A NB-IoT based intelligent combiner box for PV arrays integrated with short-term power prediction using extreme learning machine and similar days

Caigui Zhang, Zhicong Chen¹, Lijun Wu, Shuying Cheng and Peijie Lin

College of Physics and Information Engineering, Fuzhou University, Fuzhou, China

¹ Email:zhicong.chen@fzu.edu.cn

Abstract. The grid-connected photovoltaic (PV) power stations are instability and volatility due to meteorological factors. A way to improve this problem is PV power forecasting. This paper proposed an improved short-term PV power prediction model that combines an extreme learning machine (ELM) neural network and similar day method. Firstly, a narrow-band Internet of Things (NB-IoT) intelligent combiner box data monitoring system is designed to collect multivariate meteorological factors and original PV output power datasets in different seasons. Secondly, the corresponding training set is selected according to the season type of the forecast day, and the Euclidean distance (ED) between the training set and the forecasting day is calculated, and the M-day with a small Euclidean distance is selected. Then, the N-day similar day data is divided among the M days as the new training set input, and the P-day optimal similar day data and the multivariate meteorological of the prediction day are divided as test set inputs. Finally, the ELM neural network prediction model is used to predict the output power of the predicted day. The experimental results show that the proposed method has the highest prediction accuracy in contrast to other two prediction models.

1. Introduction

With the increasing pollution of environmental pollution and fossil energy, the installed capacity of PV power generation has developed rapidly in recent years. However, the instability of meteorological factors leads to the strong intermittence, fluctuation and variability in the power output of PV power generation, which may lead to an increase in the cost of photovoltaic operation and maintenance. In order to improve this problem, many PV output power prediction methods are proposed [1], including long-term prediction and short-term prediction [2-3]. The long-term prediction is based on the historical data recorded by photovoltaic power stations and numerical weather prediction (NWP), and the prediction results are obtained by regression statistics. Instead, the short-term prediction has become the focus on research in recent years by combining machine learning technology. Wang et al. [4] combined the convolutional neural networks (CNN) with weather classification model to realize short-term prediction of photovoltaic power generation. Similarly, the literature [5] used a multilayer perceptron combined with an artificial neural network (ANN) for short-term power prediction. Shi et al. [6] used the Support Vector Machine (SVM) and weather classification to predict PV output power. Similarly, the literature [7] used the K-Nearest Neighbour (KNN) algorithm and similar day method to achieve short-term prediction of photovoltaic power. The CNN and the ANN prediction models have good prediction effects. The CNN prediction model needs a large number of training samples and parameter adjustments. Instead, the ANN prediction models require a large number of artificial neural
network parameters. Although the proposed SVM prediction model has better prediction performance than the traditional ANN model, it is insensitive to the missing data. KNN is a lazy learning algorithm.

In this study, a novel NB-IoT intelligent combiner box data monitoring system is designed to collect multivariable meteorological factors and the original PV output power datasets, and then an improved ELM neural network prediction model combined with similar day method is proposed for predicting the PV output power. Firstly, Correlation analysis is performed on the collected data to determine the correlation coefficient ($R^2$) between multivariable meteorological factors and PV output power. Secondly, the collected data sets are divided into different seasonal types (spring, summer, autumn, winter). The Euclidean distance between the characteristic parameters of training sets and the predicted day is calculated according to the type of predicted day and season, and the M-day with smaller Euclidean distance is selected. Then, the data onto the N-day similar day and the P-day optimal similar day is divided. Finally, the predicted output power of each day of each season is predicted by the ELM neural network forecasting model. Through the comparison of different prediction model training results, it is verified that the ELM neural network prediction model has higher training performance.

The structure of this paper is organized as follows: Section 2 describes the NB-IoT intelligent combiner box data monitoring system; Section 3 describes the data set selection and analysis processing, and illustrates the correlation between collected meteorological data and PV output power; Sections 4 introduce improved similar days combined with ELM neural network prediction models; experimental results and analysis are drawn in Section 5.

