基于在线序列极限学习机的光伏发电中长期功率预测方法. 结合在线序列极限学习机学习速度快、泛化能力强的特点,通过对大量气象数据和历史发电数据综合处理,对光伏发电系统的输出功率进行预测. 同时,由于实时数据的不断输入,该方法能够对预测模型进行在线更新. 算例仿真研究表明,该预测方法与反向传播神经网络、支持向量机方法相比,能够有效提高预测精度,满足在线应用的需求,具有较好的应用前景.

**[关键词]** 光伏预测,相关性分析,在线序列极限学习机,数据更新

**[中图分类号]** TM615 **[文献标志码]** A **[文章编号]** 1672-1292(2020)01-0008-07

## Medium and Long Term Photovoltaic Power Generation Forecasting Based on OS-ELM

Qian Ziwei<sup>1,2</sup>, Sun Yichao<sup>1,2</sup>, Wang Qi<sup>1,2</sup>, Ji Shunxiang<sup>1,2</sup>, Zhou Min<sup>1,2</sup>, Zeng Baichen<sup>1,2</sup>

(1.School of NARI Electrical and Automation, Nanjing Normal University, Nanjing 210023, China)

(2.Jiangsu Key Laboratory of Gas and Electricity Interconnection Integrated Energy, Nanjing Normal University, Nanjing 210023, China)

**Abstract:** In order to further improve the accuracy of PV output prediction, a medium and long term power prediction method based on online sequential extreme learning machine (OS-ELM) is proposed. Combined with the characteristics of fast learning and generalization ability of OS-ELM, the output power of photovoltaic power generation system is predicted by comprehensively processing a large number of meteorological data and historical power generation data. At the same time, due to the continuous input of real-time data, the method can update the prediction model online. The simulation study shows that compared with the back propagation (BP) neural network and support vector machine (SVM) method, the prediction method can effectively improve the prediction accuracy and meet the needs of online applications, and it has a good application prospect.

**Key words:** photovoltaic forecasting, correlation analysis, online sequential extreme learning machine (OS-ELM), data update

化石能源的储量不容乐观,近年来,人类正在面临着资源匮乏的危机. 太阳能作为一种无污染的可再生能源,逐渐引起社会的关注,光伏发电因其价格低廉,安装方便等优点也得到了快速的发展. 然而光伏发电也存在着随机性、间歇性等问题,大规模的光伏并网给电力调度部门安排电源、协调光伏发电带来了困难<sup>[1]</sup>. 因此,精准地预测光伏发电的功率,能够有效地帮助电力调度部门合理安排电网运行模式,提高对光伏能源的利用效率,也能够进一步提升电能质量,维持电网安全稳定运行<sup>[2]</sup>.

按照预测时间的长短不同,可将光伏发电功率预测分为中长期预测、短期预测和超短期预测 3 类. 文献[3]中将确定性因素和不确定因素分离建模,分析了光伏发电的中长期随机特性,未给出明确的光伏预测模型. 文献[4]中将前馈方向传播和广义回归神经网络相结合,用来预测一年的光伏发电功率,但考虑的时间尺度过长,缺乏实际应用的价值. 文献[5]通过模糊理论和神经网络相结合,实现对某一个月的光伏发电力进行预测,但忽略了时间的连续特性,应考虑预测周期内每日气象数据间的相关性. 精准的中长期

收稿日期:2019-03-18.

基金项目:江苏省研究生科研与实践创新计划项目(SJCX19\_0386).

