Short-term Power Prediction Method of Photovoltaic Based on Output Clustering in Smart Grid

Xiaoyang Wang [1], Siming Chen [2], Yunlin Sun [3], and Sichou Chen [4, a]

[1] Beijing Institute of Technology, Zhuhai, Guangdong, China;

[2] HuaJu Testing Tech. Co., Ltd., Shunde District, Foshan, Guangdong, China;

[3] XiNing YueGuang Solar Energy Technology Co., Ltd, Foshan, Guangdong, China;

[4] Hailong Intelligent Technology Co., Ltd, Zhuhai, Guangdong, China.

[a] 971868890@qq.com

Abstract. Accurate photovoltaic power prediction can ensure the smart grid's safe and stable operation, as well as reasonable energy scheduling. Weather influences photovoltaic output, which is irregular and unstable, and photovoltaic output is similar under similar meteorological conditions. The paper proposes a solar power forecast model based on Mean-Shift clustering, support vector machine (SVM), and residual neural network (ResNet) in this regard. Firstly, the Mean-Shift algorithm is used to cluster similar days. Then, the SVM model is constructed to learn the similarity between the data of each meteorological type, and the similar day matching is performed on the forecast day. Finally, the short-term photovoltaic output prediction based on ResNet algorithm is carried out for the corresponding weather type. The suggested method's high forecast accuracy and stability are confirmed by experimental examination of a commercial photovoltaic power station.

Keywords: Photovoltaic power prediction; smart grid; Mean-Shift clustering; support vector machine; residual neural network.

1. Introduction

Clean energy for the smart grid can be produced using photovoltaic technology [1, 2]. The output power data exhibit significant randomness and uncertainty [3-5] due to the ease with which meteorology can influence the power generation of photovoltaic power plants. As a result, precise solar power forecasting can lessen the effect of access to renewable energy on the power system.

Photovoltaic power generation has a high degree of similarity when the climatic kinds are the same, and the power generation is strongly tied to different meteorological characteristics [6-11]. For various types of weather data, cluster analysis of historical data and selection of related day sample data can therefore effectively increase prediction accuracy. Reference [12] divided the historical dataset using improved Kmeans and then created a hybrid model using LSTM and convolutional neural network to increase prediction accuracy and adjust the estimated power value. The accuracy of the power prediction is improved in Reference [13] by using the Residual Network (ResNet) to forecast the daily average wind speed at sea. Skip links are employed by the residual blocks inside ResNet to address the gradient disappearance issue brought on by deep neural network depth [14].

Therefore, this paper proposes a photovoltaic power prediction method based on similar weather classification in order to combine the benefits of historical power and meteorological parameters to predict photovoltaic power, taking historical power generation data as one of the influencing factors.

The rest of this paper is organized as follows : Section 2 briefly describes the data sources and theoretical principles, including Mean-Shift clustering, SVM and ResNet, and proposes experimental methods. Section 3 gives the experimental results and discussion. Finally, the conclusion is given.

2. Materials and Methods

2.1 Data Source and Preprocessing

This study makes use of the actual dataset for photovoltaic output and weather from a photovoltaic power plant in Dagou, China, from April 2019 to March 2020. Temperature, humidity, wind speed, and radiation irradiance are all part of meteorological information. The daily time range for all data sampling is 07:30-18:00, with a total of 43 sampling points.

In order to solve the negative impact of different dimensions of different factors on the algorithm, the dataset is first standardized ; secondly, due to equipment limitations or data collection techniques, some individual data points have missing or abnormal values, which are filled in using the average value method.