2. Data monitoring system

![Diagram of data monitoring system]

The schematic diagram of the data monitoring system is shown in Figure 1. The data monitoring system is mainly composed of a meteorological data acquisition system, a NB-IoT combiner box system, a message queue telemetry transmission (MQTT) server and an Android App client. The meteorological data acquisition system is capable of monitoring PV output power, air temperature and relative humidity, global horizontal radiation and diffuse horizontal radiation. The system transmits
data onto the LoRa module through the wireless transparent transmission. Then, the LoRa module sends the acquired data onto the NB-IoT combiner box, and the combiner boxes summarize and encode the data and uploads it to the MQTT cloud server. The MQTT server can implement data transmission and reception in a million-level message queue service. Local customers have direct access to local data, while mobile client users can remotely receive the multivariate meteorological data and the original PV output powers data onto the PV power system by connecting to the Internet. The physical map is shown in Figure 2.

3. Data set selection and analysis processing
The data set collects multivariate meteorological data for different seasons of PV arrays in a university in Fuzhou in 2017 and 2018. The data monitoring system can continue to collect multivariate meteorological data in the absence of sunlight on rainy days and nights, but the output power of PV power generation is small. Therefore, meteorological data and PV output power data during the daytime from 8:00 am to 16:00 pm in different seasons were selected, and one sampling point was selected as the data set for PV power prediction within 5 minutes.

In order to study the relationship between multivariate meteorological data and PV output power, this paper uses $R^2$ to analyze the correlation between them. The definition of $R^2$ is shown in Eq. (1) [8]:

$$R^2 = \frac{N \sum_{i=1}^{N} (X_{i}Y_{i} - \sum_{i=1}^{N} X_{i} \sum_{i=1}^{N} Y_{i})^2}{\sqrt{N \sum_{i=1}^{N} X_{i}^2 - (\sum_{i=1}^{N} X_{i})^2} \sqrt{N \sum_{i=1}^{N} Y_{i}^2 - (\sum_{i=1}^{N} Y_{i})^2}}$$

Where X and Y are multivariate meteorological factors and PV output power respectively. N is the sampling point numbers.

The $R^2$ results from multivariate meteorological data and PV output power are shown in Table 1. The higher the $R^2$ value, the greater the influence of the meteorological factors on the PV output power. The result indicated that global horizontal radiation, diffuse horizontal radiation and air temperature are positively correlated with PV output power. Therefore, solar irradiance and temperature will serve as important predictive model input variables due to their high positive $R^2$. The relative humidity is negatively correlated with PV output power. For the spring and autumn rainy season, relative humidity is one of the important factors affecting PV output power.

| Meteorological factors | Global horizontal Radiation | Diffuse horizontal Radiation | Air Temperature | Relative Humidity |
|------------------------|-----------------------------|-----------------------------|-----------------|-------------------|
| $R^2$                  | 0.9979                      | 0.9356                      | 0.6370          | -0.4670           |

4. Improved prediction model
The proposed PV power generation prediction model is based on an improved similar day method combined with an ELM neural network to predict the PV output power of a certain day in different
seasons. Firstly, the characteristic parameters that have a large influence on the PV output power are extracted according to the $R^2$. Then, according to the season type of the forecast day, the feature parameter data set of the corresponding season is selected, and the training set and the test set of the prediction model are divided according to the size of the Euclidean distance. Finally, we train meteorological data onto the ELM neural network prediction models to obtain the PV output power of the predicted day. The implementation process of the ELM prediction model is shown in Figure 3.

### 4.1. Improved similarity day method

The similar day method is to search for historical days similar to the forecasting day type, mainly including various characteristic parameters similar to the forecasting day [9]. In this study, there are mainly four-parameter indicators, including global horizontal radiation, diffuse horizontal radiation, air temperature and relative humidity. For the selection of characteristic parameters in different seasons, the main influencing factor of the spring and autumn rainy season is relative humidity. For the summer and winter seasons, the main characteristic parameters are air temperature and solar irradiance. This is also the basis of the selection of similar methods of this study. The specific method is as follows:

1. Step 1: The Euclidean distance between the feature parameter of the prediction day and the feature parameter of the training data set is calculated according to the season category. The Euclidean distance formula is defined as follows:

   \[ d = \sqrt{\sum_{i=1}^{n} (X_{ij} - X_{ij})^2} \]  

   Where $d$ is the Euclidean distance. $X_i$ and $X_j$ are the two characteristic parameter sequences.

