Short-Term Photovoltaic Generation Forecasting Based on LVQ-PSO-BP Neural Network and Markov Chain Method

Xiran Wang1,*, Xue Ma2, Suhua Lou1, Fei Peng2 and Song Wu1

1State Key Laboratory of Advanced Electromagnetic Engineering and Technology (College of Electrical and Electronic Engineering, Huazhong University of Science and Technology), Wuhan 430074, Hubei Province, China
2State Grid Qinghai Electric Power Company Economic Research Institute, Xining 810000, Qinghai Province, China

*E-mail: 1131250221@qq.com

Abstract. With the rapid development of solar photovoltaic generation, the effective prediction of photovoltaic is of great significance to mitigate its impact on power system. According to the analysis of main factors which affect power output of photovoltaic system, a short-term power forecasting model based on back propagation(BP) neural network and LVQ-PSO-BP neural network and Markov chain method was established. The weather is clustered and distinguished by using learning vector quantization(LVQ) and the particle swarm optimization(PSO) is used to optimize BP neural network weights and thresholds, improving forecasting network training speed. Finally, daily predictive value is corrected by Markov chain method to improve short-term photovoltaic generation forecasting precision. The simulation results indicate that the proposed method can accelerate the speed of searching optiums, improving the classification accuracy of weather types and the precision of the photovoltaic generation output effectively.

1. Introduction

Photovoltaic output is affected by many meteorological factors, such as solar irradiance and atmospheric temperature. The uncertainty of photovoltaic output brings a series of problems to power balance and safe operation of power system. Therefore, the effective prediction of photovoltaic output is of great significance for realizing the combined operation of photovoltaic and traditional power sources.

Reference[1]-[2]establishes the short-term photovoltaic generation forecasting model without irradiation by adopting the back propagation(BP) neural network with humidity, temperature and other meteorological factors. However, the BP neural network is sensitive to the initial weight and is prone to converge on local optimum. Reference[3]-[4] establish the photovoltaic output model based on support vector machine(SVM)considering the factors affecting the output of photovoltaic. However, the selection of model parameters had a great influence on the prediction effect and the method only using a single model to predict different weather types will also reduced the prediction accuracy.

In order to solve these problems, a short-term photovoltaic generation forecasting method is proposed based on LVQ-PSO-BP neural network and error correction of Markov theory. First, the weather is clustered and distinguished by using learning vector quantization(LVQ) and then particle swarm optimization(PSO) is used to optimize back propagation(BP) neural network weights and...
thresholds, improving forecasting network training speed. Finally, the daily predictive value is corrected by Markov chain method to improve short-term photovoltaic generation forecasting precision.

2. Correlational analyses between photovoltaic output and meteorological factors

The photovoltaic output is related to meteorological factors such as solar radiation intensity, and temperature. If all the meteorological factors are taken as input variables of the prediction model, the complexity of the model will increase and the convergence speed will slow down. Therefore, in this paper we use the grey correlation coefficient method to figure out the meteorological factors which have great correlation with the output of photovoltaic output.

| VARIABLE | RA   | TEMP | HR   | SPD  | STP  | KT   | PRCP | P_solar |
|----------|------|------|------|------|------|------|------|---------|
| RA       | 1    | 0.765| 0.650| 0.684| 0.705| 0.715| 0.514| 0.720   |
| TEMP     | 0.764| 1    | 0.682| 0.692| 0.689| 0.612| 0.553| 0.709   |
| HR       | 0.645| 0.678| 1    | 0.752| 0.654| 0.544| 0.604| 0.590   |
| SPD      | 0.684| 0.693| 0.756| 1    | 0.657| 0.556| 0.618| 0.598   |
| STP      | 0.705| 0.690| 0.660| 0.657| 1    | 0.710| 0.470| 0.666   |
| KT       | 0.715| 0.613| 0.550| 0.556| 0.710| 1    | 0.407| 0.783   |
| PRCP     | 0.514| 0.554| 0.610| 0.618| 0.470| 0.407| 1    | 0.447   |
| P_solar  | 0.720| 0.709| 0.596| 0.598| 0.666| 0.783| 1    | 0.447   |

From the Table 1 we can see that the maximum correlation between output power and clear sky index of photovoltaic system is 0.783; the correlation with solar radiation intensity is 0.720, followed by temperature correlation, which is 0.709. Therefore, we select the key factors such as clear sky index, solar radiation intensity and temperature as input data for short-term prediction of photovoltaic output.

