Applied Energy 206 (2017) 683–696

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Day-ahead forecasting of photovoltaic output power with similar cloud space fusion based on incomplete historical data mining

Xiuyuan Yang[a], Minglu Xu[a], Shouchen Xu[b], Xiaojuan Han[c,*]

[a] Automation Institute, Beijing Information Science and Technology University, Beijing 100085, China
[b] Department of Logistics, State Grid Management Institute, Beijing 102299, China
[c] School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

HIGHLIGHTS

- Similar-day matrixes under different weather conditions are obtained by statistical indicators.
- The similar cloud correct interval is determined by forward cloud generators.
- The prediction model of photo-voltaic output power based on the Markov chain is proposed.
- A space fusion forecasting model of the photo-voltaic output power is established by backward cloud generators.

ARTICLE INFO

Keywords:
Cloud interval fusion
Similarity day
Cloud modeling theory
Photovoltaic power day-ahead forecasting

ABSTRACT

Affected by many meteorological factors, the output power of photovoltaic power generation systems is random and fluctuating, so it is uncontrollable for a large power grid. With the increase of the capacity of photovoltaic connected to grid, its impact on the large power grid can't be ignored. Due to the limited and incompleteness of historical photovoltaic output power and meteorological data, a day-ahead forecasting method of the photovoltaic output power with similar cloud space fusion based on incomplete historical data mining is proposed. Through statistical analysis of historical photovoltaic power data, the statistical indicators under different weather conditions are used to obtain similar-day matrixes by Euclidean distance clustering. The similar cloud interval is determined by forward cloud generators, which is used to correct the longitudinal predicting values obtained by Markov chain prediction model. Regarding the photovoltaic power value of the previous day as the output of the persistent prediction model, combined with the predicting value obtained by similarity cloud interval correction, a space fusion forecasting model of the photovoltaic output power is established by backward cloud generators to realize a day-ahead accurate forecasting of the photovoltaic output power. The simulation tests based on the measured data of the photovoltaic systems at a photovoltaic power station in China verify the effective and correctness of the proposed method. The results show that the model has good forecasting accuracy, and has certain practicability and feasibility.

1. Introduction

With the persistent progress of solar power generation technology and the rapid growth of global installed capacity of photovoltaic power generation, grid-connected solar power generation technology has become an important development direction of solar photovoltaic power generation [1]. However, the randomness, volatility and intermittent of photovoltaic power generation make the large-scale photovoltaic grid-connected power generation have a great negative impact on the safety, stability and economic operation of the grid [2,3]. Therefore, accurately predicting the output power of the photovoltaic power station can effectively reduce the adverse effects of grid-connected photovoltaic generation on the grid, and help the grid dispatching department adjust the scheduling plan in real time, and arrange the coordination of photovoltaic power generation and other power sources in order to achieve the rational operation requirements of the grid [4].

In term of different classification criteria, the prediction methods of the photovoltaic output power are different [5]. According to different time scales [6], it can be divided into short-term and ultra-short-term power prediction. The time scale of the ultra-short-term power prediction is from 0.5 to 6*h*, and the time scale of the short-term power prediction is from 1 to 2 days. According to the different principle of

* Corresponding author.
E-mail address: wmhxj@163.com (X. Han).

http://dx.doi.org/10.1016/j.apenergy.2017.08.222
Received 17 April 2017; Received in revised form 9 August 2017; Accepted 27 August 2017

forecasting models, it can be divided into physical method and statistical method. The physical method is to take the weather forecast data as the input using the physical equation to achieve the prediction. Statistical method is to find out the inherent law of the prediction through the statistical analysis of historical data. At present, the researches on the short-term prediction of the photovoltaic output power are mainly focused on the following two aspects: ① **Indirect prediction:** The solar radiation intensity and weather forecast information are used to predict the solar radiation intensity of the surface, and then the prediction of the photovoltaic output power is realized by the photovoltaic output power system model. However the indirect prediction method depends on complex solar radiation intensity model and accurate weather forecasting information, which is very difficult to be implemented because of the influence of the photovoltaic array conversion efficiency, installation angle and other parameters. ② **Direct prediction:** The historical data of the photovoltaic power generation system and weather forecast information are used to directly predict the output power of the photovoltaic power station. The direct prediction method mainly includes neural network, support vector machine, time series and Markov chain and so on.

