Journal of Computers Vol. 32 No. 2, 2021, pp. 74-85
doi:10.3966/199115992021043202007

Short-term Photovoltaic Power Prediction Based on IFCM and BA-Elman

Yun Wu[1], Qiang Wang[1*], Lishan Bao[2]

[1] School of Computer Science, Northeast Electric Power University, Jilin 132012, China
{838558160, 992665724}@qq.com

[2] Information and Communication Branch of State Grid Jiangsu Electric Power Co. Ltd., Nanjing 221000, China
138013800@qq.com

Received 5 April 2020; Revised 7 September 2020; Accepted 21 October 2020

Abstract. The traditional fuzzy C-means clustering algorithm (FCM) based on Euclidean distance is only applicable to clustering of spherical structures. When applied to text clustering of photovoltaic data to select similar days, it fails to take into account the difference in the importance of meteorological factors on photovoltaic power, resulting in a decrease in the accuracy and efficiency of data clustering. To solve the above problems, a short-term photovoltaic power prediction method based on improved fuzzy C-means clustering (IFCM) and bat algorithm optimization Elman neural network (BA-Elman) combination model was proposed. First, the improved fuzzy C-means clustering algorithm is used to select training samples with higher similarity to the forecast day, and then the Elman neural network prediction model optimized by bat algorithm trained by selected training samples. Finally, according to the actual data of a photovoltaic power station in Qinghai Province, a simulation experiment is carried out to verify the effectiveness of the method and model.

Keywords: fuzzy C-means clustering, bat algorithm, neural network, photovoltaic power generation, short term power prediction

1 Introduction

In recent years, with the adjustment of China's energy structure, solar energy, as a clean and renewable new energy, has developed rapidly, and it is gradually developing from an independent system to a large-scale grid connection [1-2]. However, the output power of photovoltaic power generation system has great volatility and randomness, and its large-scale photovoltaic grid connection will increase the instability of the power system and affect the safe operation of the power grid [3]. If the variation law of photovoltaic output can be predicted and timely fed back to the power dispatching department, the dispatching plan can be arranged in advance and the reserve capacity of the system can be reduced, thus improving the stability and economy of power grid operation.

Photovoltaic power prediction can be divided into indirect method and direct method [4]. Due to the need for the prediction of illumination amplitude, indirect prediction method needs high accuracy and relatively detailed meteorological information, which will lead to high prediction cost, so it is seldom used in China. The direct prediction method generally takes historical photovoltaic power generation data and meteorological data as the input of the model, and uses relevant theories and models for prediction. The prediction model has strong universality and low cost, which is the mainstream method for photovoltaic power generation power prediction at present [5].

Literature [6] calculated the correlation coefficient between historical day data through Euclidean distance, and defined the "near far" degree of photovoltaic power and day type, and finally used the

* Corresponding Author

sample set of similar days as the training sample of BP neural network model. Literature [7] first used FCM clustering algorithm for similar daily clustering, then established CG-DBN prediction model according to the clustering category, and finally used this model for short-term photovoltaic power prediction. The literature [8] uses the FCM algorithm to cluster the meteorological data of the similar day and the forecast day, and uses Elman neural network algorithm to establish a fuzzy clustering-Elman neural network prediction model to predict photovoltaic output. However, none of the above methods takes into account the difference in importance of different meteorological factors on photovoltaic power generation. It is not enough to reflect the essential characteristics of daily meteorological data to select similar days only by using correlation degree or Euclidean distance. The training samples selected are not accurate enough to affect the final photovoltaic power prediction accuracy.

Firstly, while some researchers also use clustering combination prediction method of clustering algorithm and Elman neural network to predict photovoltaic power, for the multi-attribute mixed data such as photovoltaic weather data, the traditional FCM algorithm can not measure the heterogeneity between samples and classes, and will ignore the influence of meteorological factors on photovoltaic power. Secondly, as the Elman neural network adopts the error back propagation algorithm (BP algorithm) to correct the weights, it is inevitable that there will be defects such as falling into the local optimal value and slow convergence rate [9]. Therefore, the key research point of this paper is how to improve the FCM algorithm to improve its clustering performance of photovoltaic meteorological data, select more accurate model training samples, and use bat algorithm to optimize the Elman neural network to easily fall into local optimal value, slow convergence speed and other defects. The main technical contributions of this paper are summarized as follows:

Firstly, the weight values of relevant factors of each meteorological feature are calculated. Combined with this weight value, a membership calculation formula combining Euclidean distance and covariance coefficient is proposed, and the performance of the improved algorithm is analyzed.

Bat algorithm is used to optimize the parameters of Elman neural network.

Simulation experiments are conducted based on real data. The experimental results show that the prediction accuracy of the proposed method is significantly improved compared with that of the traditional method.