3. Forecasting model of photovoltaic generation

3.1. Weather clustering based on LVQ

The structure of LVQ neural network is shown in Figure 1. LVQ neural network consists of input layer, competitive layer and output layer. When a sample enters the LVQ network, the neurons in the competitive layer generate the winning neurons through the competition rule and the winning neurons have a state of 1 and the others are 0. The output neurons connected with the winning neurons are 1 and the others are 0. Each output neuron represents different classes, and the neuron with state 1 gives the classes of samples so as to achieve the purpose of sample classification.

![Figure 1. The structure of LVQ neural network.](image-url)
According to the classification criteria of weather types formulated by the China Meteorological Administration (CMA), the weather conditions can be divided into 33 different types [5]. In view of the fact that too many classifications will lead to few samples and complicated calculation process, this paper divides 33 weather types into four types: sunny, cloudy, overcast and rainy.

3.2. PSO-BP neural network

Although BP neural network has good ability to approximate non-linear mapping, it’s essentially a gradient descent method. BP neural network has a slow convergence speed and it’s prone to converge on local optimum. Therefore, this paper uses PSO algorithm to optimize BP neural network by training the network to get the initial weights and thresholds. On this basis, BP algorithm is used to further refine and optimize the network parameters until the optimal network parameters are obtained.

The process of PSO algorithm optimizing the BP neural network is as follows:

1) Design the structure of BP neural network;
2) Determine the number of particles \( N \), threshold \( c_1 \), \( c_2 \), inertia weight \( \omega \).
3) Optimize the network with PSO. The position vector of PSO includes the connection weight and threshold of BP network.
4) Take the optimized network weights and thresholds as the initial weights and thresholds of BP algorithm.
5) Implement BP neural network optimization until the network achieves the performance index. Save the connection weights and thresholds and carry out prediction simulation.

3.3. Error correction based on Markov chain

Markov theory combines the initial state of the system and predicts the state of the system at a certain time in the future through the transition probability matrix. It is suitable for tracking and predicting the random discrete time series. Considering that abrupt weather may affect the prediction results of photovoltaic power station output, this paper corrects the prediction values of BP neural network based on multi-order weighted Markov chain model, so as to ensure the prediction accuracy of abrupt weather. The detailed process of multi-order weighted Markov chain can refer to the reference [6].

4. Training and evaluation of model

4.1. LVQ clustering network training

Step 1: Set the input data and Initialization. Determine the input vector, learning efficiency \( \eta \) and iteration times. In this paper, the input data include clear sky index, solar radiation intensity, temperature and precipitation. Initialize the weights with smaller random values and normalize the weights and input vectors.

Step 2: Find the winning neuron \( j^* \). Calculate the scalar product of the input samples and the connecting weight, the largest of which can be served as the winning neuron \( j^* \). Step 3: update the weights. Adjust the weights according to whether the classification is correct or not.

If the classification is correct, the weight matrix moves in the direction of the input vector, the formula of which can be expressed as:

\[
\omega_j(t + 1) = \omega_j(t) + \eta(t) \left[ P - \omega_j(t) \right]
\]  

If the classification error occurs, the weight matrix moves away from the input vector, the formula of which can be expressed as:

\[
\omega_j(t + 1) = \omega_j(t) - \eta(t) \left[ P - \omega_j(t) \right]
\]  

Step 4: determine whether it convergences. Determine whether the number of iterations reaches the preset value. If the maximum number of iterations is not reached, then go back to step 3, otherwise the calculation will be completed.
After the training, using the trained network to identify the weather types of the day with undefined labels to determine the appropriate prediction model.