2. Step 2: The first M-days of the smallest Euclidean distance in the training datasets were selected.
Step 3: From the M-day data, the N-day similar day is selected as the new training datasets, and the P-day optimal similar day is selected as the input of the test datasets.

4.2. ELM neural network prediction model
The ELM algorithm is proposed to the shortcomings of the single-hidden layer feedforward neural network (SLFN) algorithm. It mainly consists of an input layer, a hidden layer and an output layer [10]. The ELM neural network structures are shown in Figure 4. The ELM neural network has the characteristics of fast learning efficiency and good generalization performance. Compared with the traditional backpropagation neural network (BPNN), it is not necessary to continuously adjust the weights and thresholds, which can be set randomly. In view of the advantages of ELM, ELM was chosen as the forecasting model in this paper. The specific solution flow is as follows:

Step 1: The number of hidden layer neurons is determined, and the connection weight \( \omega \) between the deviation of the input layer hidden layer and the hidden layer neuron \( b \) is randomly set.

Step 2: The activation function of the hidden layer neurons is selected to calculate the hidden layer output matrix \( H \). The output function of ELM is shown in Eq. (3).

\[
T_j(x) = \sum_{i=1}^{\ell} \beta_i g(\omega_i \cdot x + b) = \sum_{i=1}^{\ell} \beta_i h_i(x)
\]

(3)

Where \( \beta_i \) is the weight vector of the \( i^{th} \) output node. \( h_i(x) \) is the output of the \( i^{th} \) hidden node.

\[
H = \begin{pmatrix}
g(\omega_{a_1} \cdot x_1 + b_i) & \cdots & g(\omega_{a_n} \cdot x_1 + b_i) \\
\vdots & \ddots & \vdots \\
g(\omega_{a_1} \cdot x_0 + b_i) & \cdots & g(\omega_{a_n} \cdot x_0 + b_i)
\end{pmatrix}
\]

(4)

Step 3: The connection weight \( \beta \) is solved by the least-squares solution to obtain the minimum output power error, and calculate the output layer weight \( \hat{\beta} : \hat{\beta} = H^T \).

5. Experiments and result analysis
In this study, in order to accurately evaluate the performance of the proposed prediction model for PV power prediction, including the standard deviation (\( \sigma \)), the root means square error (RMSE) and \( R^2 \) was used as evaluation criteria. \( R^2 \) is shown in Eq. (1). The equations for \( \sigma \) and RMSE are as follows [11].

\[
\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2}
\]

(6)

Where \( N \) is the number of samples. \( x_i \) represents the individual. \( \mu \) is the average.

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (f_i - \hat{f}_i)^2}
\]

(7)

Where \( f_i \) and \( \hat{f}_i \) are the real and predicted PV output power at i hour. \( N \) is the sample point numbers.
The datasets for this study used multivariate meteorological data and original PV output power collected by the data monitoring system from June 2018 to May 2019. In order to better verify the superiority over the ELM prediction model, the sample data for 4 days per season is selected respectively, including July 25, 2018 (summer), September 14, 2018 (autumn), December 29, 2018 (winter), and April 10, 2019 (spring). The forecasting PV power values and real PV power values are provided from 8:00 am to 16:00 pm every day with 1-hour time interval. The forecast result is shown in Figure 5. It is clearly shown that all prediction models can achieve fairly good prediction accuracy, especially the power predicted by the improved ELM prediction model is better than the traditional BPNN and GRNN neural network prediction model. For the spring and autumn rainy seasons and low winter temperatures, the improved ELM prediction model can better predict the output power of each season.

![Figure 5. Comparison of forecasting and measured power per hour in different seasons.](image)

### Table 2. Comparison of $R^2$ and RMSE results.

| Models | Index | Spring | Summer | Autumn | Winter | Average | $\sigma$ |
|--------|-------|--------|--------|--------|--------|---------|---------|
| ELM    | RMSE  | 0.06950 | 0.05159 | 0.07443 | 0.06998 | 0.06638 | 0.01010 |
| BPNN   | RMSE  | 0.09487 | 0.06054 | 0.08979 | 0.09407 | 0.08482 | 0.01634 |
| GRNN   | RMSE  | 0.12487 | 0.08309 | 0.12958 | 0.08405 | 0.10540 | 0.02528 |
| ELM    | $R^2$ | 0.99264 | 0.99829 | 0.99647 | 0.99787 | 0.99632 | 0.00257 |
| BPNN   | $R^2$ | 0.99349 | 0.99545 | 0.99464 | 0.98771 | 0.99282 | 0.00350 |
| GRNN   | $R^2$ | 0.98652 | 0.99465 | 0.99291 | 0.99554 | 0.99241 | 0.00407 |