The rest of the paper is structured as follows: Section 2 discusses the strengths and weaknesses of existing writings. Section 3 briefly describes the principle of FCM algorithm, calculates the weight of relevant factors of each meteorological feature, and gives the specific improvement steps and performance analysis of FCM algorithm. Section 4 gives the detailed steps of bat algorithm to optimize Elman neural network. Section 5 builds the experiment to carry on the simulation analysis. At last, the thesis is summarized and the future work is prospected.

2 Related Work

FCM algorithm gets the final clustering result according to the membership degree of each subclass of the sample, and proper modification of membership degree is conducive to improving the clustering performance of the algorithm.

Literature [10] used Markovnikov distance to replace Euclidean distance in FCM algorithm, and applied this algorithm to the clustering of high-dimensional texts, which improved the clustering effect of the algorithm. Literature [11] introduced fuzzy entropy constraint, improved FCM algorithm, resolved membership degree and clustering center formula, and improved the noise resistance of the algorithm.

The original design of FCM algorithm is to cluster the point sets in space. In practice, incomplete data and other forms of data, such as mixed data and interval data, need to be processed. The basic FCM algorithm cannot directly cluster these data.

For mixed data, literature [12] extended Euclidean distance and applied it to the clustering of mixed data to measure the heterogeneity between samples and classes, so that it can reflect the heterogeneity between objects and classes more accurately under the same framework and improve the clustering performance of the algorithm for mixed data. Literature [13] proposed a similarity measure for mixed data, which took data information into full consideration and reduced the impact of noise on clustering results. The algorithm proposed in literature [14] (hereinafter referred to as SFCM), based on the information entropy within and between classes, weighted different types of attributes of mixed data, took into account the difference in importance of different attributes, and improved the clustering effect of the algorithm. When FCM algorithm is applied to the clustering analysis of photovoltaic meteorological data, different meteorological factors have great differences in the influence on photovoltaic output power, which is a key factor affecting the clustering effect of the algorithm. Based on the above literature, the improved algorithm proposed in literature [14] is more suitable for the clustering analysis of photovoltaic meteorological data than other algorithms.

Although the algorithm proposed in literature [14] carries out weighted processing for different types of attributes of data, and the difference in importance of different attributes is considered, this algorithm still used Euclidean distance to calculate the sample membership degree. For the sample data with spherical space, good clustering effect can be obtained, while the weighted sample space is non-spherical, which results in the limited clustering performance of this algorithm in the photovoltaic meteorological data. In order to solve this problem, a membership calculation formula combining Euclidean distance and covariance coefficient is proposed after determining the weight coefficient of meteorological characteristic factors. The validity of the algorithm is proved by experiments.

3 Data Clustering Based on IFCM

3.1 Basic Principles of Fuzzy C-Means

The FCM algorithm is a data clustering method based on the optimization of the objective function [15-16]. The basic principle of FCM is as follows: set a data sample $X=(x_1, x_2, \cdots, x_n)$, and each sample has m indicator indicating its characteristics, that is $x_i=\{x_{i1}, x_{i2}, \cdots, x_{im}\}(i=1,2,\cdots,n)$. Among them, a subset of sample set X is the fuzzy cluster set $Z_1, Z_2, ..., Z_c$. The objective function of FCM is:

$$J_{FCM}\left(U,Z,X\right)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{b}\left(d_{ij}\right)^{2}, \tag{1}$$

where U is the membership matrix of each sample and Z is the cluster center matrix; X is the sample matrix; u_{ij} is the i th cluster center membership of the j th sample; b is a weighted index representing the data clustering ambiguity; d_{ij} is the Euclidean distance from the i th cluster center to the j th sample, and is defined as follows:

$$d_{ij}=\sqrt{\left(\sum_{e=1}^{m}(x_{je}-z_{ie})^{2}\right)}, \tag{2}$$

(1) Initialize u_{ij}. Randomly selects value b, at the same time, for each point x_i and each cluster Z_j, the membership degree u_{ij} takes a value between 0 and 1, and u_{ij} satisfies the following constraints:

$$\begin{cases}\sum_{i=1}^{c}u_{ij}=1 & 1\le j\le n\\ u_{ij}\in[0,1] & 1\le i\le c, 1\le j\le n\\ 0<\sum_{j=1}^{n}u_{ij}<n & 1\le i\le c\end{cases}, \tag{3}$$

(2) Calculate the centroid. For cluster Z_i, the corresponding centroid z_i is defined by:

$$z_i=\frac{\sum_{j=1}^{n}\left(u_{ij}\right)^{b}x_j}{\sum_{j=1}^{n}\left(u_{ij}\right)^{b}}, \tag{4}$$

(3) Update fuzzy pseudo-division.

$$u_{ij} = 1 / \sum_{p=1}^{c} \left(\frac{d_{ij}}{d_{pj}} \right)^{\frac{2}{b-1}} . \tag{5}$$

The FCM continuously calculates the centroid and fuzzy pseudo-division of each cluster until the termination condition is reached (the absolute value of all u_{ij} changes is less than the specified threshold φ).