4.2. PSO-BP neural network training
BP neural network uses a three-layer network structure with a single hidden layer. The input layer is meteorological data with 18 neuron nodes, including daily clear sky index, minimum/maximum temperature and solar radiation intensity at 7:00-21:00. The output layer is the photovoltaic output power with 15 neuron nodes. And the optimal number of nodes in the hidden layer is determined to be 12 by the equation (3).

\[ p = \sqrt{m+n+a} \]  

where \( m \) is the number of neuron nodes in the output layer; \( n \) is the number of neuron nodes in the output layer; \( a \) is the constant in the range of \([1,10]\).

According to the structure of BP neural network, there are 12 neurons in the hidden layer and 15 neurons in the output layer, so the total number of connection weights \( W_{ij} \) are 216(18×12) and the number of weights \( W_{jk} \) are 180(12×15). In addition, there are 12 thresholds \( \theta_j \) and 15 thresholds \( \alpha_k \). Therefore, the space dimensions of initialization swarm are 423, which includes all the connection weights and thresholds of BP neural network.

In this paper, the absolute value of prediction error is used as the individual fitness function to train the error sum of samples, which can be expressed as in equation (4). Then search for the optimal particle by iteration calculation as the initial connection weight and threshold of BP neural network.

\[ F = g \left[ \sum_{k=1}^{n} \text{abs}(Z_k - Y_k) \right] \]  

where \( g \) is constant coefficient; \( n \) is the number of nodes in the output layer; \( Z_k \) is the expected output of the output node \( K \); \( Y_k \) is the forecasting output power of the output node \( K \).

PSO has the better local search ability with the smaller inertia weight and the better global search ability with the larger inertia weight.

4.3. Evaluation indicators for predicted performance evaluation indicators
In this paper, we use Mean Absolute Percent Error (MAPE) and Root Mean Square Error (RMSE) to evaluate the accuracy of photovoltaic power generation forecasting.

\[ \text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{P_{mi} - P_{mi}}{P_{mi}} \right| \times 100\% \]  

\[ \text{RMSE} = \frac{1}{P_m} \left( \frac{1}{N} \sum_{i=1}^{N} (P_{mi} - P_{mi})^2 \right)^{1/2} \]  

where \( P_{mi} \) is the forecasting output power at \( i \)-th time point; \( P_{mi} \) is the actual power at \( i \)-th time point of the photovoltaic power station; \( N \) is the total number of predicted samples.

5. Results and discussion

5.1. Weather clustering based on LVQ
In the case study, the historical meteorological data are provided from January 2014 to December 2014 in a certain area of Qinghai Province. Taking summer as an example, the actual effective data consist 92 days, of which there are 31 sunny days, 26 cloudy days, 15 overcast days and 20 rainy days respectively. We adopt the cross-validation method to carry out four clustering calculations with dividing training set and test set in proportion of 3:1.

The comparison results of weather clustering based on K-means algorithm and LVQ network are
shown in Table 2. We can see that the accuracy of weather clustering by K-means algorithm is lower than that of LVQ network and the sample data can easily deviate from other clustering centers when using K-means.

### Table 2. Comparison of clustering results of LVQ and K-means.

| Actual weather | K-means cluster | LVQ cluster |
|----------------|-----------------|-------------|
| Sunny          | 19, 5, 7, 0     | 31, 0, 0, 0 |
| Cloudy         | 9, 11, 6, 0     | 2, 24, 0, 0 |
| Overcast       | 0, 4, 8, 3      | 0, 0, 14, 1 |
| Rainy          | 0, 2, 1, 17     | 0, 1, 2, 17 |

5.2. Discussions of simulation results

Taking the photovoltaic power forecasting model in summer as an example, select the sub-models of corresponding weather types and input the meteorological data of the predicted day into the prediction system shown in Figure 1 to obtain the forecasting output power. In order to verify the effectiveness of the proposed method, we adopt the traditional BP neural network and PSO-BP neural network as comparison models. Table 3 lists the errors of three forecasting models.

