Power forecast research of photovoltaic system based on double-level neural network

K Cheng[1], Q Q Sun [2, 3] and X Y Ma[2]

[1]Solar Energy Research Institute, Northwestern Polytechnical University, Xi'an 710072, China
[2]School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, China

E-mail: 1564015758@qq.com

Abstract. A photovoltaic generation system is affected by many factors. The traditional method of forecasting photovoltaic power generation is more complicated. A double-level neural network is designed with consideration of these factors: the first level network calculates a real solar radiation with input of theory radiation and weather coefficient; the second level network calculates the final result generation power with input of real solar radiation which comes from the first network and the highest temperature which comes from weather forecast. The final power results indicate this method has good forecast capacity, calculate results are very close to measure value and errors are less than single level network.

1. Introduction

Solar photovoltaic power generation is developing rapidly, and has great potential for application. The accurate prediction of the power generation capacity of the photovoltaic system can reduce the impact of the fluctuation of the photovoltaic system on the power grid and make the photovoltaic system operate efficiently. At present, the methods for predicting the output power of a photovoltaic system mainly include the following:

The first is to establish the overall energy balance of solar cells, that is, the energy absorbed by photovoltaic cells is equal to the sum of radiation heat transfer, convection heat transfer and power generation, and then solved by means of iteration. The problem is that many Solar cell physical parameters need to be solved, such as the material's emissivity, absorption rate and convection heat transfer coefficient, these parameters are generally difficult to obtain the exact value, and each manufacturer's photovoltaic cells are not the same [1].

The second one is the mathematical model of the equivalent circuit of solar cells, which is summarized by the related theories of solid state physics and electronics. However, because of the complexity of the model and the large number of parameters, it is difficult to obtain accurate data. The I-V characteristics of any irradiance and temperature can be obtained by the simplification of the model and the fitting of the mathematical curve [2, 3].

The third is a self-learning algorithm using neural networks: the purpose of training the prediction with existing experimental data. The advantage of this approach is that the measured data that is used to train itself contains all the information of the system, thus avoiding the analysis of complex model details [4].

Comparing the advantages and disadvantages of the above methods, in order to simplify the prediction and avoid the analysis of complex mathematical models, this paper selects the neural network method.

Further, the existing single-stage neural network uses the power sequence of the previous day, the two-day weather type index and the two-day maximum temperature to train the network. Therefore, when forecasting the power of a certain day, it is required to have the previous day's measured data. However, this paper designed a two-level neural network method, when the network training is good, you do not have to rely on the measured value, only the theoretical calculations and the weather forecast. The result proves that this method is more flexible in application while ensuring the accuracy of prediction.

In this paper, the experimental results of two kinds of networks are validated and compared with the measured data of the photovoltaic power plant in Shanxi University of Science and Technology.

2. Neural network model

2.1. Neural network
An artificial neural network is a network system that processes the information in a parallelly way and nonlinearly by simulating the way the human brain processes information. Among them, BP (back propagation) neural network is a kind of forward-oriented network that uses error back propagation algorithm and its structure is shown in figure 1:
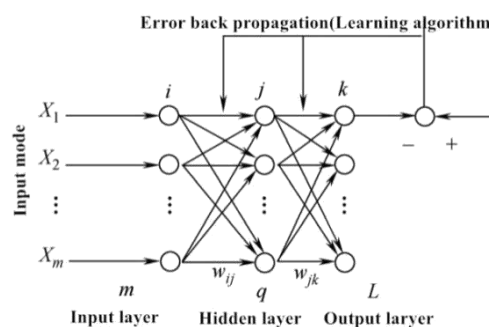
Figure 1. BP neural network structure.

Where Wij is the connection weight between the input layer and the hidden layer node, Wjk is the connection weight between the hidden layer node and the output layer node, and the input of the hidden layer and output layer node is the weight of the output of the previous layer node and [5].

The three-layer neural network has been able to approximate any complex function, so this paper also uses three-layer neural network.
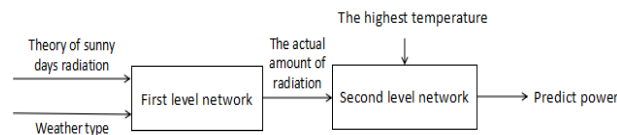
2.2. Neural network input
There are many factors affecting the output of PV system, including the type of weather, location, ambient temperature, installation angle, array conversion efficiency, etc. The biggest advantage of neural network method is that the training sample data itself can contain all the information of the system without going in to deep analysis of each variable. For example, the power of a certain day at a certain point contains information such as the amount of irradiation with time, the geographical location, the installation angle of the photovoltaic array, the conversion efficiency and the like. Therefore, when selecting inputs, for a given PV system, only the variables that are independent and directly affected are selected.