2.2 Methodology

2.2.1 Photovoltaic power clustering based on Mean-Shift

Under similar weather types, the photovoltaic output has obvious regularity. Therefore, the Mean-Shift clustering technique is used to separate the photovoltaic power station's historical power dataset into multiple clusters, with each cluster having unique power characteristics and serving as the training set for the prediction model to adapt to varied weather types. The clustering process is shown in Fig. 1.
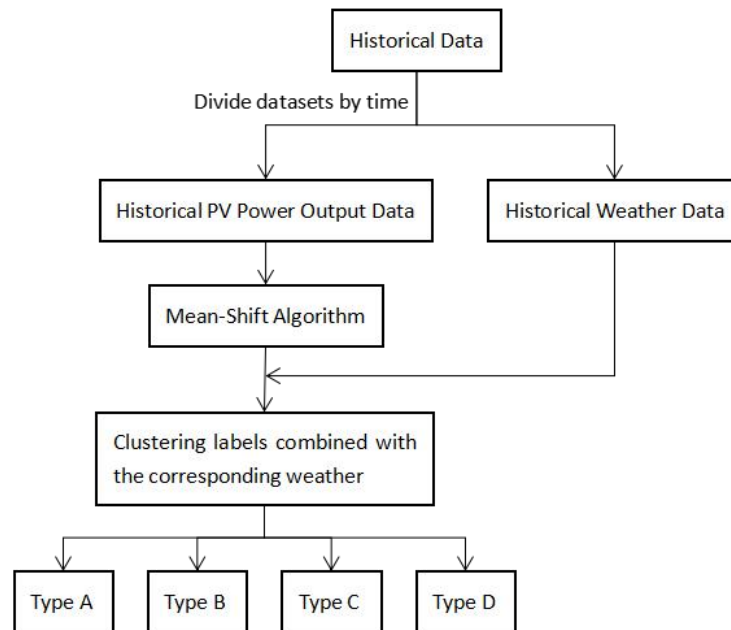

Fig. 1 Mean-Shift clustering process

Mean-Shift is a non-parametric clustering algorithm based on sliding window, which aims to find the dense area of different types of photovoltaic output [15, 16]. Compared with other clustering algorithms such as K-means, Mean-Shift does not need to set the number of clustering centers in advance.

Assuming that n sample points in D-dimensional space are x_i (i = 1, 2, ... n), the basic form of the Mean-Shift drift vector of any point x in the space is as follows :

$$M_r = \frac{1}{K}\sum_{x_i \in S_r}(x_i - x) \tag{1}$$

$$S_r(x) = \{y: \|y - x_i\|^2 < r^2\} \tag{2}$$

which y is a variable with in the spatial scaler; K is the number of n sample points distributed in the region S_r; S_r is a high-dimensional spherical region with x as the center and r as the radius.

In the process of clustering, the drift vector is iteratively calculated according to the following formula, and the clustering center is finally obtained:

$$\hat{x} = x + M_r(x) = \frac{1}{r}\Sigma_{x_i \in S_r} \, x_i \qquad (3)$$

In order to optimize the clustering effect, the Gaussian kernel function is introduced, and the appropriate r value (r is bandwidth) is selected to meet the requirements of the greater weight obtained by the closer distance between the two variables. Therefore, the Mean-Shift after the kernel function improvement is:

$$M_r = \frac{\left[K\left\{\frac{x_i-x}{r}\right\}*(x_i-x)\right]}{\Sigma_{x_i \in S_r}\left[K\left\{\frac{x_i-x}{r}\right\}\right]} \qquad (4)$$

$$K\left\{\frac{x_i-x}{r}\right\} = \frac{1}{\sqrt{2}r}e^{-\frac{(x_i-x)^2}{2r^2}} \qquad (5)$$

2.2.2 Weather type matching based on SVM

To adapt the solar output prediction algorithm to diverse weather kinds and increase photovoltaic output forecast accuracy. In the prediction of photovoltaic output, the SVM technique is used to match the types of similar days of photovoltaic output on the forecast day.

Support Vector Machines [17, 18] is a binary classification model for small sample, nonlinear and high-dimensional data. It converts multi-classification into several binary classifications to solve multi-classification (The number of categories is more than 2) problems. SVM maximizes the distance between two samples by finding the optimal decision hyperplane.

2.2.3 Photovoltaic output prediction based on ResNet

After training the photovoltaic output clustering and weather type matching models, this study employs a ResNet-based photovoltaic output prediction model for training and prediction to put the above Mean-Shift-based photovoltaic output clustering and SVM-based weather type matching methods to the test.

ResNet [19, 20] is a deep convolutional neural network constructed by introducing multiple block residual blocks. Its main feature is to obtain local features of data by local connection and sharing weights. ResNet can reduce information loss and effectively improve the convergence speed of the network.