### Table 3. Prediction errors of different models.

| Model         | Weather | BP   | PSO-BP | PSO-BP-Markov |
|---------------|---------|------|--------|---------------|
| MAPE, RMSE    | MAPE, RMSE | MAPE, RMSE | MAPE, RMSE |
| Sunny         | 0.0815, 0.1452 | 0.0463, 0.0948 | 0.0373, 0.0551 |
| Cloudy        | 0.109, 0.1746 | 0.0955, 0.1499 | 0.0873, 0.1045 |
| Overcast      | 0.3685, 0.3975 | 0.1247, 0.2459 | 0.1077, 0.1489 |
| Rainy         | 0.4476, 0.6771 | 0.1903, 0.2962 | 0.1279, 0.1781 |

Figure 2 shows the forecast results of typical weather types in summer. Comparing the prediction results of the three models, we can see that the prediction effect of model 3 is better than that of model 2 and model 1. The forecasting errors of the three models are relatively small but the forecasting accuracy in cloudy day decreases. In overcast and rainy days, the fluctuation of photovoltaic output makes it more difficult to predict. The forecasting output of model 1 deviates from the actual output seriously with the error increasing sharply. Compared with model 1, the forecasting result of model 2 is improved but still not ideal. However, the model 3 based on error correction of Markov chain is sensitive to the variation of error, which can accurately track the drop of photovoltaic output and further improve the forecasting accuracy of photovoltaic output.

(a) Forecasting results of sunny and cloudy day in summer.
6. Conclusion
This paper analyses the correlation between photovoltaic generation system and many meteorological factors and establishes a short-term photovoltaic output forecasting model based on LVQ-PSO-BP neural network and Markov chain error correction. We adopt the LVQ network to cluster all weather types according to season, which effectively improves the clustering accuracy of each weather type. And the PSO algorithm is used to optimize the initial weights and thresholds of BP neural network, which effectively overcomes the deficiencies of the traditional BP neural network of slow convergence speed and easily falling into local optimum. Moreover, we adopt the multi-order Markov chain correct the forecasting errors for the further improvement of the forecasting accuracy, making up for the shortcomings of the low accuracy prediction of traditional methods. Therefore, the short-term forecasting model proposed in this paper, which is based on LVQ-PSO-BP network and Markov chain error correction, has a better accuracy and applicability.

Acknowledgement
This paper is supported by State Grid Corporation Science and Technology Project (522830170009).

References
[1] Lorenz E, Hurka J, Heinemann D, et al. Irradiance forecasting for the power prediction of grid connected photovoltaic systems[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2009, 2(1):2-10
[2] Amrouche B, Pivert X L. Artificial neural network based daily local forecasting for global solar radiation[J]. Applied Energy, 2014, 130(5): 333-341
[3] Shi J, Lee W J, Liu Yongqian, et al. Forecasting power output of photovoltaic systems based on weather classification and support vector machines[J]. IEEE Transactions on Industry Applications, 2012, 48(3):1064-1069
[4] Gao K L, Liu J M and Xu R Z 2011 A hybrid security situation prediction model for information network based on support vector machine and particle swarm optimization[J]. Power System Technology 35(4):176-182
[5] WANG S X, ZHANG N. Short-term output forecast of photovoltaic based on a grey and neural network hybrid model[J]. Automation of Electric Power Systems, 2012, 36(19):37-41
[6] T Jin, D Changhong, Y Wei, et al. Ultra-short-term Photovoltaic Power Forecasting in Microgrid Based on Adaboost Clustering[J]. Automation of Electric Power Systems, 2017, 41(21): 33-39