The $R^2$ and RMSE results for each prediction model are shown in Table 2. From table 2, it can clearly see that the average RMSE of the improved ELM prediction model is 0.06638, which is the smallest compared to the BPNN and GRNN prediction models. For the prediction of four different seasons, the improved ELM prediction model shows better prediction superiority than the other two prediction models. The RMSE mean values of the proposed prediction model is 21.74% and 39.02% higher than the other two prediction models, respectively.

Furthermore, the degree of correlation between the predicted power curve and the real power curve is also an important indicator to measure the prediction effects. It can be seen from Table 2 that the $R^2$ average of the improved ELM prediction model reaches 0.99632, which is the most relevant among all
prediction models, indicating the improved ELM prediction model works best. The standard deviations from RMSE and $R^2$ for the improved ELM prediction model are 0.01010 and 0.00257, respectively, which is the smallest of all prediction models. Therefore, the error of the ELM prediction model proposed to this paper is very small, which can meet the accuracy requirements for prediction.

6. Conclusions
In this paper, a new NB-IoT intelligent combiner boxes data monitoring system is designed to collect multivariate meteorological data and PV output power data in real-time, based on which a novel ELM based short-term power prediction model is proposed to perform accurate power prediction for grid stability. The experimental results demonstrate that in contrast to the two other common prediction models based on BPNN and GRNN, the proposed improved ELM power prediction model achieves the highest prediction accuracy.

Acknowledgment
The authors would like to acknowledge the financial supports in part from the National Natural Science Foundation of China (Grant Nos. 61601127 and 61574038), the Fujian Provincial Department of Science and Technology of China (Grant Nos. 2019H0006 and 2018J01774), and the Foundation of Fujian Provincial Department of Industry and Information Technology of China (Grant No. 82318075).

References
[1] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison F J & Antonanzas-Torres F 2016 Review of photovoltaic power forecasting Solar Energy 136 78-111
[2] Lin P, Peng Z, Lai Y, Cheng S, Chen Z & Wu L 2018 Short-term power prediction for photovoltaic power plants using a hybrid improved Kmeans-GRA-Elman model based on multivariate meteorological factors and historical power datasets Energy Conversion and Management 177 704-717
[3] Sobri S, Koohi-Kamali S & Rahim N A 2018 Solar photovoltaic generation forecasting methods: a review Energy Conversion and Management 156 459-497
[4] Wang F, Zhang Z, Liu C, et al 2019 Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting Energy conversion and management 181 443-462
[5] Chen C, Duan S, Cai T & Liu B 2011 Online 24-h solar power forecasting based on weather type classification using artificial neural network Solar energy 85(11) 2856-2870
[6] Shi J, Lee W J, Liu Y, Yang Y & Wang P 2012 Forecasting power output of photovoltaic systems based on weather classification and support vector machines IEEE Transactions on Industry Applications 48(3) 1064-1069
[7] Chen C R 2017 k-Nearest Neighbor Neural Network Models for Very Short-Term Global Solar Irradiance Forecasting Based on Meteorological Data Energies 10(2) 186
[8] Bugala A, Zaborowicz M, Boniecki P, Janczak D, Koszela K, Czekala W & Lewicki A 2018 Short-term forecast of generation of electric energy in photovoltaic systems Renewable and Sustainable Energy Reviews 81 306-312
[9] Xu Z, Chen Z, Zhou H, Wu L, Lin P & Cheng S 2019 A short-term photovoltaic power prediction model based on a radial basis function neural network and similar days IOP Conference Series: Earth and Environmental Science 227(2) 22-32
[10] Huang G, Huang G B, Song S & You K 2015 Trends in extreme learning machines: A review Neural Networks 61 32-48
[11] Chowdhury D, Sarkar M, Haider M Z & Alam T 2018 Zone Wise Hourly Load Prediction Using Regression Decision Tree Model International Conference on Innovation in Engineering and Technology 1-6