3.2 Calculation of Weights of Factors Related to Meteorological Characteristics

When calculating the similarity of daily feature vectors, if the average weight is used, this average weight has a local similarity tendency. In the case of discrete measurement values, the point with the large point similarity measurement value determines the overall similarity, causing calculation errors [17]. In order to solve the local tendency caused by the average weight coefficient in the selection process of the training samples of the prediction model, the entropy weight method is used to calculate the weight coefficient in each meteorological factor in this paper.

The entropy weight method is one of the most widely used methods for objectively determining weights. Its basic principle is as follows: there are sample data for n historical days, each sample has m meteorological parameters, forming an $n \times m$ order data matrix $A = [a_{ie}]_{nm}$, where a_{ie} represents the value of the e th meteorological parameter in the i th historical days. The entropy of the e th meteorological parameter is:

$$E_e = \delta \cdot \sum\nolimits_{i=1}^{n} h_{ie} Inh_{ie}, e = 1, 2, \cdots, m , \tag{6}$$

where $\delta = -\frac{1}{Inn}, h_{ie} = \frac{a_{ie}}{\sum\nolimits_{i=1}^{n} a_{ie}}$.

When $h_{ie} = 0$, make $h_{ie} Inh_{ie} = 0$, then the weight of the e th meteorological parameter is:

$$\omega_e = \frac{1 - E_e}{\sum\nolimits_{e=1}^{m} (1 - E_e)} . \tag{7}$$

where $\omega_e \in [0, 1], \sum_{e=1}^{m} \omega_e = 1$.

In literature [18], by analyzing the correlation between photovoltaic power generation and meteorological factors [18] the main meteorological factors affecting photovoltaic power generation are atmospheric temperature, solar irradiance, relative humidity and wind speed. This paper takes the actual data of a photovoltaic power station in Qinghai Province as an example, and selects meteorological characteristics related factors: daily maximum temperature ($T_{\max}$), daily minimum temperature ($T_{\min}$), daily average temperature (T_{avg}), daily relative humidity (R_H), daily average wind speed (V), daily average Irradiance (I_{avg}) and the calculation results are shown in Table 1.

Table 1. Meteorological feature weight coefficient

Meteorological factors	$T_{\max}$	$T_{\min}$	T_{avg}	R_H	V	I_{avg}
Weighting coefficient	0.044	0.016	0.073	0.012	0.021	0.834

3.3 Improved Fuzzy C-means Algorithm

The FCM algorithm calculates the membership function value of the sample according to Euclidean distance. In the calculation process, the weight coefficient of each sample attribute is set to be the same, that is the data space is a spherical space. In this case, the FCM algorithm can get better clustering results. When selecting similar days for photovoltaic prediction models, the influence of various meteorological factors on photovoltaic power is different, so the data space may appear non-spherical. In view of the above problems, after determining the weight coefficients of the meteorological characteristics, formula u'_{ij} for calculating the membership degree combining Euclidean distance and covariance coefficient is proposed.

$$S_{ij} = \alpha d'_{ij} + \beta r'_{ij} , \tag{8}$$

$$d'_{ij} = \sqrt{[\sum_{e=1}^{m} \omega_e^2 (x_{je} - z_{ie})^2]} , \tag{9}$$

$$r'_{ij} = \frac{\sum_{e=1}^{m} \omega_e^2 (x_{je} - \overline{x_{je}})(z_{ie} - \overline{z_{ie}})}{[\sum_{e=1}^{m} \omega_e^2 (x_{je} - \overline{x_{je}})^2 \sum_{e=1}^{m} \omega_e^2 (z_{ie} - \overline{z_{ie}})^2]^{\frac{1}{2}}} , \tag{10}$$

$$u'_{ij} = 1 / \sum_{p=1}^{c} \left(\frac{S_{ij}}{S_{pj}} \right)^{\frac{2}{b-1}} . \tag{11}$$

where ω_e is the weight value of the e th feature of the sample, $e = (1, 2, \cdots, m)$; α and β are the weight coefficients of d'_{ij} and r'_{ij} , and $\alpha + \beta = 1$ (there, $\alpha = 0.5, \beta = 0.5$).

After clustering the samples, the category of the new samples can be identified by pattern recognition: select the new cluster center, and then calculate the membership degree between the new sample and each cluster center according to formula (11). The category with the largest membership is taken as the category of the new sample.

3.4 Evaluation of FCM Algorithm Before and After Improvement

In order to evaluate the effectiveness of the fuzzy clustering algorithm, the partition entropy coefficient PE, the partition coefficient PC, and the improved partition coefficient MPC were selected to compare and analyze the clustering effect of the FCM algorithm before and after improvement [19].

(1) The calculation formula of the divided entropy coefficient PE is as follows:

$$PE = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{C} u_{ij} \log(u_{ij}) , \tag{12}$$

where n is the number of cluster samples, C is the number of cluster centers, the range of PE values is $[0, \log(C)]$, the closer the value is to 0, the better the clustering effect is, and the closer the value is to $\log(C)$, the more blurred the clustering effect is.

