Forecasting the Electricity Production from Renewable Energy in Shanghai

Yingjie Tian, Naiwang Guo

State Grid Corporation of China in Shanghai, Shanghai, China

Keywords: Renewable Resources; Cluster Analysis; Data Screening; Neural Network

Abstract: The development of new and renewable energy in China is still in its infancy. Comprehensive consideration can confirm that there is great potential for future development, but there are also many problems restricting development. From the point of view of the development of new technologies, this paper applies cluster analysis to the field of photovoltaics, and establishes a prediction model of photovoltaic power generation based on neural network. Aiming at the three weather types of Shanghai sunny, cloudy and rainy days, the abnormal samples in historical data are screened by cluster analysis, and the back propagation (BP) neural network prediction model is established by using the selected samples as training data. By comparing the results of the forecasting models before and after screening, it can be seen that the forecasting model based on the data filtered by clustering analysis has higher accuracy. Therefore, the combination of clustering analysis and BP neural network is an effective method to improve the forecasting accuracy of photovoltaic power generation.

1. Introduction

The development and utilization of new and renewable energy is still in the primary stage of development in China, and it accounts for a small proportion of the total primary energy consumption. From the point of view of energy utilization, with the gradual scarcity of fossil fuel resources and the consideration of serious environmental pollution caused by the application of fossil fuels, the development of alternative fuels becomes more and more urgent [1]. Among them, new energy and renewable energy have gradually become the focus of development and utilization because of their huge resource stock, $sustainable utilization, $environmentally friendly advantages. Therefore, it is necessary to make some prediction and scenario analysis on the future development of new and renewable energy in China, in order to facilitate the development of new and renewable energy in China. Be prepared [2]. There are many problems in the long-term forecast of the development of new and renewable energy in our country-many new energy technologies are still in the stage of demonstration projects and are difficult to be accepted by the market.'Our government has not yet formulated a sound policy support system for the development of new and renewable energy, and how to control the development of new and renewable energy in the future [3].There are great uncertainties in the development of bio-energy. Therefore, the long-term prediction of new and renewable energy development at this stage requires a lot of simplification and assumptions for complex situations [4].

2. New Energy Forecasting Theory

With the increasing installed capacity of photovoltaic power generation system year by year, the proportion of photovoltaic power generation in the power grid is also increasing [5]. The randomness of output fluctuation of photovoltaic system will impact the power grid system and affect the safe and stable operation of the power grid system. Therefore, the effective prediction of the power generation situation of photovoltaic system is of great significance to ensure the safe and stable operation of the power grid system [6]. The popularization of new and renewable energy in China can be regarded as a process of new technology entering the market. From this point of view, the method suitable for the popularization process of new technology can be used to predict [7]. Many experiences have shown that the growth of new technologies actually follows the

  DOI: 10.25236/scmc.2019.023

development law of growth curve, among which logical growth curve is one of the most widely used ones. In the initial stage, the popularity increases slowly with time, then accelerates gradually, and then decreases gradually when the growth rate reaches its peak. Finally, the popularity tends to saturate gradually, and the approaching speed becomes slower and slower [8]. At the same time, some studies show that the process of renewable energy technology development and commercialization development can be divided into the following stages: research and development stage'"technology demonstration and commercialization demonstration stage". Although the stages of utilizing new energy technologies in China are different, those potential new energy products will inevitably undergo a stage of large-scale popularization and price reduction, which usually meets the learning curve [9].

At present, there are several prediction models for photovoltaic power generation, such as support vector machine model [10], ARMA model, grey theory model and so on. Neural network model has good prediction performance, so this paper uses this model to predict the power generation of photovoltaic system. In the process of BP neural network prediction, the quality of input sample data will directly affect the prediction accuracy of the model. In the actual production process, errors will occur in the data acquisition and storage process of photovoltaic power generation system and abnormal samples will be formed. These abnormal samples deviate from the actual situation, and the model built on this basis can not accurately reflect the corresponding weather type, so it is necessary to screen the sample data. Because of the large amount of data recorded in the history of photovoltaic power plants, it is very difficult to screen the abnormal samples manually. This paper aims at the statistical data of different weather types, carries out cluster analysis and screens the historical data, and then establishes three types of photovoltaic generation in sunny, cloudy and rainy days by using the screened data based on BP neural network of MATLAB neural network toolbox. Electricity forecasting model.

3. Renewable Energy Forecast

3.1. Prediction model of photovoltaic power generation

In this model, clustering analysis and BP neural network are combined to screen the data recorded by the photovoltaic power station data collector, and the filtered data are used as the input parameters of the neural network to establish the photovoltaic power generation forecasting model. The photovoltaic power generation system is installed on the roof of the plant with a total installed capacity of 1.2 MW. the solar radiation and photovoltaic power generation were recorded hourly from 6:00 to 19:00 every day, and these data were taken as original samples.