通讯作者:孙毅超,博士,讲师,研究方向:电力电子技术及其在电力系统中的应用. E-mail:yichao.sun1987@gmail.com

预测能够给电网规划提供依据,为电网安全稳定的运行提供参考,及时发现电网运行中存在的薄弱环节,以便电力部门能够提前做好预防工作。

按照预测的原理不同可分为物理法、统计法和人工智能法。物理法主要是通过搭建物理模型进行预测,文献[6]主要通过分析光伏电池和逆变器的等效模型,结合光伏电站的实际位置,来实现光伏发电功率预测。物理法仅需要知道光伏电站详细的物理信息就能够对光伏发电功率进行预测,但是预测的精度受所建立模型的精度影响。与物理法不同的是,统计法则需要大量光伏电站运营时的数据,通过对数据的提取,寻找输入变量与被预测值之间的联系,从而对光伏发电功率进行预测。常用的方法有聚类分析、相似时段、马尔可夫链等。文献[7]提出了将近邻传播聚类 and 回声状态网络相结合的方法,构建回声网络状态方程来实现光伏发电功率预测。文献[8]通过结合分时段神经网络模型和相似时段筛选法,将分时段神经网络模型作为相似时段筛选法在相似度不够时的补充,从而进行光伏发电功率的预测。文献[9]中提出了一种基于回归分析和马尔科夫链的预测模型来进行光伏发电功率的预测。文献[10]先将训练样本按照隶属度的指标来选取相似日,再基于交叉熵理论提出了光伏发电功率的组合预测模型从而进行预测。统计法需要大量历史数据,在实际运用过程中难以满足数据的需求。人工智能法则主要是计算机的机器学习,通过计算机对历史数据进行处理分析,从而得到各个输入变量与预测结果之间的联系,最终实现光伏发电功率的预测。主要包括:人工神经网络、支持向量机等。文献[11]提出一种适用于小样本的双层神经网络单步光伏预测的方法,在有效减少对训练样本数量需求的同时保证了光伏出力预测的准确度。文献[12]将一种集合经验模态分解的方法与常见的支持向量机方法相结合,先按照天气类型进行分类,再进行光伏发电处理预测。人工智能法预测精度较高,但是需要大量时间来进行计算机的机器学习,当预测模型需要更新时,训练时间和成本较高。因此,为提高光伏发电功率预测精度,充分考虑实时更新的样本数据,本文提出一种基于在线序列极限学习机(online sequential extreme, OS-ELM)的光伏发电中长期功率预测模型。首先,采用相关性分析的方法对光伏发电功率的影响因子进行分析,选取具有代表性的影响因素作为模型的训练样本,然后构建 OS-ELM 光伏发电预测模型。通过算例分析,验证模型在新数据不断输入的情况下,精度能够不断提升。同时,将本文所提方法与反向传播神经网络和支持向量机预测模型进行对比分析,验证了本文所提 OS-ELM 光伏发电预测模型的正确性和优越性。

# 1 光伏发电功率的影响因素分析

对光伏发电功率造成影响的因素有很多,在工程上通常将光伏发电功率按照如下公式进行计算<sup>[13]</sup>:

$$P=\eta SI[1-0.005(T+25)]. \tag{1}$$

式中, $\eta$  为光电的转换效率; $S$  为光伏面板的总面积; $I$  为日照强度; $T$  为环境温度。在进行光伏预测时通常将转换效率和光伏面积当作常数,光伏发电功率主要与日照强度和环境温度相关。然而,式(1)只是工程上计算光伏输出功率的简化公式,在预测时还需要考虑其他多种因素,如风速、湿度、气压等。

本文利用 Pearson 系数法<sup>[14]</sup>分析各个气象因子和光伏功率之间的关系,分析结果如表 1 所示。

表 1 Pearson 相关系数分析结果  
Table 1 Results of Pearson correlation coefficient analysis

变量	相关系数					
	$P$	$I$	$T$	$H$	$p_0$	$v$
$P$	1	0.874	0.640	-0.571	-0.052	-0.171
$I$	0.874	1	0.660	-0.636	-0.175	-0.127
$T$	0.640	0.660	1	-0.289	-0.503	-0.036
$H$	-0.571	-0.636	-0.289	1	-0.101	0.093
$p_0$	-0.052	-0.175	-0.503	-0.101	1	-0.139
$v$	-0.171	-0.127	-0.036	0.093	-0.139	1

从表 1 中可知,与光伏发电功率  $P$  相关系数较大的三个影响因子分别是日照强度  $I$ ,环境温度  $T$  和湿度  $H$ ,而大气压强  $p_0$  和风速  $v$  与功率相关系数较小,因此选择日照强度、温度和湿度作为光伏发电功率的

影响因子. 考虑到气象观测站无法提供日照强度数据,故采用日照时长来代替.

