2022 21st International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Hybrid Intelligent Machine Learning based Ultra-short Term Generation Prediction of Photovoltaic Systems

Yongguang Wang
State Grid Hulun Buir Power Supply Company
Hulun Buir City, Inner Mongolia, China
rruhanlin1@163.com

Zhimin Wo
State Grid Hulun Buir Power Supply Company
Hulun Buir City, Inner Mongolia, China
rruhanlin1@163.com

Yang Bai
State Grid Hulun Buir Power Supply Company
Hulun Buir City, Inner Mongolia, China
rruhanlin1@163.com

Chuncheng Cao
State Grid Hulun Buir Power Supply Company
Hulun Buir City, Inner Mongolia, China
rruhanlin1@163.com

Songtao Tian
State Grid Hulun Buir Power Supply Company
Hulun Buir City, Inner Mongolia, China
rruhanlin1@163.com

Xu Tai
State Grid Hulun Buir Power Supply Company
Hulun Buir City, Inner Mongolia, China
rruhanlin1@163.com

***Abstract*—This work developed an ultra short-term photovoltaic power prediction model based on hybrid intelligent technology. The proposed model adopts a series of data processing technologies, including input variable selection based on statistical analysis, attribute reduction based on principal component analysis (PCA) and feature subset division based on the K-means clustering algorithm, to obtain a more relevant and effective data as input information for prediction. The model uses an adaptive neural fuzzy inference system (ANFIS) to train and learn the input information to obtain the output prediction results. The particle swarm optimization (PSO) algorithm is adopted in the training process to optimize the ANFIS parameters to reduce the prediction error. The proposed solution is evaluated through simulation experiments and the numerical results demonstrate that it can achieve effective prediction accuracy and has good adaptability.**

Keywords- Hybrid intelligent machine learning; ultra-short term prediction; training process; photovoltaic systems

I. Introduction

Photovoltaic power prediction system is of great significance to the operation of power systems connected with a large number of PV. The power system is a complex dynamic system. It is the responsibility of the power grid to maintain the power balance between power generation, transmission and consumption. For the power system without PV, the power grid dispatching organization can formulate the power generation plan according to the daily load curve to meet the power demand of the next day. The output power of the photovoltaic field is fluctuating and intermittent. The large-scale access to photovoltaics leads to a great increase in the difficulty of making power generation plans. Photovoltaic brings great challenges to the dispatching and operation of the power system.

Photovoltaic power prediction is to establish a prediction model of the output power of the photovoltaic field based on the historical power, historical photovoltaic speed, terrain and landform, numerical weather forecast, operation status of photovoltaic units, etc. the data of photovoltaic speed, power, numerical weather forecast, etc. are used as the input of the model, and the future active power of the photovoltaic field is predicted in combination with the equipment status and operating conditions of the photovoltaic field units. The basic requirements of the power grid dispatching department for photovoltaic power prediction include two aspects: one is daily prediction, that is, the prediction from 0:00 to 24:00 of the next day, with a time resolution of 15 minutes. The second is real-time (ultra short-term) forecast, that is, the forecast for the next 15 minutes to four hours from the reporting time, with a time resolution of no less than 15 minutes. The daily forecast requires the grid-connected photovoltaic field to submit the photovoltaic active power forecast data and startup capacity of 96 time nodes every 15 minutes from 0:00 to 24:00 of the next day to the power grid dispatching agency before the specified time every day. The real-time (ultra short-term) forecast requires the grid-connected photovoltaic field to report the photovoltaic power forecast data and real-time photovoltaic speed and other meteorological data in the next 15 minutes to four hours every 15 minutes (e.g., [1]-[3]).