Work flow of photovoltaic power generation forecasting model: Firstly, the theoretical solar radiation, actual solar radiation and photovoltaic power generation are clustered based on three typical weather types: sunny day, cloudy day and rainy day of shanghai , and outliers are screened according to clustering spectrum; then, the outliers are selected from January 2012 to January 2014. The historical data in January of 2001 were used as training samples. The theoretical solar radiation, power generation and maximum and minimum temperature of each day were used as input variables of BP neural network. The neural network model was trained. Based on the data of different weather types before and after screening, the prediction models of photovoltaic power generation were established respectively. In this paper, the screening data from February 2014 to January 2015 are substituted into the photovoltaic power generation prediction model, and the nonlinear learning ability of BP neural network is used to predict the power generation of photovoltaic power generation system under different weather types. Finally, the predicted power generation of photovoltaic power generation system is calculated and its average absolute percentage error (MAP) is calculated. E) Comparing with the actual power generation MAPE, the performance of the photovoltaic power generation prediction model is evaluated to verify the necessity of data screening.

3.2. Cluster analysis model

Clustering analysis is a process of dividing observation objects into groups or classes according to some data characteristics. This process makes the data objects within the same class have high similarity, but the similarity between different types of data objects is low. The key of clustering problem is to gather similar things together.

According to the different clustering methods, clustering analysis can be roughly divided into system clustering method, addition method, decomposition method and dynamic classification method. The purpose of clustering analysis in this paper is to deal with "outlier data" (abnormal samples) in the original data and improve the quality of the original data. Therefore, the system clustering method is selected in this paper. The main calculation steps of the algorithm are: (1) transforming the centralized sample objects; and (2) dividing the processed objects into n categories, each of which has its own characteristics. Include one sample; calculate the distance between each two classes according to the metrics; merge the two closest classes into one new class; _calculate the distance between the new class and the current classes, then merge the two closest classes into one class, and continue to run this step until all classes (or samples) are combined into one class; Draw the pedigree of the sample set; _According to the number of classifications, get the corresponding classification results. In this paper, the clustering results of 8 groups of rainy day sample data selected randomly are shown in Figure 2. The samples can be clustered into 2 to 5 categories. As shown in Figure 2, when the distance between classes is 5, the samples can be grouped into three categories, of which 2-4, 6, 7 are classes a, 1, 8 are classes b, and 5 are classes C.

The selection of metrics directly affects the quality of clustering results. Therefore, statistical analysis of historical power generation data in the database is needed to determine the metrics of the system clustering method. Taking cloudy weather as an example, Fig. 3 shows the scatter plots of actual solar radiation and photovoltaic power generation during the experimental period, and Fig. 4 shows the scatter plots of theoretical solar radiation and actual solar radiation during the experimental period. It is found that part of the sample data points are far away from the main data set, forming outliers, and the data distribution is approximately a slant line. Choosing cosine coefficient as the criterion, the data near the oblique line are classified as valid data, and the data far from the oblique line are classified as outliers. After eliminating outliers, the optimal samples are formed, and then the degree of dispersion (cosine similarity coefficient) of the optimized samples is calculated.

3.3. Prediction model evaluation of photovoltaic power generation

In this paper, based on Levenberg-Marquardt (LM) algorithm, a two-layer feedforward neural network is built and trained by using the neural network toolbox of MATLAB 2010b. Finally, the simulation results are obtained. The training data and test data (validation data) of the neural network are randomly selected, of which 90% are training data and 10% are validation data. The validation data is not involved in training, but to ensure the credibility of the simulation results. The average absolute percentage error is used to calculate the simulation results, and the whole photovoltaic power generation system is evaluated based on the calculation results:

$$MAPE = \left(\frac{100}{N}\right)\sum_{z=1}^{N}\left|\frac{P_f^z - P_a^z}{P_a^z}\right|\% \tag{1}$$

In the formula: N is the total data; Pf is the actual photovoltaic power generation; Pa is the predicted photovoltaic power generation; Z is the data serial number. According to equation (2), the lower the MAPE is, the better the prediction effect of the model is. Clustering process and results This paper extracts the original data of three weather types, sunny, cloudy and rainy days, and uses statistical analysis software SPSS to cluster different weather types based on cosine coefficient. Taking cloudy weather as an example, according to clustering pedigree, the number of samples contained in each cluster is compared. There are 10 samples in category a, 30 samples in category b, 61 samples in category C and 67 samples in category D. Then, several samples with the least clustering number are screened. Clustering analysis reduces the degree of data dispersion.