(2) The calculation formula of the division coefficient PC is as follows:

$$PC = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{C} u_{ij}^2 , \tag{13}$$

where the value range of PC is $[1/C, 1]$, the closer the value is to 1, the better the clustering effect is, and the closer the value is to $1/C$, the more fuzzy the clustering effect is.

(3) Optimized the monotonic tendency of PC and PE by using the improved partition coefficient MPC. The calculation formula is as follows:

$$MPC = 1 - \frac{C}{C-1}(1 - PC) . \tag{14}$$

(4) Objective function value and number of iterations: The objective function value of the FCM algorithm can measure whether the cluster center can accurately represent the indicators of each

meteorological feature center. The smaller the objective function value, the better the clustering effect, and the less the number of iterations indicates that the algorithm is more efficient.

The parameters of the FCM algorithm are set as follows: fuzzy index $b=2$, iteration termination threshold $\varphi=1.0\times10^{-6}$, maximum number of iterations $D=100$, number of clustering centers $c=3$, FCM algorithm, the algorithm proposed in literature [14] and IFCM algorithm evaluation indicators, and the objective function values and iteration times are shown in Table 2 and Fig. 1.

Table 2. Comparison of evaluation indexes of FCM, the algorithm proposed in literature [14] and IFCM

Algorithm	PE	PC	MPC
FCM	0.7961	0.5401	0.6934
The algorithm proposed in literature [14]	0.7512	0.5821	0.7231
IFCM	0.7265	0.6214	0.7476

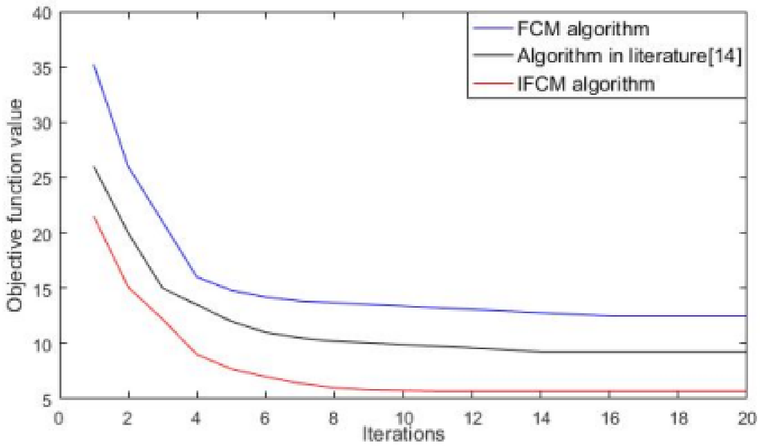

Fig. 1. The objective function value change curve of FCM, the algorithm proposed in literature [14] and the IFCM algorithm

As can be seen from Table 2, the indexes of both the algorithm proposed in literature [14] and the IFCM algorithm are superior to the traditional FCM algorithm, while the indexes of the algorithm proposed in literature [14] are also improved.

As can be seen from Fig. 1., both the algorithm proposed in literature [14] and the IFCM algorithm are superior to the traditional FCM algorithm in the number of iterations and the minimum value of the objective function. The number of iterations when the objective function value of IFCM algorithm reaches the minimum value (10) is less than that of the algorithm proposed in literature [14] reaches the minimum value (14). At the end of algorithm iteration, the minimum objective function of IFCM (5.7) is also smaller than that of the algorithm proposed in literature [14] (9.23). Comparing the objective function value and iteration number of the IFCM algorithm and the algorithm proposed in literature [14] shows that the IFCM algorithm can achieve a smaller objective function value in a smaller number of iterations than the algorithm proposed in literature [14] algorithm, and its convergence efficiency and clustering effect are both has seen an increase.

4 Photovoltaic Power Prediction Model

4.1 Elman Neural Network

Elman neural network is a local recursive internal delay feedback neural network. It contains input layer, hidden layer, receiving layer and output layer, has the characteristics of self-organization and self-learning, and its unique receiving layer adds hidden layer and output layer node feedback, enhancing the accuracy of network learning [20]. The Elman network structure is shown in Fig. 2.

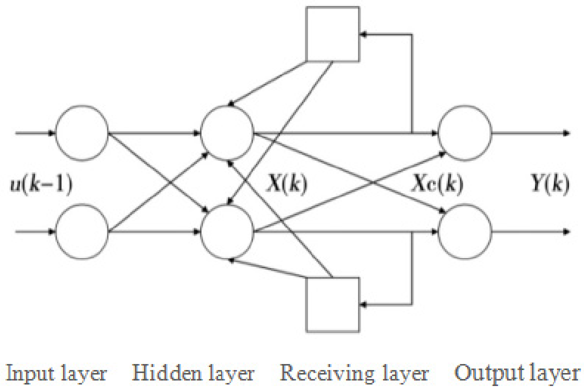

Fig. 2. Structure of Elman-NN

The mathematical model of Elman neural network is:

$$X(k)=F\left(W^1X_c(k)+W^2u(k-1)\right), \tag{15}$$

$$X_c(k)=\lambda X_c(k)+X(k-1), \tag{16}$$

$$Y(k)=g\left(W^3X(k)\right). \tag{17}$$

where In the formula: $Y(k)$-output vector; $g(*)$-output neuron activation function; λ-self-connected feedback gain, $X_c(k)$-receive unit output at time k; $X(k)$-hidden unit output at time k; W^1-receiving-hidden layer node weight matrix at time k, W^2-input-hidden layer node weight matrix at time k, W^3-hidden-output layer node weight matrix at time k; $F(*)$-sigmoid function; $u(k)$-input vector.

Elman neural network has strong robustness, objectivity and versatility, and it has good application value and prospect in the field of prediction containing feature quantities of nonlinear time series. When the Elman neural network is used to build the prediction model, the initial weights and thresholds are randomly generated, so the algorithm has a slow convergence rate, and it is easy to fall into a local minimum [21], which ultimately leads to low prediction accuracy of photovoltaic power. Therefore, it is necessary to optimize its initial weights and thresholds.

4.2 Basic Principles of the Bat Algorithm

The bat algorithm is a heuristic algorithm based on bat echolocation proposed by Yang in 2010 [22]. This algorithm simulates the changing law of pulse emission frequency and pulse loudness during bat foraging, using an adjustable frequency technique is used to enlarge the search range of solution space, and the balance between global search and local search is realized by automatic scaling.

Assume that the speed of a bat at time t is v_θ^t, its position is l_θ^t, and the best solution among all bats in the current population is l^*. According to the idea of echo localization, the speed and position update formula of the bat is as follows:

$$f_\theta=f_{\min}+\left(f_{\max}-f_{\min}\right)\mu, \tag{18}$$

$$v_\theta^t=v_\theta^{t-1}+\left(l_\theta^{t-1}-l^*\right)f_\theta, \tag{19}$$

$$l_\theta^t=l_\theta^{t-1}+v_\theta^t, \tag{20}$$

where $f_{\min}$ and $f_{\max}$ are the minimum and maximum frequencies of the pulse emission, respectively; $\mu\in[0,1]$ is a random vector subject to uniform distribution.

For local search, once a solution is selected from the existing optimal solutions, the next solution l_{new} of each bat is generated nearby in a random walk, that formula is:

$$l_{new} = l_{old} + \eta A^t , \quad \textbf{(21)}$$

where η is a random number in the interval $[0,1]$; $A^t = \langle A_\theta^t \rangle$ is the average loudness of all bats at this moment; l_{old} is any one of the current optimal solutions for each bat individual.

In order to control the global search ability and local search ability of the algorithm, the pulse emission loudness A_θ and the pulse emission rate R_θ are introduced during the iteration. When bats find food, A_θ decreases and R_θ increases. The expression formula is as follows:

$$A_\theta^{t+1} = \delta A_\theta^t , \quad \textbf{(22)}$$

$$R_\theta^{t+1} = R_\theta^0 \left[1 - \exp(-\gamma t)\right]. \quad \textbf{(23)}$$

where γ is the pulse emissivity increase coefficient, δ is the pulse loudness attenuation coefficient, all are constant; R_θ^0 is the maximum pulse emissivity, and R_θ^{t+1} is the pulse emissivity at time $t+1$.

4.3 BA Optimized Elman Neural Network

Step 1: Initialize the neural network structure. First, initialize the number of nodes in the input layer, the hidden layer, the receiving layer and the output layer, and then import the training data.

Step 2: Initialize the bat population. Initialize the population size N the position l_θ^t and the velocity $v_\theta^t \ (\theta = 1, 2, \cdots, N)$ of the θ th individual bat, the pulse emission loudness A_θ, the pulse emission rate R_θ, the pulse loudness attenuation coefficient γ, the maximum iteration times S and search accuracy ζ. The range of the pulse frequency is $f_{\min} \sim f_{\max}$.

Step 3: Confirm the fitness function $Fitness(\theta)$ of the bat algorithm:

$$Fitness(\theta) = \frac{1}{q} \sum_{\theta=1}^{q} \sum_{a=1}^{n'} \left(y'_{\theta,a} - y_{\theta,a} \right)^2 . \quad \textbf{(24)}$$

where q is the number of samples; $y'_{\theta,a}$ is the predicted value of the a th output node of the θ th sample; $y_{\theta,a}$ is the actual value corresponding to the a th output node of the θ th sample.

Step 4: Calculate the current optimal fitness F_{best} of the fitness function $Fitness(\theta)$ and the current optimal position l_{best}.

Step 5: Generate new solutions according to equations (18)~(20).

Step 6: Generate a random number $rand$. If $rand > R_\theta^t$, select an individual's position as the global optimal individual position l_{best}, and use formula (21) to generate a local individual near it, and calculate the fitness value F_{new} of the local individual.