Photovoltaic power prediction methods are generally divided into physical methods and statistical methods. In recent years, the combination prediction method combined with artificial intelligence technology has attracted more and more attention. Through the mining of historical data information, an intelligent prediction model is established. For example, based on the adaptive neuro-fuzzy inference system (ANFIS) model for ultra short-term prediction, ANFIS has stronger self-learning ability, robustness and adaptability. In recent years, some processing methods for input data have been proposed and validated (e.g., [4]-[7]). These prediction methods have improved the prediction accuracy and qualification rate of ultra-short-term photovoltaic power to a certain extent, but they provide important support for the safe, economic and high-quality

2473-3636/22/$31.00 ©2022 IEEE
DOI 10.1109/DCABES57229.2022.00037

operation of the power system. The ultra-short-term prediction accuracy needs to be further improved. To this end, this work explores the hybrid intelligent machine learning based ultra-short term generation prediction solution for photovoltaic systems. The main technical contributions made in this work can be summarized as follows:

(1) The mixed prediction model is used to reasonably mine and analyze the data, seek the internal laws, and design an effective model structure; Make comprehensive use of the effective technologies of supervised learning and unsupervised learning, and make adaptive improvements to the existing related technologies according to specific practical problems.

(2) The principal component analysis method is used to extract the feature dimension of the original data and obtain the contribution rate of each component in turn. The contribution rate is used as the weighting coefficient of the clustering algorithm to further compress the historical samples. It can greatly reduce the training burden of neural networks and improve the generalization ability while retaining the original information.

(3) The fuzzy inference neural network is adopted, and the network is trained by classification. A network outputs the training result corresponding to the center of a cluster and the adjacent sample set.

The rest of the work is organized as follows: Section II firstly formulates the problem; Section III presents the proposed solution of hybrid intelligent machine learning based ultra-short term generation prediction of photovoltaic systems. Section IV carries out the simulation experiments and presents the numerical results. Finally, the conclusive remarks are given in Section V.

II. SYSTEM MODELS

In this paper, an ultra-short-term photovoltaic power prediction model based on hybrid intelligent technology is proposed to solve the challenge of ultra-short-term photovoltaic power generation. Based on the available raw data, the proposed model adopts a series of data processing technologies, including input variable selection based on statistical analysis, attribute reduction based on principal component analysis (PCA) technology and feature subset division based on the K-means clustering algorithm to obtain more relevant and effective condensed data as input information for prediction. The proposed hybrid model uses an adaptive neural fuzzy inference system (ANFIS) to train and learn the input information to obtain the output prediction results. In the training process, the particle swarm optimization (PSO) algorithm is used to optimize the ANFIS parameters to reduce the prediction error. The overall framework of the proposed model is illustrated in Figure 1.

The hybrid model applies the effective statistical method of preprocessing and mining multi-dimensional historical data. The unsupervised learning technology is used to mine historical data, and the feature dimension and sample size of data are reduced and optimized without prior knowledge.

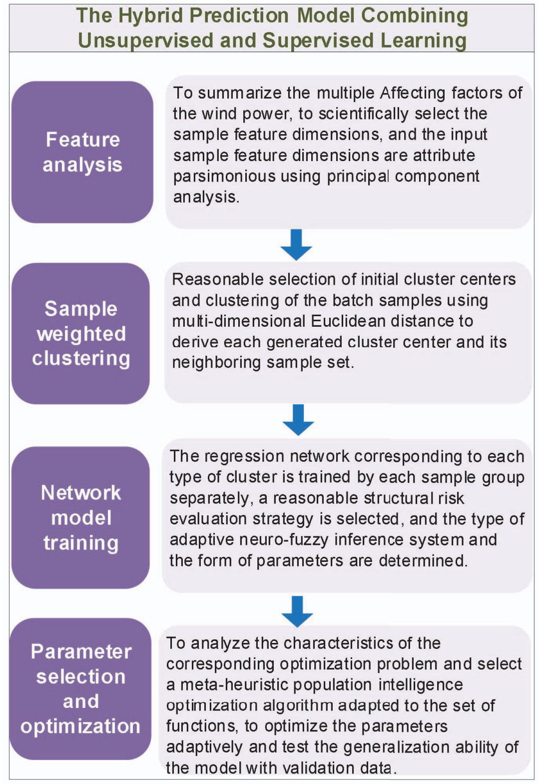


Figure 1. Flowchart of the proposed solution framework

Feature attribute reduction: the feature analysis of historical photovoltaic power data and weather factors shows that the meteorological factors affecting the output power of the photovoltaic field include photovoltaic speed, photovoltaic direction, air temperature, air pressure and humidity at different times. However, these attributes have inherent physical laws and show a certain correlation. To reduce the number of input variables (the number of feature dimensions) of the neural network to improve the calculation efficiency without affecting the prediction accuracy, it is necessary to reduce the redundant feature attributes reasonably. Principal components analysis (PCA) can transform the above-mentioned multiple meteorological variables into a few principal components (i.e., comprehensive variables). The principal components are not related to each other. These principal components can not only reflect most of the information of the original variables but also contain information that is not complementary and overlapping.