Table 1 The clustering result statistics for all kinds of sample data

Weather category	sunny	Cloudy	rain	Total
Before screening	150	168	61	379
After screening	134	146	48	328
Screening ratio(%)	10.7	13.1	21.3	13.6

The clustering results of various sample data are shown in Table 1. For sunny days, the weather condition is relatively stable and the data dispersion is small, so the number of anomalous samples screened out is less, which is 10.7% of the total number of samples; for cloudy weather, the change of cloud layer is random, so the number of anomalous samples screened out is 13.1%; for rainy days, the weather condition is not stable enough, the number of anomalous samples is not enough. According to the large degree of dispersion, the number of anomalous samples screened out is 21.3%.

Table 2 MAPE values of various weather neural network models

Weather category	sunny	Cloudy	rain	Average
Before screening	32.4	45.6	85.8	46.8
After screening	20.7	22.6	34.5	23.6

Before data screening, no matter which weather type the model is based on, the difference between the predicted value and the actual value of photovoltaic power generation is larger; after data screening, the difference between the predicted value and the actual value of photovoltaic power generation is smaller. Evaluation results of neural network model prediction (Table 2): For sunny days, the prediction results of the two models before and after screening are close. Compared with the MAPE of the pre-screening neural network model, the MAPE of the screened neural network model decreases by 11.7% and the prediction effect is relatively good; for cloudy days, the prediction results of the two models before and after screening are relatively good. The results are quite different. Compared with the pre-screening neural network model, the MAPE of the screened neural network model decreases by 23%. This is because the thickness, cloud shape, cloud amount and cloud location are random in cloudy weather, and the uncertainty of shading the sun is high. For rainy days, the two models are screened before and after the screening. Compared with the pre-screening neural network model, the MAPE of the screened neural network model is reduced by 51.3% and the prediction error is larger. This is because the weather changes in rainy days are more complex. For different rainy days, there are great differences in cloud, rain and water vapor, and photovoltaic power generation system is too large. There are also great differences in the absorption of solar radiation energy, resulting in larger relative errors. In addition, the average value of the filtered neural network model MAPE is 23.6%, which is less than the average value of the filtered neural network model MAPE. Therefore, the prediction model of photovoltaic power generation based on clustering analysis data is more accurate, which shows that it is necessary to screen the data.

4. Conclusion

In this paper, clustering analysis and BP neural network are combined to screen the abnormal samples in the historical monitoring data of photovoltaic power plants. Then, based on the screening results, a prediction model of photovoltaic power generation is established, and the prediction results are evaluated and analyzed. The results show that using clustering analysis method to screen historical data can effectively eliminate the more discrete operation data (abnormal samples); using the samples filtered by clustering data as training data, and establishing BP neural network prediction model can effectively improve the accuracy of prediction results. The research results of this paper can provide an effective method for the prediction technology of photovoltaic power plant, which is conducive to the popularization and application of photovoltaic power generation technology.

Acknowledgement

Supported by Science Project of State Grid corporation of China in Shanghai.

References

[1] Ellabban O, Abu-Rub H, Blaabjerg F. Renewable energy resources: Current status, future prospects and their enabling technology. Renewable and Sustainable Energy Reviews, 2014 39 748-764.

[2] Alotto P, Guarnieri M, Moro F. Redox flow batteries for the storage of renewable energy: A review. Renewable and Sustainable Energy Reviews, 2014 29 (7) 325-335.

[3] Hamid Elsheikh M, Shnawah D A, Sabri M F M, Mohamad M. A review on thermoelectric renewable energy: Principle parameters that affect their performance. Renewable and Sustainable Energy Reviews, 2014 30 337-355.

[4] Sen R, Bhattacharyya S C. Off-grid electricity generation with renewable energy technologies in India: An application of HOMER. Renewable Energy, 2014 62 388-398.

[5] Bhandari B, Poudel S R, Lee K T, Ahn S H.. Mathematical modeling of hybrid renewable energy system: A review on small hydro-solar-wind power generation. International Journal of Precision Engineering and Manufacturing-Green Technology, 2014 1 (2) 157-173.

[6] Scarlat N, Dallemand J F, Monforti-Ferrario F, Banja M, Motola V. Renewable energy policy framework and bioenergy contribution in the European Union – An overview from National Renewable Energy Action Plans and Progress Reports. Renewable and Sustainable Energy Reviews, 2015 51 969-985.

[7] Chauhan A, Saini R P. A review on Integrated Renewable Energy System based power generation for stand-alone applications: Configurations, storage options, sizing methodologies and control. Renewable and Sustainable Energy Reviews, 2014 38 99-120.

[8] Introducing the Temporal Distortion Index to perform a bidimensional analysis of renewable energy forecast. Energy, 2016 94 180-194.

[9] Ogimoto K. Generation Forecast of Renewable Energy and System Technology. Ieej Transactions on Power & Energy, 2014 134 (6) 473-476.

[10] Li C, Liu X, Cao Y, Zhang P, Shi H, Ren L, Kuang Y. A Time-Scale Adaptive Dispatch Method for Renewable Energy Power Supply Systems on Islands. IEEE Transactions on Smart Grid, 2015 7 (2) 1-10.