2.2.4 Combination algorithm structure

In this paper, Mean-Shift clustering, SVM classification and ResNet combined prediction algorithm are used to predict short-term photovoltaic power generation. The specific process is shown in Fig. 2.
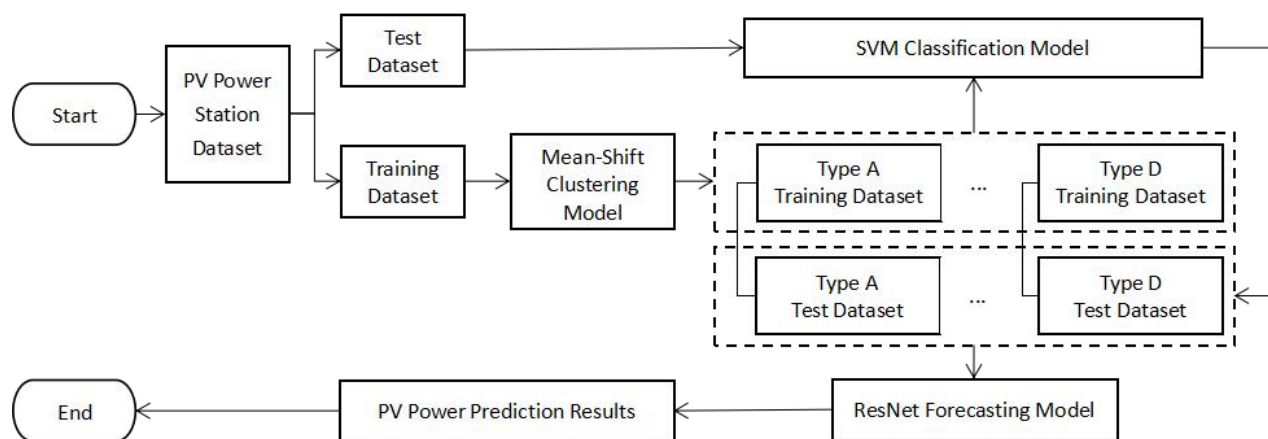
Fig. 2 Combination forecasting process

3. Results & Discussion

3.1 The Results of Clustering Analysis

The Mean-Shift clustering method is used to cluster the data at each time point. The bandwidth of the model is adaptively adjusted to 4 clusters, and the data and clustering center in January 2020 are visualized. The abscissa is 07:30-18:00, a total of 43 sampling time points. The ordinate is the actual active power of photovoltaic output, showing the distribution of photovoltaic output power at each time point this month.
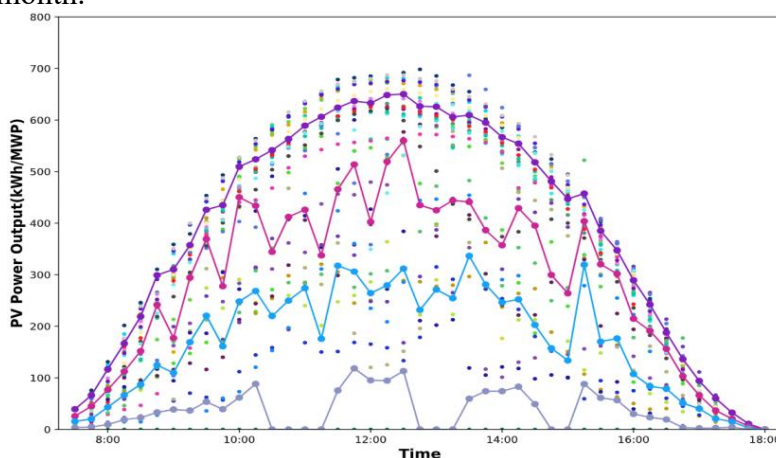

Fig. 3 Mean-Shift clustering result graph

The clustering results from Fig. 3 can be seen as four weather type A, B, C, and D. Type A to D represent the arrangement of weather type data from the most suitable for photovoltaic output to the most unsuitable for photovoltaic output.

3.2 Results and Analysis of Precision

In this study, the historical daily dataset was divided into training datasets for each time period according to 07:30-08:45, 09:00-10:45, 11:00-12:45, 13:00-14:45, 15:00-16:45, 17:00-18:00.