## 2 光伏发电预测模型

### 2.1 ELM 算法

ELM 算法是一种快速学习的人工智能算法,ELM 的结构与单隐层神经网络相类似,主要包括输入层、隐含层和输出层.

如图 1 所示,对于某一标准极限学习机网络模型,假设有  $N$  组任意的训练样本  $(X_i, Y_i)$ , 其中  $X_i = [x_1, x_2, \dots, x_n]^T$  为训练样本中的输入,  $Y_i = [y_1, y_2, \dots, y_m]^T$  为训练样本中输入所对应的输出. 隐含层的输出函数有以下定义:

$$o_j = \sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i). \quad (2)$$

式中,  $j=1, \dots, N$ ;  $\beta_i$  为输出权重;  $g(x)$  为激励函数;  $\omega_i = [\omega_{i,1}, \omega_{i,2}, \dots, \omega_{i,n}]^T$  为输入权重;  $b_i$  是第  $i$  个隐含层单元的阈值.

ELM 算法的目标是使得模型的输出与实际理论上所对应的输出之间的差最小,即

$$\sum_{i=1}^N \|o_i - y_i\| = 0. \quad (3)$$

即存在  $\beta_i, \omega_i$  和  $b_i$ , 使得

$$\sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i) = y_j, j=1, 2, \dots, N. \quad (4)$$

也可以通过矩阵的形式来表示,即

$$H\beta = T. \quad (5)$$

式中,  $H$  是隐含层节点的输出,  $\beta$  是输出权重,  $T$  是期望输出, 具体形式为

$$H(\omega_1, K, \omega_L, b_1, K, b_L, X_1, K, X_N) = \begin{bmatrix} g(\omega_1 \cdot X_1 + b_1) & L & g(\omega_L \cdot X_1 + b_L) \\ M & 0 & M \\ g(\omega_1 \cdot X_N + b_1) & L & g(\omega_L \cdot X_N + b_L) \end{bmatrix}_{N \times L}. \quad (6)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ M \\ \beta_L^T \end{bmatrix}_{L \times m}, T = \begin{bmatrix} T_1^T \\ M \\ T_L^T \end{bmatrix}_{L \times m}. \quad (7)$$

在 ELM 算法中,一开始就随机给定了输入权重  $\omega_i$  和阈值  $b_i$ . 并且隐含层与输出层之间的输出权值不需要重复进行迭代,可以通过求解方程组的解来一次性确定,大大减少了算法运算的时间. 并且文献[15]已经证明全局最优输出权值可以写为:

$$\hat{\beta} = H^* T. \quad (8)$$

式中,  $H^* = (H^T H)^{-1} H^T$  为隐含层输出矩阵  $H$  的 Moore-Penrose 广义逆.

### 2.2 OS-ELM 模型

在各种传统模型预测过程中,输入样本数据的质量会直接影响预测模型的预测精度. 在实际生产过程中,各类气象数据在采集、存储的过程中会发生错误而产生异常样本. 同时光伏发电的历史数据数量庞大,难以将异常样本全部剔除. 故这些偏离实际情况的异常样本将会对预测结果造成影响. OS-ELM 在新数据输入过程中,可将输入样本数控制在合理的范围内,对训练样本进行有效的筛选,确保不存在异常样本. 随着新数据的不断输入,历史数据中异常样本对预测结果所造成的影响也将不断削弱,从而提高了预测模型的鲁棒性. OS-ELM 在具体实现过程中主要分为两个阶段.