For a single-stage neural network, as in [5], the input is the previous day power sequence, the two day weather type and the highest temperature. Since the solar irradiance of two consecutive days is highly correlated, the power of the previous day has included this information, and the biggest impact

is the change in weather types. The type index of the weather should be given a quantification, and it becomes a value between 0 and 1. This paper is divided into clear, cloudy, cloudy, rain four types.

The design idea of the double-level neural network is to get rid of the high dependence on the measured data of the previous day. Therefore, the input of the double-level neural network adopts the calculated value or the predicted value, and learns the existing measured value through the network so as to correct the theoretical value and get the prediction result [6]. The double-layer neural network input and output is shown in figure 2:

Figure 2. Double-level neural network input output diagram.

There are a large number of monographs and literature on the theory of a sunny day's radiation, which is not the scope of this article, and will not be repeated here.

Since the commonly used incentive functions are both S-type function (Logarithmic S function logsig, tangent type S function tansig), the input / output range of the function is between (-1,1) or (0,1), so the input and output matrix must be normalized so that the data can fall within the larger range of the function intermediate gradient to avoid neuronal saturation and improve the accuracy and generalization ability. A commonly used normalization method is:

$$T_n = \frac{t_n - t_{\min}}{t_{\max} - t_{\min}} \tag{1}$$

Where: Tn - the normalized data;tn- Data before normalization;tmax, tmin- the maximum and minimum values in the data matrix

2.3. Hidden layer neurons and the choice of incentive function
The number of hidden layer neurons has a great influence on the network. Generally speaking, the larger number of neurons can reduce the error and improve the accuracy. However, too many neurons, the network is too complicated, will increase training time, prone to over-fitting phenomenon and the generalization of the network will be poor.

At present, the theoretical calculation method of the number of neurons has not been determined.It can only be determined through experience and trial and error. For a simple three-level BP neural network, according to the Kolmogorov theorem:

$$n_2 = \sqrt{n_1 + m + 1} + a \quad _{a = 1 \sim 10} \tag{2}$$

where n2: the number of hidden layers; n1: Enter the number of layers; m: The number of output layers

From (2) we can roughly determine the hidden layer neuron range, and then try to determine the best way through the trial and error, this article is also uses this method. The number of hidden neurons finally determined is 25 for single-level networks, the first level of a double-level neural network is 20, and the second level is also 20.

As mentioned earlier, the incentive stimulus functions are logarithmic and tangent functions of type S, and the hidden layer is a logarithmic type of logsig function with output exactly between (0,1). Therefore, the output layer is a tangent tansig function.

3. Photovoltaic power generation system
The experimental data used in this study all come from Xi'an University of Science and Technology

Shaanxi Photovoltaic System. The system is installed on the roof of teaching building. According to the additional photovoltaic design of buildings, 3680 blocks 240 Wp polysilicon photovoltaic modules are used, with a total installed capacity of 883.2 KWp; 5 three-phase grid-connected inverters with a total capacity of 900 KW; Three-phase low-voltage AC distribution network, spontaneous use; and set the appropriate lightning protection, grounding and monitoring equipment.

Remote data logging can be achieved with the provincial data monitoring platform data upload network construction, remote monitoring and communication functions. The main control configuration of a data collector, the use of high-performance industrial control PC as the host of the system monitoring, data acquisition through the RS485 interface with the inverter, combiner boxes, environmental tester communication can record the PV module output voltage and current characteristics, PV inverter output voltage, current, power, power generation, lighting, emission reduction and other parameters as shown in figure 3.

Figure 3. Photo of photovoltaic power station.

The system started trial operation in February 2013 and was operated from March so far. The system runs normally and the data monitoring record is complete. In this study, the data of June and July were selected as the original samples.

4. Neural network model training and results

4.1. Model training
The selection of training samples is crucial for the quality of training and the generalization ability of the network. The samples must be large enough and representative enough to be as complex as the actual situation, but not too large. If the sample is too large the redundant information is too much. Not only the training time will be longer, but also easy to fall into the local minimum [7].

For the first level network, because the input has two days before and after the weather index, so selected training samples include a variety of weather combinations, such as continuous sunny, continuous rain, from clear to negative, etc., to ensure the representativeness of the sample.

The same is true for double-level neural networks, including various weather types.

In addition, the sample selection should pay attention to short periods of time excluding the sudden change of the shower, as well as obvious inconsistencies with common sense, for example, the power generation in rainy days is greater than the previous day's sunny day. Such samples are unfavorable for network regularity learning.

The training samples should be normalized as before.