When preparing for SVM weather type matching training, the training dataset at each time period is used to eliminate the active power feature column of photovoltaic power generation, and the category label after Mean-Shift clustering is added to the last column to form a new sub-training dataset. Each sub-training dataset is input into the SVM algorithm to construct a weather type matching model for each time period. The test dataset is divided according to the historical data time period. After inputting the weather type matching model of each time period, the weather type

of the test dataset is matched according to the various meteorological characteristics of each time period, and the sub test dataset in the form of (Time period * Weather type) is divided.

The ResNet model adjusts the main parameters for optimization, sets the learning rate to 0.1, the number of iterations to 20, and the remaining parameters to the default value.In order to verify the prediction effect, this paper uses the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE) to evaluate the prediction results of the traditional regression algorithm and the combination algorithm of this study.

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N} \frac{\left|Y_{forecasting,i} - Y_{true,i}\right|}{Y_{true,i}} \times 100\% \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(Y_{forecasting,i} - Y_{true,i}\right)^2} \tag{7}$$

which N is the number of values to be tested; $Y_{forecasting, i}$ is the predicted value of photovoltaic outputand; $Y_{true, i}$ is the real value of photovoltaic output.

The time-phased prediction results are shown in Table 1, and the analysis of the overall forecast results for the day is shown in Table 2.
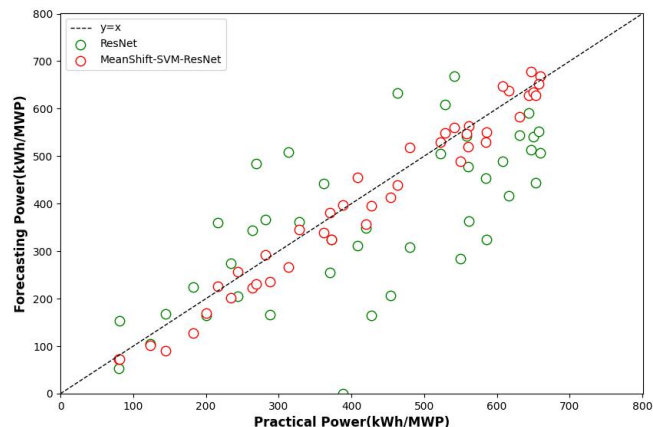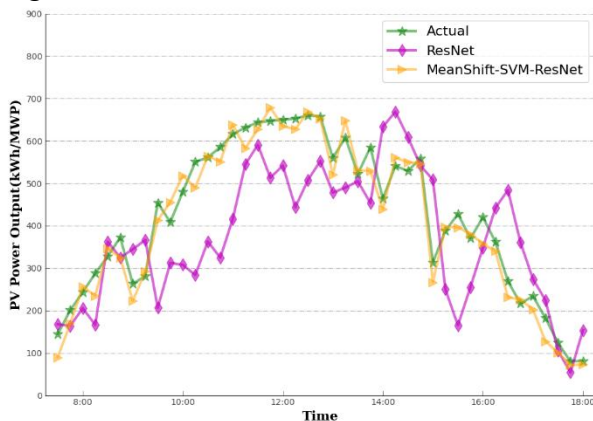
Table 1. Comparison of prediction results of time-phased

Time(C)	Index	ResNet	MeanShift-SVM-ResNet Model
7:00-8:45	MAPE(%)	19.24	15.88
	RMSE(W)	59.89	39.83
9:00-10:45	MAPE(%)	37.91	8.20
	RMSE(W)	191.11	38.87
11:00-12:45	MAPE(%)	20.35	3.35
	RMSE(W)	140.43	25.20
13:00-14:45	MAPE(%)	17.25	4.94
	RMSE(W)	105.97	31.34
15:00-16:45	MAPE(%)	54.99	8.34
	RMSE(W)	208.87	34.03
17:00-18:00	MAPE(%)	35.40	16.33
	RMSE(W)	44.02	30.37

Table 2. Comparison of overall prediction results

Index(C)	Resnet	MeanShift-SVM-ResNet Model
MAPE(%)	31.08	8.73
RMSE(W)	146.27	33.56

The prediction results of the above two methods and the actual values are visualized as shown in Fig. 4.

a.   Visual curve comparison   b. Forecast error

Fig. 4 Comparison of Actual and forecast power

From the comparative analysis of Table 1, Table 2 and Fig. 4, it can be seen that the prediction results of meteorological classification through Mean-Shift clustering are more stable than those using only ResNet algorithm, and by dividing the time period, it can adapt to the unpredictable weather of one day and improve the accuracy of the prediction model.