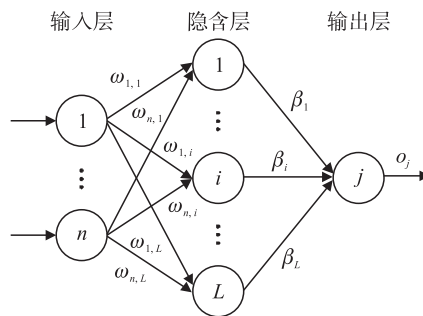


图 1 ELM 结构

Fig. 1 The structure of ELM

### (1) 初始化阶段

通过给定网络的初始训练样本、隐含层神经元个数、激励函数,随机生成输入权重和阈值,从而确定初始网络中隐含层输出权重  $\beta_0$  以及隐含层输出矩阵  $H_0$ 。

### (2) 在线顺序学习阶段

当新的一批数据到来时,隐含层输出矩阵和输出权重向量可根据下面表达式更新<sup>[16]</sup>:

$$H_{t+1} = \begin{bmatrix} g(\omega_1^T \cdot X_1^{(t+1)} + b_1) & L & g(\omega_L^T \cdot X_1^{(t+1)} + b_L) \\ M & 0 & M \\ g(\omega_1^T \cdot X_{N_{t+1}}^{(t+1)} + b_1) & L & g(\omega_L^T \cdot X_{N_{t+1}}^{(t+1)} + b_L) \end{bmatrix}_{N \times L} \quad (9)$$

$$\beta_{t+1} = \beta_t + Q_{t+1} H_{t+1} (R_{t+1}^T - H_{t+1}^T \beta_t). \quad (10)$$

式中,

$$Q_{t+1} = Q_t - \frac{Q_t H_{t+1} H_{t+1}^T Q_t}{I + H_{t+1}^T Q_t H_{t+1}}, \quad (11)$$

$$Q_0 = (H_0^T H_0)^{-1}, \quad (12)$$

$N_{t+1}$  表示第  $t+1$  批次样本的个数,  $X_{N_{t+1}}^{(t+1)} = [x_1^{(t+1)}, x_2^{(t+1)}, \dots, x_{N_{t+1}}^{(t+1)}]$  表示第  $t+1$  批次样本的输入向量。

通过上述公式结合新加入的数据不断更新参数  $H$  和  $\beta$ ,直到所有数据都输入后,完成 OS-ELM 模型的训练。

## 3 光伏发电功率预测模型

### 3.1 数据的预处理

由于数据之间存在较大的差异,为了减小因量纲不同而带来的误差,需要对数据进行归一化处理,其公式为:

$$\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}. \quad (13)$$

式中,  $\hat{x}_i$  为进行归一化后得到的数据,  $x_i$  为原始数据,  $x_{\max}$  和  $x_{\min}$  分别为原始数据的最大值和最小值。

### 3.2 基于 OS-ELM 的预测模型

基于 OS-ELM 预测流程图如图 2 所示。首先选择一批初始样本,包括气象影响因子的数据和历史光伏发电数据,通过标准的 ELM 算法得到一个初始的预测模型,生成  $\beta_0$  和  $H_0$  两个参数。之后不断导入新一批的数据包括气象数据和光伏发电功率数据,训练网络,更新  $\beta$  和  $H$  两个参数,更新预测模型,直到所有数据全部导入,得到最终的光伏功率预测模型。

### 3.3 评价指标的选取

为了更加精准地评价模型的准确性和有效性,本文选用均方根误差 (Root Mean Squared Error, RMSE) 和平均绝对百分比误差 (Mean Absolute Percent Error, MAPE),其中  $e_{\text{RMSE}}$  和  $e_{\text{MAPE}}$  的计算公式分别为:

$$e_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\hat{i}} - y_i)^2}. \quad (14)$$

$$e_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_{\hat{i}}}{y_i} \right| \times 100\%. \quad (15)$$