Finally, 11 days of data meeting the above requirements were selected from the original sample to train the network.
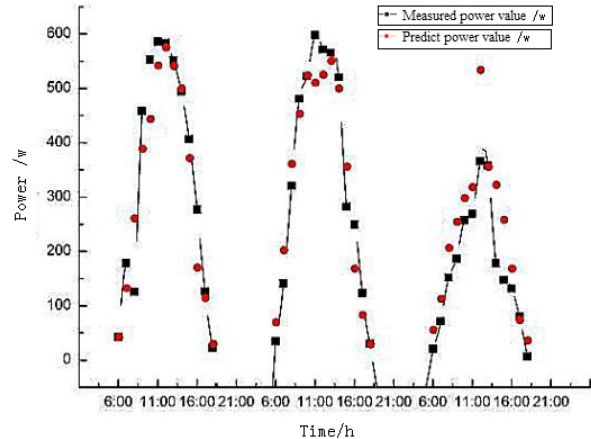
4.2. Forecast results

All of the network establishment, training and simulation of results in this paper are done by the neural network toolbox of Matlab 2010b.

The network has been trained to test, select three days to predict. One of the first two days is sunny, the third day becomes cloudy. These three days of data for the network is brand new, the only way to truly test the generalization of the network capacity.

Figures 4 and 5 show the power prediction results of single-stage network and two-stage network, respectively. It can be seen that no matter which network is accurate for predicting the trend of power variation in one day, and for the change of weather type good predictive ability.
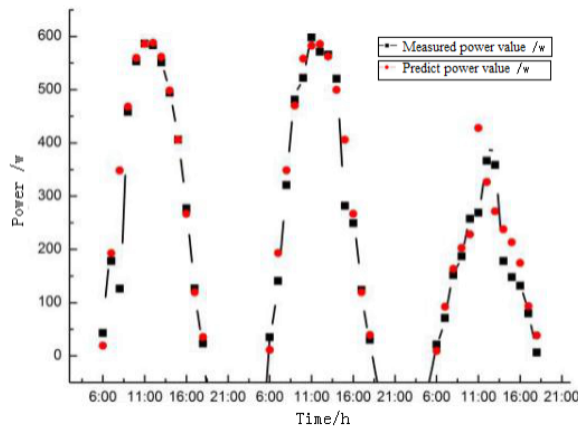
Figure 4. Single-stage network prediction results.

Figure 5. Double-level neural network forecast results.

4.3. Results analysis

In order to quantify the quality of the network prediction, we introduce a correlation evaluation. The correlation coefficient is a measure of how similar two things are. If we compare two things with the sequences x (n) and y (n) and compare the similarity of the two sequences in the range of n = a to b, then the correlation coefficient r is:

$$r = \frac{\sum_{n=a}^{b} x(n) y^*(n)}{\left\{ \sum_{n=a}^{b} |x(n)|^2 \cdot \sum_{n=a}^{b} |y(n)|^2 \right\}^{1/2}} \tag{3}$$

Where the "*" sign indicates conjugate to y (n).

Table 1 shows the correlation coefficients of the two network prediction results. It can be seen that the prediction results are highly correlated with the measured results. Generally, the single-level network is slightly better than the double-level neural network.

Table 1. Predictive results and the measured correlation coefficient.

	first day	the next day	the third day	average
single-stage network	0.9582	0.9823	0.9343	0.9583
double-level neural network	0.9594	0.9727	0.9270	0.9530