Sample subset Division: the sample data is taken from the record results of several years, covering the effects of different seasons, seasons, days and nights, etc. there are internal stable differences in the data samples. The clustering method (K-means) is used to cluster objects with high similarity into one class to further explore the internal rules of data samples. In the hybrid model of photovoltaic power prediction, the historical data are divided into several categories with high similarity to distinguish the climate factors in the large-scale time range. Based on the clustering results of the samples, the data far from the cluster center is eliminated to reduce the burden of small probability extreme cases on network training.

Adaptive neuro-fuzzy inference system: it uses the information of historical data mining to conduct supervised learning for different types of sample groups. Through classification training, the prediction network with higher accuracy is obtained, and different types of networks take into account their generalization ability. Adaptive neuro-fuzzy inference system (ANFIS) is a data-based modeling method. Its fuzzy membership function and fuzzy rules are obtained through supervised learning of a large number of historical data, rather than given based on human experience or intuition. Therefore, ANFIS takes into account the nonlinearity and adaptability of neural networks and the advantages of fuzzy inference systems in dealing with complex systems and is suitable for the prediction of multivariable nonlinear systems.

III. PROPOSED SOLUTION

A. Selection of Feature Input

Due to the complex operating environment of the photovoltaic field, many factors affect the accuracy of the photovoltaic power prediction in the photovoltaic field, among which the numerical weather forecast has the greatest impact. The numerical weather forecast forecasts meteorological data such as PV speed, PV direction, air temperature and air pressure, which is the basis and input for PV power prediction in the PV field. Whether an accurate numerical weather forecast can be obtained has a great impact on the accuracy of PV power prediction. However, due to the randomness and uncertainty of photovoltaics, as well as the fact that many photovoltaic fields in China are built in remote areas, the terrain difference is large, and there will be rapid changes in photovoltaic in a short time, which makes it difficult to predict the photovoltaic speed near the ground. We all know that the photovoltaic power is proportional to the third power of the photovoltaic speed, so the accuracy of the numerical prediction of the photovoltaic speed will directly affect the accuracy of the photovoltaic power prediction.

The original input vector is constructed according to the historical power record data and the public numerical weather forecast information, and the features included are power and photovoltaic speed statistics. For the current time $t+\tau$ is the power value at the time to be predicted, the selected input vector is as shown in formula (1):

$$x(t,\tau)=\left[P(t-1),P(t),v(t+\tau),\Delta v_p(t+\tau),\Delta v_b(t+\tau),V(t+\tau)\right] \tag{1}$$

$$\begin{cases}\Delta v_p(t+\tau)=v(t+\tau+1)-v(t+\tau)\\ \Delta v_b(t+\tau)=v(t+\tau)-v(t+\tau-1)\\ V(t+\tau)=\left[\operatorname{mean}(V_{t,\tau}),\min(V_{t,\tau}),\max(V_{t,\tau}),\operatorname{std}(V_{t,\tau})\right]\\ V_{t,\tau}=\left\{v(t+\tau-\Delta\tau),v(t+\tau-\Delta\tau+1),\dots v(t+\tau+\Delta\tau)\right\}\end{cases} \tag{2}$$

Where $P(t)$ and $v(t+\tau)$ are the average power value in the current period and the NWP PV speed value in the prediction period, respectively. Here, $\operatorname{mean}(\bullet),\min(\bullet),\max(\bullet),\operatorname{std}(\bullet)$ represent the mean, minimum, maximum and standard deviation of the time series, and are the radius of the range of the selected statistical series. For ultra short-term prediction, the time resolution is usually required to be 15 minutes to predict the photovoltaic power sequence in the next 4 hours, so the prediction step size $\tau=1,2,\dots,16$ is desirable.