式中,  $n$  为测试样本数,  $y_{\hat{i}}$  为预测光伏输出功率,  $y_i$  为实际光伏输出功率。

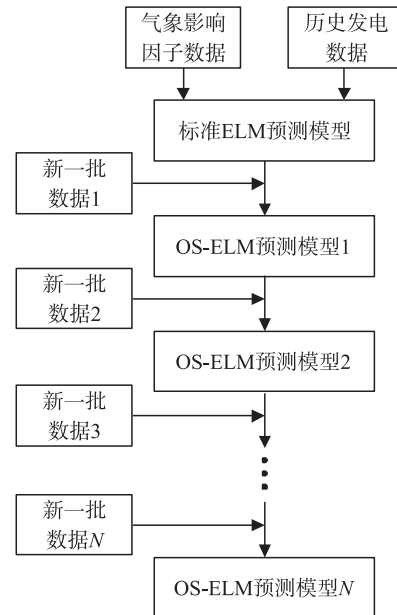


图 2 光伏功率预测流程

Fig. 2 Photovoltaic power forecasting process



4 算例分析

4.1 实验方案设计

为了验证本文所提的基于 OS-ELM 的光伏发电功率预测模型的准确性和有效性,以江西南昌某光伏发电系统为例(装机容量为 45 kW),选取 2013 年 1 月至 12 月的运行数据作为训练样本,气象数据来自于当地气象观测站,各种数据的时间分辨率间隔为 1 h. 由于本文所分析的是光伏发电中长期功率预测,故将每天每时段的光伏发电功率求和,从而得到每天光伏发电总功率. 本算例采用的是 Intel i5-4258U 2.40 GHz CPU 和 4 GB RAM 的 PC 平台进行分析.

4.2 基于 OS-ELM 的预测模型的测试与分析

OS-ELM 随着新数据的输入,能够不断优化光伏预测模型,减小历史异常数据对预测结果的干扰,实现预测精度的提升. 为了进一步观察 OS-ELM 在训练过程中对预测精度的影响,现从训练样本中随机选出 40 个数据作为需要预测样本,将剩下的训练样本分为 4 组,每组中各包含 40 个数据,任选一组作为初始样本,形成初始预测模型,剩下 3 组训练样本逐次输入更新模型,观察各个模型的训练结果. 图 3~图 6 分别表示初始的 OS-ELM 网络的训练结果、第一次更新后的训练结果、第二次更新后的训练结果和第三次更新后的训练结果. 本次 OS-ELM 均采用“sig”作为核函数.