At present, most of the researches on photovoltaic power generation mainly focus on solar energy monitoring system, short-term prediction of the photovoltaic power generation and performance prediction of grid-connected photovoltaic systems. Main prediction methods [7–12] include Persistent model, Artificial Neural Networks (ANNs), Markov chain and Support vector machine and so on. In [9], the artificial neural network (ANN) was adopted to mimic the nonlinear correlation between the metrological parameters and energy generated by the photovoltaic system in order to find that short-term prediction performance is comparable with real-time prediction performance when ahead solar angles are applied to the predictions. However, its application is based on a large number of a prior data, which limits its engineering applications. The Markov chain is suitable for the time series forecasting with larger random fluctuation which can reveal the stochastic performance of the system affected by many complicated factors. In [10], a forecasting method based on Markov chain is proposed to predict the output power of a photovoltaic power station directly by historical power data modeling. In [11], a Markov-switching forecasting model was used to establish the nonlinear model of wind speed time-series. Compared with traditional forecasting methods, the Markov-switching model can offer both of the point and interval forecasting.

According to above forecast characteristics of the Markov chain, a longitudinal prediction model of the Markov chain is proposed to realize a day-ahead prediction of the photovoltaic output power. The state transition matrix at the same time from the previous day to the next day is obtained by probability statistics on historical data at the same time every day, and the power value of the corresponding time on the predicting day is calculated based on the state transition matrix at each time. However, the fluctuations of the photovoltaic output power are strongly correlated with the meteorological condition. When the meteorological conditions change drastically, the prediction accuracy of the model needs to be improved. How to find the historical day with the highest similarity to the meteorological feature vector as the similar-day of the predicting day is the focus of the study. Similar-day classification modeling according to different weather types is an effective means to improve the predicting accuracy of the photovoltaic output power under various weather statuses. In [13], a prediction model based on the similar-day selection algorithm and Elman neural network was provided to achieve the prediction of the photovoltaic output power, in which the similar-day is searched by the meteorological feature vector. In [14], a novel short-term photovoltaic power forecasting method based on knowledge mining and intelligent algorithm was proposed. The climatic conditions to be predicted and historical climatic conditions were classified by fuzzy C-means method, and the nearest similarity was searched. The relationship between weather conditions and photovoltaic output power was established by Bayesian neural network.

However, the weather type of historical data is missing in some cases (here called incomplete data), which will cause great difficulties to similar-day classification modeling. In [15], to identify the missing the weather type of historical data, a solar irradiance feature extraction and support vector machines based weather statuses pattern recognition model for the short-term prediction of the photovoltaic output power was presented. To ensure the feasibility and reduce the workload of classification modeling, four generalized weather classes covering all weather types were constituted with input features extracted from solar irradiance data. Due to the changes of the radiation intensity and temperature in different weather conditions, the photovoltaic output power is different. Meanwhile statistical indicators that reflect the relationship between the photovoltaic output power and weather information or radiation intensity can be mined through the statistical analysis method, and a similar-day curve will be determined by the statistical indicators of the photovoltaic output power in the case of unknown weather information. For the randomness and volatility of the photovoltaic output power in different weather conditions, only using a similar-day curve to predict the photovoltaic output power will result in a certain predicting error, a typical similar-day correction interval based on cloud modeling theory is used to eliminate the influence, which is a new approach to represent the randomness and fuzziness between uncertainty language and precise numerical value and has been applied to data mining, decision analysis, intelligent control and image processing [16–18].

A novel day-ahead forecasting method with similar cloud space fusion of the photovoltaic output power based on the incomplete historical data mining is proposed in this paper. According to the historical data of the photovoltaic output power in different seasons, six statistical indexes in the same season were obtained. By the Euclidean distance, the similarity matrix is obtained under different weather conditions in different seasons, and the similar-day curves and distribution intervals are formed by forward cloud generators. Based on Markov chain theory, the transition state matrix at the same time in the corresponding season in the next year can be calculated to predict the photovoltaic output power in the next 24 h. According to weather forecast information, the category of the day to be predicted is determined, and the predicting value is modified by the distribution curve of the similar day to form a new predicting value, combining with the value obtained by the persistent forecasting method, a space fusion forecasting model is established by backward cloud generators to realize a day-ahead prediction of the photovoltaic output power based on the historical data mining taking into account the randomness and volatility of the photovoltaic output power. A simulation test of the operation data at a photovoltaic power station with the installed capacity of 40 MW is carried out and the simulation results verify the effective and correctness of the proposed method. Compared with the traditional method, it can be seen that the proposed method can further improve the prediction accuracy of the photovoltaic output power and has better engineering application prospect.

2. Statistical clustering of photovoltaic output power with similar cloud distribution interval based on cloud modeling theory

2.1. Descriptive statistical indicators of the photovoltaic output power

There are many factors that affect the output power of a photovoltaic power station, such as solar radiation intensity, the conversion efficiency of the inverter, solar panel installed angle, temperature, relative humidity, rainfall and air pressure and so on. Monthly cumulative average of the total solar radiation intensity at a meteorological station from 1985 to 2015 is shown in Table 1.

It can be seen from Table 1 that the total solar radiation intensity was the highest in late spring and the whole summer, that is, and the lowest in winter, which is consistent with the regional astronomical