B. Principal Component Analysis (PCA)

To reduce the number of input variables of the ANFIS network and reduce the computational complexity without affecting the prediction accuracy, the principal component analysis is applied to reduce the dimension of the original input vector to realize attribute reduction, and the characteristic variables in the above steps are converted into a few principal components (i.e., comprehensive variables). The principal components are uncorrelated, which can reflect most of the information of the original variables, and the information contained is not complementary and overlapping. The principal component analysis is a widely used feature selection method at present. It was first proposed by Pearson et al. It transforms the data set from the original space to the principal component space. Each principal component in the principal component space represents the effective new features after transformation. The covariance matrix or correlation matrix of the data set is calculated, the dimension of the principal component is determined by the eigenvalue of the matrix, and the direction of the principal component is determined by the eigenvector of the matrix. This method has a clear meaning and is easy to operate. In general, before the principal component analysis, the data are standardized and preprocessed to eliminate the differences between variables due to different dimensions.

C. K-means Clustering Partition Subset

When studying some attributes of things, we can cluster the objects with high similarity into one class by the clustering method to study and master the internal laws of things. When forecasting photovoltaic power, historical data can be divided into several categories with high similarity to distinguish climate factors in the large-scale time range. It realizes the subset division of the sample data and reduces the network training burden. When clustering data objects, it is necessary to evaluate the degree of difference between objects. Among them, the Euclidean distance function is the most commonly used measurement method. The clustering technology of the proposed model is the k-means algorithm.

D. Adaptive neuro-fuzzy inference system (ANFIS)

Adaptive neural fuzzy inference system (ANFIS) is a data-based modeling method. The fuzzy membership function and fuzzy rules of the system are obtained by learning a large amount of historical data, rather than given based on human experience or intuition. It takes into account the nonlinear and self-adaptability of neural networks and the advantages of fuzzy inference systems in dealing with complex systems and is suitable for the prediction of multivariable nonlinear systems. The historical sample data is divided into *k* subsets according to the clustering process. This historical sample data is divided into k groups according to the cluster center, and *K* ANFIS networks are trained with the sample data in *K* groups. Compared with a single network, ***k*** independent sub-networks can effectively reduce the local optimal solution, making it easier for the optimization algorithm to determine the optimal parameters of each network. The ANFIS network structure is illustrated in Fig. 2.

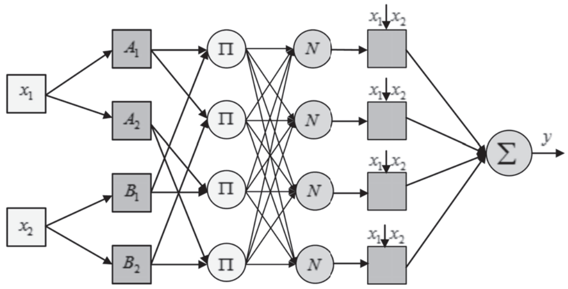
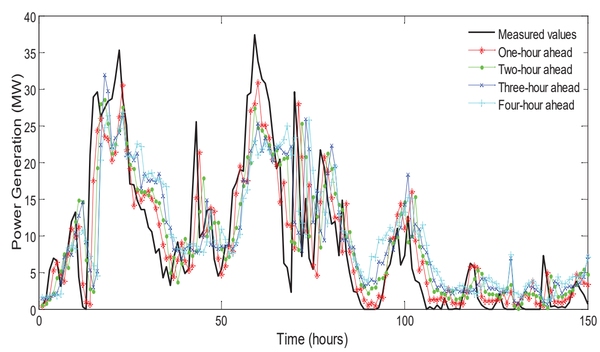


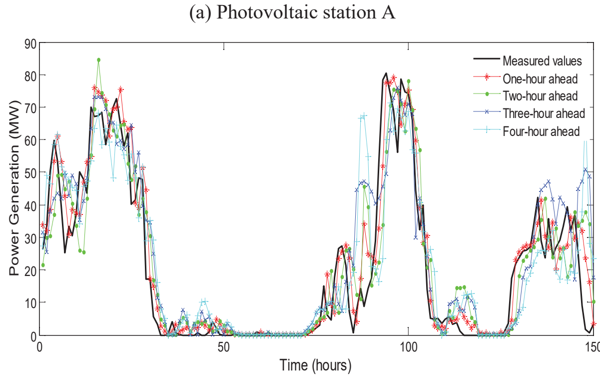
Figure 2. Typical ANFIS network structure