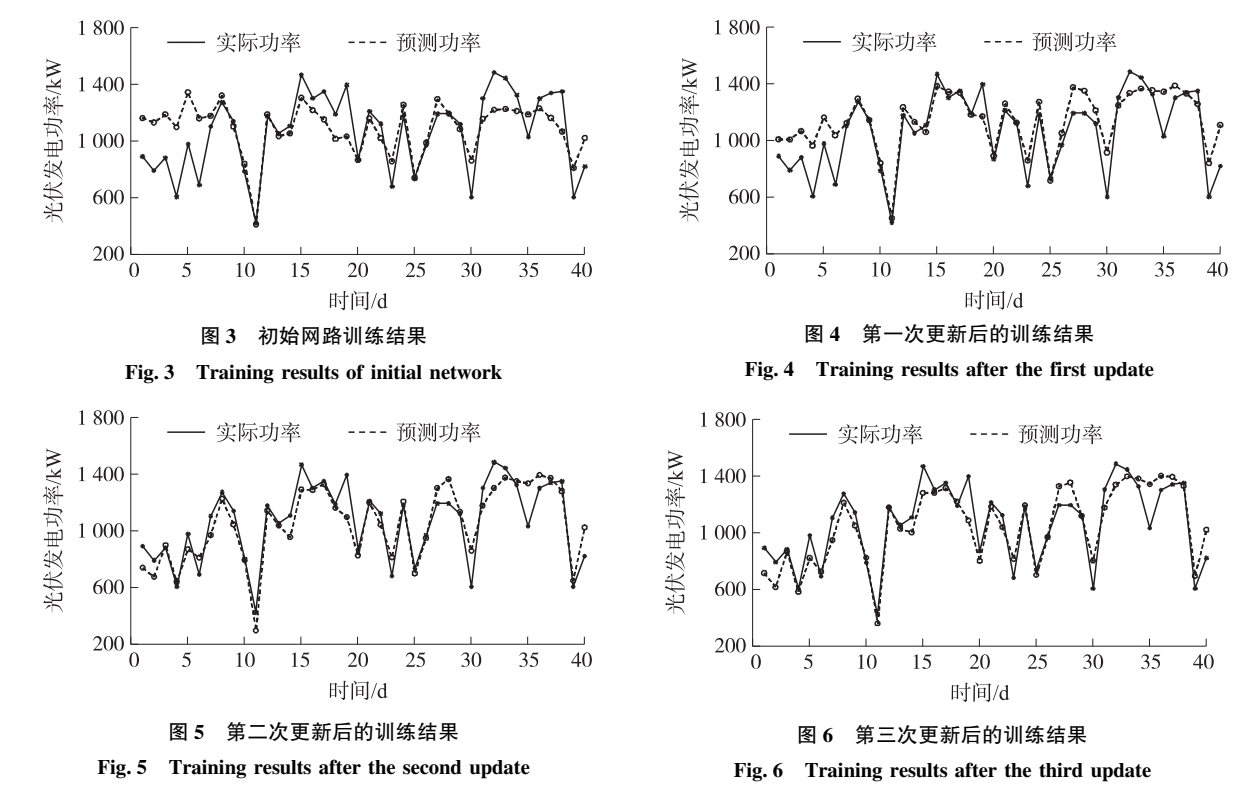


图 3 是仅加入初始训练样本得到的预测结果,由于初始训练样本中存在异常样本且样本数较少,大部分预测数据与实际数据存在着较大的误差. 图 4 是再次加入一组训练样本更新模型后的预测结果,预测结果有了较好的提升,但部分数据的误差仍然较大. 图 5 和图 6 分别是第二次和第三次加入新数据后的预测结果,与初始预测结果相比较有了明显的提升,大部分的预测结果都接近实际光伏发电功率,但仍有少数数据存在较小的误差.

表 2 给出的是模型更新后的统计误差. 由表 2 可以看出,在模型不断更新的过程中,两个评价指标都在不断的减小. 由最初加入训练样本得到的初始网络中,MAPE 值为 12.531 2%,RMSE 的值也偏大. 直到三次更新后,预测模型预测误差 MAPE 值仅为 6.801 3%,误差减小了 45%,RSME 值也减小了将近一半,预测精度有了明显的提升.

表 2 模型更新前后统计误差		
Table 2 Statistical error before and after model update		
模型	MAPE/%	RMSE/kW
初始网络	12.531 2	202.032 0
第一次更新	9.757 9	154.348 9
第二次更新	7.219 8	122.125 0
第三次更新	6.801 3	121.890 8

4.3 算法性能比较

为了了解 OS-ELM 中 3 种核函数对预测精度的影响,现在同一平台上,将 OS-ELM 的 3 种不同核函数分别对同一样本进行运算,观察预测计算时间和预测精度.此外,通过将 OS-ELM 与常见的 BP 神经网络和 SVM 方法进行比较,进一步验证 OS-ELM 的优势所在.

表 3 给出的是不同算法的性能对比.由表 3 中可以看出,OS-ELM 的 3 种核函数性能各不相同,“sin”耗时最少,而精度则是“sig”相对较高,但总体上与 BP 神经网络和 SVM 方法相比,耗时都远小于两种传统算法,同时精度也有所提升.由此可见 OS-ELM 与传统预测方法相比具有耗时短,预测精度高的特点.各种单一方法对比如图 7 所示,其中 OS-ELM 采用的是“sig”核函数,总体上 OS-ELM 的预测结果更加接近实际的光伏输出功率.