Step 7: If F_{new} is better than F_{best} and $rand < A_\theta^t$ in Step 6, then set the solution as the current global optimal individual, record its fitness value, and use formulas (22) and (23) to decrease A_θ^t and increase R_θ^t.

Step 8: Determine whether the termination condition is reached, and output the result if the termination condition is met, otherwise return to Step 5 to continue the iteration.

Step 9: As parameters of Elman neural network, the weights and thresholds corresponding to the global optimal position of bats are taken, and a photovoltaic power prediction model will be establish.

5 Forecast Examples and Results Analysis

5.1 Forecast Example

Using historical data from a photovoltaic power plant in Qinghai Province from March 1, 2015 to May 31, 2015. According to historical data, the generation power is basically 0 from 19:00 to 8:00 the next day, so the sampling time is from 8 to at 19:00 (with a sampling interval of 15 minutes), May 15, 2015 (sunny day) and May 26, 2015 (rainy day) were selected as the forecast days. The meteorological parameters are shown in Table 3.

Table 3. Forecast daily weather parameters

Forecast day	maximum temperature	Lowest temperature	average temperature	Relative humidity	average wind speed	Average irradiance
2015.5.15	24℃	8℃	16℃	31.1%	4.5m/s	737W/m^2
2015.5.26	18℃	4℃	13℃	68.2%	6.4m/s	523W/m^2

First, the cluster sample is formed after normalization of the historical day meteorological factors outside the forecast day by formula (25) [23]. Then, IFCM was used to classify the cluster samples into three categories. Finally, the new sample classification identification is used to determine the category to which it belongs, which is the training sample of the prediction model.

$$Q' = \frac{Q - \min(Q)}{\max(Q) - \min(Q)}. \tag{25}$$

where Q' is the normalized data, Q is the original data.

Relevant parameter settings: the activation function of the Elman neural network uses the sigmoid function, the learning rate is 0.01, the number of input nodes is 4, the number of nodes in both hidden and receiving layers is 5 and the number of output nodes is 1. The sum number of the weights and thresholds is 4*5+5+5*1+1+5*5=56, of which the weight is 50 and the threshold is 6. So the bat individual dimension is 56, the range of each component in the individual is [-1, 1], the bat population size is 25, the pulse emission loudness is 0.3, the pulse emission rate is 0.5, the pulse loudness control coefficients are all 0.98, the echo frequency range $[f_{\max}, f_{\min}]$ is [0, 2], the maximum number of iterations is 500, and the expected error is 0.001.

Using all historical data through the algorithm proposed in literature [14] and Elman neural network combined forecast method (SFCM-Elman), improved FCM and Elman neural network combined prediction method (IFCM-Elman), and improved FCM and BA optimized Elman neural network combined prediction method (IFCM-BA-Elman) predicted the photovoltaic power generation on May 15, 2015 (sunny day) and May 26, 2015 (rainy day). The prediction results are shown in Fig. 3. and Fig. 4.