IV. SIMULATION EXPERIMENTS AND NUMERICAL RESULTS

This paper proposes an ultra short-term photovoltaic power prediction method based on hybrid intelligent technology. The historical power time series and the public numerical weather forecast (NWP) are used as the model input data, and the prediction network is obtained through the mining and training of the training sample data. For the real photovoltaic field, the test data is June 2022, and the training month data is from March to May. For ultra short-term prediction requirements, the time resolution is 15 minutes in the next 4 hours, so the optional prediction step size is 1-16. The statistical results of monthly root mean square error (RMSE) and monthly average percentage error (MAPE) of the two photovoltaic fields are shown in Table 1. Cut off part of the duration in the test month, and select the representative period prediction curve as shown in Figure 3.



(a) Photovoltaic station A



(b) Photovoltaic station A

Figure 3. Prediction curve of representative time slots

Table 1. Prediction errors of different steps

Steps	Photovoltaic station A		Photovoltaic station B	
	RMSE (%)	MAPE (%)	RMSE (%)	MAPE (%)
#1	4.78	2.84	4.04	2.34
#2	6.67	4.10	6.12	3.71
#3	8.04	5.13	7.39	4.47
#4	9.64	6.12	8.37	5.22
#5	10.44	6.87	9.86	6.28
#6	11.18	7.43	10.89	6.93
#7	11.32	7.82	11.55	7.34
#8	12.02	8.38	12.05	7.78
#9	12.17	8.68	12.50	8.02
#10	12.54	8.99	13.19	8.57
#11	12.80	9.21	13.55	8.85
#12	12.88	9.36	14.23	9.28
#13	13.06	9.55	14.27	9.39
#14	13.17	9.65	14.29	9.33
#15	13.24	9.70	14.54	9.68
#16	13.51	9.97	14.71	9.65

V. CONCLUSIVE REMARKS

This work proposed an ultra short-term power prediction model based on hybrid machine learning models. The proposed model adopts a series of data processing technologies, including input variable selection based on statistical analysis, attribute reduction based on principal component analysis (PCA) and feature subset division based on the K-means clustering algorithm to obtain more relevant and effective data as input information for prediction. The model uses an adaptive neural fuzzy inference system (ANFIS) to train and learn the input information to obtain the output prediction results. The numerical results show that the proposed solution can achieve good prediction performance with adaptability.

REFERENCES

[1] B. Wang, W. Cui, C. Wang and H. Xu, "Regional Distributed Photovoltaic Ultra-Short-Term Prediction Method Based on Multi-task Learning," 2021 IEEE 4th International Electrical and Energy Conference (CIEEC), 2021, pp. 1-5.

[2] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," in IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841-851, Jan. 2019.

[3] P. A. G. M. Amarasinghe and S. K. Abeygunawardane, "Application of Machine Learning Algorithms for Solar Power Forecasting in Sri Lanka," 2018 2nd International Conference On Electrical Engineering (EECon), 2018, pp. 87-92.

[4] D. A. Snegirev, S. A. Eroshenko, A. I. Khalyasmaa, V. V. Dubailova and A. I. Stepanova, "Day-ahead Solar Power Plant Forecasting Accuracy Improvement on the Hourly Basis," 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2019, pp. 1088-1091.

[5] F. Serttas, F. O. Hocaoglu and E. Akarslan, "Short Term Solar Power Generation Forecasting: A Novel Approach," 2018 International Conference on Photovoltaic Science and Technologies (PVCon), 2018, pp. 1-4.

[6] Y. Cai et al., "Ultra-short-term Photovoltaic Power Prediction Based on Elman Neural Network and Satellite Cloud Images," 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2), 2021, pp. 2149-2154.

[7] C. Liu, M. Li, Y. Yu, Z. Wu, H. Gong and F. Cheng, "A Review of Multitemporal and Multispatial Scales Photovoltaic Forecasting Methods," in IEEE Access, vol. 10, pp. 35073-35093, 2022.