In order to calculate the relative error of each point and eliminate the influence of stochastic fluctuation on the result, and embody the regularity, the two sets of data are respectively fitted by quadratic polynomial, as shown in figure 6 and 7:
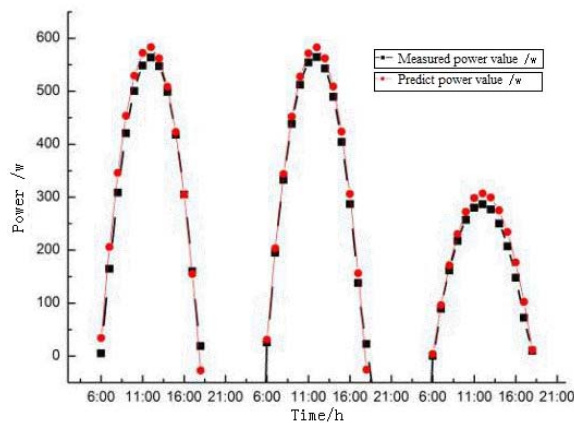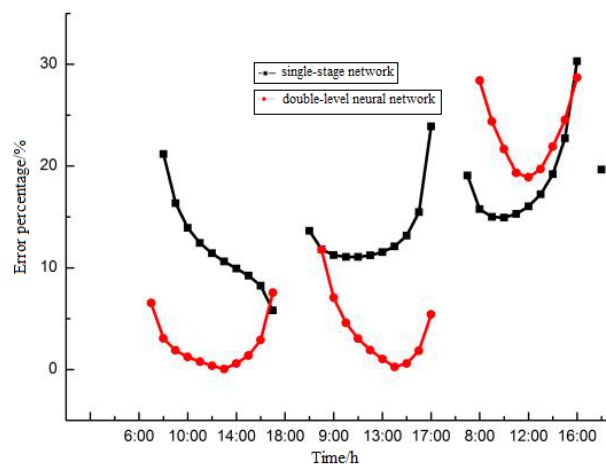
Figure 6. Single-stage network fitting results.
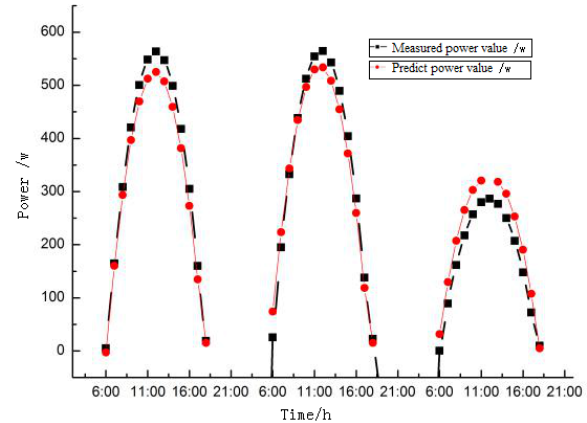
Figure 7. Double-level neural network fitting results.

Calculate the relative error of each point, discard the data point of gross error, get the relative error curve shown in figure 8:

Figure 8. Relative error curve.

As can be seen from the figure, the morning and evening errors are relatively large. This is because the power generated by morning and evening is small. Even if the absolute error is small, the calculation base is small, so the relative error becomes large, The two time periods have little effect on the power generation throughout the day. Therefore, we should mainly consider the core period of PV system operation (9: 00-16: 00), which accounts for more than 80% of the electricity generation in the whole day. It should be the most concerned time period of our forecast. The relative error of the three days is shown in table 2:

Table 2. Average relative error of three days.

	first day	the next day	The third day	average
single-stage network	11.52%	12.12%	18.84%	14.16%
double-level neural network	1.16%	2.45%	22.88%	8.22%

Three-day errors in single-stage networks are relatively average, while two-stage networks are

accurate for two sunny days and large errors in cloudy days, mainly due to error accumulation in double-level neural networks. From the training samples it can be seen, even if the day is cloudy, the fluctuations in the weather is still relatively large, so the cloudy forecast result is not very good. Although the result is a certain error, it still has great significance for guiding the actual project.

5. Conclusion

In this paper, a two-level neural network-based photovoltaic power forecasting model is designed. The structure of the network, the choice of input, the design of hidden layer and the selection of training samples are analyzed. The power of three days of different weather types is predicted. The existing single-stage network prediction results are compared and the results show that the two methods have their own advantages and disadvantages: the prediction results of the double-level neural network are highly correlated with the measured results, the correlation coefficient is 0.9530; the overall relative error is 8.22% 14.16% of the network, and has the advantage of not relying on the data measured the day before, the use of more flexible. However, the double-level neural network tocalculate the amount of radiation makes theworkload increased.

References

[1] Xu RD, Dai Y, Sun XY 2012 Temperature prediction of photovoltaic array based on BP neural network[J]. *Industry and Mine Automation* **7** 59-63.

[2] Chang YQ, Li YC, Lv Z, *et al*. 2007 Multivariable feedforward decoupling control based on double-level neural network in a fermentation process[J]. *Journal of Northeastern University (Natural Science)* **28** 925-28.

[3] He Y，Xiang LG 2004 Research and application of BP neural network model based on fuzzy clustering[J]. *Syst. Eng. Theory and Practice* 79-82.

[4] Cheng CS, Duan SX, Yin JJ 2009 Design of Photovoltaic Array Power Forecasting Model Based on Neutral Network[J].T. China Electrotechnical Soc. **24** 153-8.

[5] Zhang L, Zhang YX, Guo CM, *et al.* Photovoltaic system power forecasting based on neutral networks. *Electric Power* 2010 **43** 75-8.

[6] Cheng CS 2011 Research on Power Forecasting and Energy Management for Photovoltaic Microgrid [D]. Wuhan: Huazhong University of Science & Technology.

[7] Liao ZL, Ruan XB 2009 Non-linear engineering simplification model of silicon solar cells in arbitrary solar radiation and temperature [J]. *Acta Energiae Solaris Sinica* **30** 430-5.