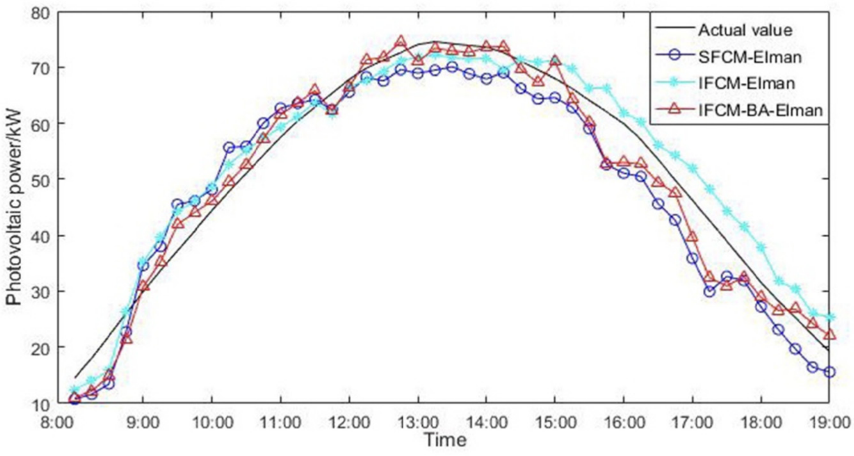

Fig. 3. PV power prediction sunny curve

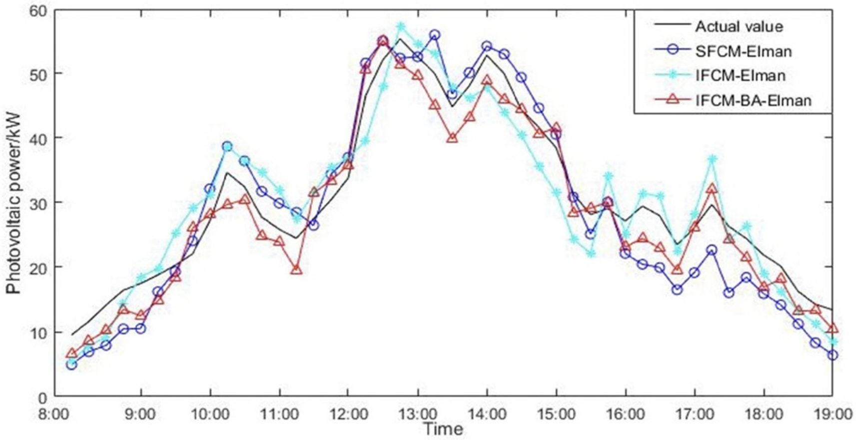

Fig. 4. PV power prediction rainy days curve

5.2 Error Analysis

This paper selects the absolute average error percentage (MAPE) and root mean square error (RMSE) as the error analysis indicators. MAPE is shown in formula (26) [24], and RMSE is shown in formula (27) [25]. The error comparison results are shown in Table 4 and Table 5.

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\frac{|S_i-Y_i|}{S_i}\times 100\%, \quad \textbf{(26)}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i-Y_i)^2}. \quad \textbf{(27)}$$

where S_i is the actual value of photovoltaic power, Y_i is the predicted value of photovoltaic power, and n is the number of predicted samples.

Table 4. Prediction error (2015.5.15)

Method of prediction	MAPE/%	RMSE
SFCM-Elman	13.01	5.41
IFCM-Elman	11.31	5.03
IFCM-BA-Elman	8.46	4.23

Table 5. Prediction error(2015.5.26)

Method of prediction	MAPE/%	RMSE
SFCM-Elman	18.79	5.32
IFCM-Elman	16.12	4.57
IFCM-BA-Elman	12.21	3.42

Fig. 3 to Fig. 4 and Table 4 to Table 5 show that:

(1) When the weather is better (sunny, for example), the absolute mean error percentage and root mean square error of IFCM-Elman method are reduced by 1.7% and 0.38 respectively compared with SFCM-Elman method; when the weather conditions are poor (such as cloudy, rainy days), the prediction accuracy of IFCM-Elman method is significantly higher than that of SFCM-Elman method, and its absolute mean error percentage and root mean square error are reduced by 2.67% and 0.75 respectively.

(2) The prediction effect of using IFCM algorithm to select similar day combined with optimized neural network compared with the prediction accuracy of the IFCM algorithm combined with a single neural network, is obviously improved. The absolute mean error percentage and root mean square error of the IFCM-BA-Elman method are reduced compared to the IFCM-Elman method 2.85% and 0.8

(sunny), 3.91% and 1.15 (rainy), respectively.

(3) On the basis of the improved FCM algorithm proposed in literature [14], this paper introduces the covariance coefficient and proposes a combination of Euclidean distance and covariance coefficient formulas of membership degree is applied to the photovoltaic meteorological data clustering. The performance comparison and photovoltaic power prediction experiments show that the proposed improved algorithm is effective and the accuracy of similar samples selected in practical application is higher.

6 Conclusion

Aiming at the problem of selecting similar days in short-term photovoltaic power prediction, this paper first calculates the influence weight of each meteorological factor on photovoltaic power generation by entropy weight method, and then puts forward a membership degree calculation formula combining European distance and covariance coefficient on the basis of traditional FCM clustering analysis, which reduces the number of training samples and their degree of difference. Based on the prediction of Elman neural network, the initial weights and thresholds are further optimized by bat algorithm, which makes the Elman neural network converge to a better solution more quickly, and further improves the efficiency and accuracy of photovoltaic power prediction model. Comparison of experimental results shows that compared with the prediction method of FCM algorithm combined with neural network, the proposed method and prediction model can effectively improve the prediction accuracy of short-term photovoltaic power. In the future, this method can be combined with data preprocessing to further improve the prediction accuracy and application scope.

Acknowledgements

This work was supported by the “Thirteenth Five-Year Plan” for scientific and technological research and planning of the Education Department of Jilin Province (JJKH20200121KJ).

References

[1] Y.-F. Gong, Z.-X. Lu, Y. Qiao, Q. Wang, An overview of photovoltaic energy system output forecasting technology, Automation of Electric Power Systems 40(4)(2016) 140-151.

[2] H. Li, F.-Q. Zhao, A. Jiao，L.-D. Chen, H.-H. Chen, L. Sun, A review of research on complementary generation of CSP and coal-fired units, Journal Of Northeast Electric Power University 39(6)(2019) 8-14.

[3] S.-Y. Li, Q.-Y. Xia, J.-D. Zhao, Q.-S. Tan, Analysis of inrush current characteristics of transformer in photovoltaic power generation system, Journal Of Northeast Electric Power University 40(2)(2020) 28-36.

[4] M. Ding, Z. Liu, R. Bi, W.-P. Zhu, Photovoltaic output prediction based on grey system correction-wavelet neural network, Power System Technology 39(9)(2015) 2438-2443.