表 3 不同算法性能对比

Table 3 Performance comparison of different algorithms

算法	计算时间/s	MAPE/%
OS-ELM(sig)	0.031 3	6.472 9
OS-ELM(rbf)	0.140 6	8.276 5
OS-ELM(sin)	0.015 6	7.325 1
BP	1.203 1	11.035 9
SVM	2.784 1	10.467 4

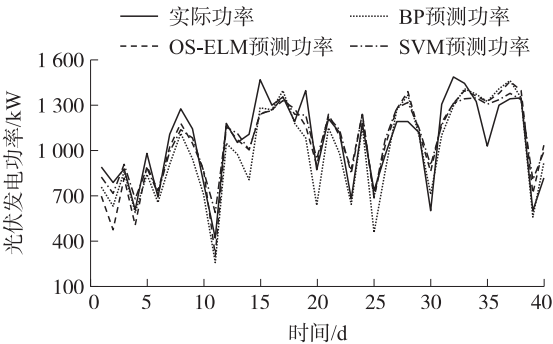


图 7 各种方法对比

Fig. 7 Comparison of various methods

5 结论

本文针对光伏发电中长期功率预测提出了一种基于 OS-ELM 的光伏发电功率预测方法,通过研究分析得到以下结论:

- (1) OS-ELM、BP 神经网络预测和 SVM 都可对光伏发电功率进行预测,但在相同的算例下,OS-ELM 的耗时更短,预测精度也更加精准.
- (2) OS-ELM 能够避免对历史数据进行重复训练,减小历史数据中异常样本对预测结果的影响,提高了光伏发电预测模型的鲁棒性.
- (3) 本文提出的光伏输出功率中长期预测方法,可以给电网规划调度提供一定的理论指导,具有一定的应用价值.

[参考文献](References)

[1] 丁明,刘志,毕锐,等. 基于灰色系统校正——小波神经网络的光伏功率预测[J]. 电网技术,2015,39(9):2438-2443.  
[2] 李卫良,黄堃,陈璐,等. 采用可控负荷平滑光伏发电功率波动的需求响应策略[J]. 江苏电机工程,2014,33(2):40-43.  
[3] 曹煜祺,张立梅. 光伏发电功率预测方法综述[J]. 黑龙江科学,2017,8(21):31-33.  
[4] AMINMOHAMMAD S,HIZAM H,RADZI M A M,et al. Modelling and prediction of photovoltaic power output using artificial neural networks[J]. International Journal of Photoenergy,2014(5):1-10.  
[5] YONA A,SENJYU T,FUNABASHI T,et al. Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead pv power output correction[J]. IEEE Transactions on Sustainable Energy,2013,4(2):527-533.  
[6] 卢静,翟海清,冯双磊,等. 光伏发电功率预测方法的探索[J]. 华东电力,2013,41(2):380-384.  
[7] 李乐,刘天琪. 基于近邻传播聚类 and 回声状态网络的光伏预测[J]. 电力自动化设备,2016,36(7):41-46.

- [8] 李建文,焦衡,刘凤梧,等. 基于相似时段的分时段光伏出力短期预测[J]. 电力自动化设备,2018,38(8):183-188.
- [9] 王继拓,王万成,陈宏伟. 基于回归——马尔科夫链的光伏发电功率预测[J]. 电测与仪表,2019,56(1):76-81.
- [10] 季顺祥,王琦,姚阳,等. 基于相似日和交叉熵理论的光伏发电功率组合预测[J]. 南京师范大学学报(工程技术版),2018,18(2):19-28.
- [11] 张程熠,唐雅洁,李永杰,等. 适用于小样本的神经网络光伏预测方法[J]. 电力自动化设备,2017,37(1):101-106,111.
- [12] 茆美琴,龚文剑,张榴晨,等. 基于 EEMD-SVM 方法的光伏电站短期出力预测[J]. 中国电机工程学报,2013,33(34):17-24.
- [13] 陈通,孙国强,卫志农,等. 基于相似日和 CAPSO-SNN 的光伏发电功率预测[J]. 电力自动化设备,2017,37(3):66-71.
- [14] 康重庆,夏清,刘梅. 电力系统负荷预测[M]. 北京:中国电力出版社,2017.
- [16] 王焱,汪震,黄民翔,等. 基于 OS-ELM 和 Bootstrap 方法的超短期风电功率预测[J]. 电力系统自动化,2014,38(6):14-19.

[责任编辑:陈 庆]