4. Summary

The example analysis shows that the prediction method proposed in this paper has the following advantages : The Mean-Shift algorithm is used to cluster the photovoltaic output, which indirectly and accurately realizes the clustering of meteorological characteristics. The clustering prediction after data division can better adapt to the impact of different meteorological conditions in different time periods and improve the prediction accuracy of short-term photovoltaic power generation. The prediction model established in this paper has obtained good prediction results and strong practicability. It has certain reference value in the research of short-term photovoltaic power prediction under smart grid.

Acknowledgment

References

[1] Li XC. (2022) Overview of grid-connected technology for large-scale photovoltaic power plants. Electric Engineering, 562: 61–64.

[2] Li X, Hu MH, Feng LD. (2023) Research on a new day-ahead photovoltaic power probability interval prediction method. Automation Application, 64: 71-74,78.

[3] Ji ZG, Li H, Ye L, et al. (2022) Short-term photovoltaic power prediction based on fluctuation characteristics mining. Journal of Solar Energy, 43: 146-155.

[4] Chen H. Risk assessment of distribution network operation considering line fault and randomness of photovoltaic output. Beijing Jiaotong University,2019.

[5] Che B, Liu YZ, Yu RY, et al. (2019) Distribution network risk assessment method considering randomness of distributed photovoltaic output. Renewable Energy Source, 37: 1685-1690.

[6] Han RB. Research on short-term load forecasting based on clustering analysis and feature learning.China University of Mining and Technology, 2022.

[7] Wei LB, Wang B, Wang Y, et al. (2022) Short-term power prediction of photovoltaic power generation based on meteorological similar day selection and lifting regression tree. Electron Device, 45: 183-188.

[8] Luo ZX. Research on photovoltaic prediction data preprocessing model based on clustering and similar day theory. North China Electric Power University, 2021.

[9] Gao XX. Photovoltaic output interval prediction based on similar day clustering. Dalian University of Technology, 2020.

[10] Zhang Q, Liu G, Wei XZ. Research on power prediction of photovoltaic power station based on similar hour and LM-BP neural network. 2021 International Conference on Power Grid System and Green Energy(PGSGE 2021)

[11] Hou SB, Wang KH, Shi KB, et al. (2018) Short-term output prediction of photovoltaic power generation system based on similar days and principal component analysis. Renewable Energy Source, 36: 15-21.

[12]]Chen ZX, Lin PJ, Chen SY, et al. (2021) Photovoltaic power prediction based on K-means ++ and hybrid deep learning. Electrical Engineering, 22: 7-13.

[13] Li S. (2021) Research on offshore wind power prediction and operation and maintenance strategy based on wind turbine state.

[14] Shafiq Muhammad,Gu Zhaoquan. (2022) Deep Residual Learning for Image Recognition: A Survey. Applied Sciences, 12: 8972-8972.

[15] Shi SQ, Zhang LP, Zhang BS, et al. (2022) Non-intrusive load identification method based on Mean-Shift clustering and siamese network. Electric drive, 52: 67-74.

[16] Chen JH, Wang L, Ju YT. (2021) Evaluation method of large-scale new energy impact on power grid line loss based on Mean-shift clustering. Renewable Energy Sources, 38: 59-64.

[17] Zhang J, Zhu XH, Huang XA, et al. (2019) A SVM photovoltaic output prediction model based on weighted Markov chain correction is proposed.. Power System Protection and Control, 47: 63-68.

[18] Guo JZ, Wang FZ, Zhang L. (2020) Short-term photovoltaic power generation prediction method based on support vector machine and deep belief network. Manufacturing Automation, 42: 29-33.

[19] Shen HL. Research on photovoltaic power generation fluctuation prediction based on deep learning. China Jiliang University, 2021.

[20] Wang Q, Ren YF, Wang L. (2022) Full convolutional network model detection of small photovoltaic hot spots. Journal of Testing Technology, 36: 199-205.