[5] Y.-F. Wang, Y.-C. Fu, H. Xue, DMCS-WNN prediction method of photovoltaic power generation by considering solar radiation and chaotic feature extraction, Proceedings of the CSEE 39(Suppl. 1) (2019) 63-71.

[6] X.-L. Yuan, J.-H. Shi, J.-Y. Xu, Short-term power forecasting of photovoltaic generation considering weather type index, Proceedings of the CSEE 33(34)(2013) 57-64+S8.

[7] Z.-M. Li, Z.-L. Gao, C.-X. Liang, Short-term prediction of photovoltaic power based on combination of FCM and CG-DBN, Modern Electric Power 36(5)(2019) 62-67.

[8] J.-J. Zhang, Q. Zhang, Y. Ma, J.-H. Ma, J.-J. Ding, Fuzzy clustering-Elman neural network model for short-term photovoltaic power generation prediction, Electrical Measurement & Instrumentation 57(12)(2020) 46-51.

[9] W.-Q. Zhao, Y. Li, MPPT based on neural network optimized by intelligent water drop algorithm, Electric Power Automation Equipment 37(7)(2017) 8-12.

[10] C.-X. Li, Y.-J. Tan, J.-S. Kong, Applying mahalanobis distance-based text clustering algorithm in automatic paper marking system, Computer Applications and Software 32(4)(2015) 80-82+86.

[11] S.-Y. Liao, J.-F. Zhang, A.-Q. Liu, Fuzzy C means clustering algorithm by using fuzzy entropy Constraint, Journal of Chinese Computer Systems 35(2)(2014) 379-383.

[12] X.-X. Chang, Y.-Q. Zhang, A Partition-Based Clustering Algorithm for Mixed Data, Computer Applications and Software 31(6)(2014) 154-157.

[13] H.-J. Sun, G.-H. Shan, Y.-L. Gao, T. Yuan, Algorithm for clustering of high-dimensional data mixed with numeric and categorical attributes, Computer Engineering and Applications 51(8)(2015) 128-133.

[14] X.-W. Zhao, J.-Y. Liang, An Attribute Weighted Clustering Algorithm for Mixed Data Based on Information Entropy, Journal of Computer Research Development 53(5)(2016) 1018-1028.

[15] D.-F. Yang, Q.-K. Ma, Y. Wang, M.-J. Peng, Optimal Configuration of Stand-Alone Microgrid Capacity Considering Volatility of Photovoltaic Output, Journal of Northeast Electric Power University 39(4)(2019) 19-26.

[16] J. Cao, L.-J. Zhang, L. Hou, Z.-H. Chen, H. Zhang, Traffic Status Recognition Based on Information Entropy Weighted FCM, Computer Applications and software 35(10)(2018) 68-73.

[17] Q. Sun, J.-G. Yao, J. Zhao, M. Jin, L.-F. Mao, T. Mao, Short-term Bus Load Integrated Forecasting Based on Selecting Optimal Intersection Similar Days, Proceedings of the CSEE 33(4)(2013) 126-134+S16.

[18] Q. Dai, S.-X. Duan, T. Cai, C.-S. Chen, Z.-H. Chen, C. Qiu, Short-term PV Generation System Forecasting Model Without Irradiation Based on Weather Type Clustering, Proceedings of the CSEE 31(34)(2011) 28-35.

[19] R.-T. Yue, N. Han, Y.-F. Zhao, Sub-health State Classification Method of Area Control Position Based on Entropy-FCM Algorithm, Safety and Environmental Engineering 27(1)(2020) 204-208.

[20] W. Zhang, D.-J. Mao, X.-Y. Dai, MPSO-Elman Wind Power Prediction Model based on Entropy Association Data Mining, Journal of Engineering for Thermal Energy and Power 34(6)(2019) 165-171.

[21] G.-Q. Mei, The Research on Improved Elman Neural Network and Parameter Optimization Algorithm, [dissertation] Southwest University 2017.

[22] Y. Wu, J.-W. Lei, L.-S. Bao, C.-Z. Li, Short-term load forecasting based on improved grey relational analysis and neural network optimized by Bat Algorithm, Automation of Electric Power Systems 42(20)(2018) 67-72.

[23] M. Yang, B.-Z. Dai, L. Liu, A review of wind power probabilistic prediction, Journal Of Northeast Electric Power University 40(2)(2020) 1-6.

[24] M. Yang, H.-Y. Liu, Y. Sun, B.-J. Li, A study on real-time prediction of wind power based on atomic sparse decomposition and support vector machine, Journal Of Northeast Electric Power University 40(3)(2020) 1-7.

[25] M. Ding, C. Zhang, B. Wang, R. Bi, L.-Y. Miao, J.-F. Che, Short-term forecasting and error correction of wind power based on power fluctuation Process, Automation of Electric Power Systems 43(3)(2019) 2-9.