Short-term wind power prediction based on nutrosophic clustering and GA-ELM

Zijun Ye1, Yunfei Ding1
1 Shanghai Dianji University, Shanghai, 201306, China
*Corresponding author’s e-mail: yezi66jiayou@163.com

Abstract. Wind farm NWP data has regularity and difference, and making full use of the information contained in NWP data is the key to wind power prediction. A short-term forecasting method for wind power based on nutrosophic clustering and GA-ELM is proposed. Firstly, the wind farm NWP data is divided into several weather types by the Chinese wisdom clustering method, and then the GA-ELM model is established for different weather types. The Gaussian index method is used to classify the forecast data, and then the different types of forecast data are substituted into the corresponding model predictions. Taking a 14MW wind farm in Northeast China as an example, the experiment shows that the nutrosophic clustering method reduces the influence of boundary points and abnormal points on the clustering center. Compared with the traditional method, the method has higher precision and universality.

1. Introduction
The short-term prediction of wind power depends mostly on the numerical weather information of the wind farm (Numerical Weather Prediction, NWP) [1]. NWP data has certain regularity and difference. NWP data with similar weather and seasons have obvious regularity. In contrast, NWP data in different weathers, different seasons, and extreme weather have significant differences. In order to make full use of the information contained in NWP data to build a more accurate prediction model, it is necessary to perform cluster analysis on NWP data.

Wind power prediction based on cluster analysis is the focus of current scholars' research. The fuzzy C-means (FCM) method is suitable for cluster analysis due to theoretical perfection. The FCM algorithm and its optimization algorithm have been widely used in the field of wind power prediction. However, although the FCM algorithm achieves soft clustering of data, it ignores the influence of boundary data and abnormal data on the category center. If the data points are hardly divided near the edges of multiple clusters, a certain deviation will occur in each cluster center. In order to solve this problem, in 2015, Yanhui Guo proposed the neutrosophic C-means Clustering Algorithm (NCM) algorithm [3], and discussed the membership of the FCM algorithm at category edge data points and abnormal data points.

Inspired by the clustering idea, this paper intends to use the advantages of neutrosophic theory in data clustering, and proposes a short-term wind power prediction method based on neutrosophic clustering and GA-ELM to improve the accuracy and stable performance of short-term wind power prediction model.

2. Basic theory
2.1. FCM algorithm
The FCM algorithm uses membership to describe the degree of membership of each category. The basic idea of the algorithm is: Divide the data set $X = \{x_1, x_2, \ldots, x_N\}$ to be clustered into $c$ categories, where $x_1, x_2, \ldots, x_N$ is the sample data to be clustered and $N$ is the number of samples, $2 \leq c \leq N$. Suppose the cluster center is $V = (v_1, v_2, \ldots, v_c)^T$, the calculation formula of the objective function $J$ is:

$$
\min J(X, U, v_1, v_2, \ldots, v_c) = \sum_{k=1}^{c} \sum_{i=1}^{N} u_{ki}^m d_{ki}^2 
$$

In the formula, $u_{ki}$ is the degree of membership of the $i$th sample belonging to the $k$th category; $U$ is the membership matrix of $u_{ki}$; $v_k$ is the $k$th cluster center; $d_{ki}$ is the Euclidean distance between $v_k$ and $x_i$, i.e. $d_{ki} = \|v_k - x_i\|$; $m_0$ is the fuzzy parameter, generally $m_0 = 2$.

The constraints of the objective function are as follows:

$$
\begin{align*}
\sum_{k=1}^{c} u_{ki} &= 1, 1 \leq i \leq N \\
0 &\leq u_{ki} \leq 1, 1 \leq k \leq c, 1 \leq i \leq N \\
\sum_{i=1}^{N} u_{ki} &= [0, N], 1 \leq k \leq c
\end{align*}
$$

Using the Lagrange method to find the optimal solution of the objective function, we get:

$$
\begin{align*}
\hat{u}_{ki} &= \frac{1}{\sum_{j=1}^{c} \left( \frac{d_{kj}}{d_{ki}} \right)^{\frac{1}{m-1}}} \\
v_k &= \frac{\sum_{i=1}^{N} (u_{ki})^m x_i}{\sum_{i=1}^{N} (u_{ki})^m}
\end{align*}
$$

In the formula, $\lambda$ is the Lagrange factor.

2.2. NCM algorithm
The objective function of the NCM algorithm is as follows:

$$
J(T, I, F, C) = \sum_{i=1}^{N} \sum_{j=1}^{C} \left( w_{ij} T_{ij} \right) \left[ \|v_c - x_i\|^2 + \sum_{j=1}^{C} \left( w_{ij} T_{ij} \right) \|v_c - \bar{x}_{ij}\|^2 + \delta \sum_{j=1}^{C} \left( w_{ij} T_{ij} \right) \right]^2
$$

In equation (5), $m$ is a constant; $N$ is the sample size; $C$ is the category number; $T_{ij}$ represents the degree of membership of class $j$; $I_i$ represents the membership of uncertain points degree; $F_i$ represents the membership of outliers or abnormal points; $0 < T_{ij}, I_i, F_i < 1$; $w_1, w_2, w_3$ respectively represent the weight coefficients of $T_{ij}, I_i, F_i$; $c_j$ represents the category center; $c_{j,\max}$ represents the average value of $c_{pi}, c_{qi}, p_i$ represents the category with the largest $T_{ij}$, and $q_i$ represents the category with the second largest value of $T_{ij}$; $\delta$ represents the regularization factor.

$$
c_{j,\max} = \frac{c_{pi} + c_{qi}}{2}
$$

$$
p_i = \arg \max(T_{ij}), j = 1, 2, \ldots, C
$$

$$
q_i = \arg \max(T_{ij}), j \neq p_i \cap j = 1, 2, \ldots, C
$$
\[
\sum_{j=1}^{C} T_{ij} + I_{i} + F_{i} = 1 \quad (9)
\]

Find the extreme value of the NCM objective function by the Lagrangian method:

\[
T_{ij} = \frac{w_{i}w_{j}(x_{i} - c_{j})^{-\frac{2}{m+1}}}{\sum_{j=1}^{C} (x_{i} - c_{j})^{-\frac{2}{m+1}} + (x_{i} - c_{\text{max}})^{-\frac{2}{m+1}} + \phi^{-\frac{2}{m+1}}} \quad (10)
\]

\[
I_{i} = \frac{w_{i}w_{i}(x_{i} - c_{\text{max}})^{-\frac{2}{m+1}}}{\sum_{j=1}^{C} (x_{i} - c_{j})^{-\frac{2}{m+1}} + (x_{i} - c_{\text{max}})^{-\frac{2}{m+1}} + \phi^{-\frac{2}{m+1}}} \quad (11)
\]

\[
F_{i} = \frac{w_{i}(\phi)^{-\frac{2}{m+1}}}{\sum_{j=1}^{C} (x_{i} - c_{j})^{-\frac{2}{m+1}} + (x_{i} - c_{\text{max}})^{-\frac{2}{m+1}} + \phi^{-\frac{2}{m+1}}} \quad (12)
\]

\[
e_{j} = \frac{\sum_{i=1}^{N} (w_{i}T_{ij})^{\alpha_{i}}x_{i}}{\sum_{i=1}^{N} (w_{i}T_{ij})^{\alpha_{i}}} \quad (13)
\]

The specific steps of the NCM algorithm are as follows:

1. Initialize various membership degrees \( T_{ij}^{(0)}, I_{i}^{(0)}, F_{i}^{(0)} \);  
2. Initialization parameters \( C, m, \delta, \varepsilon, w_{i}, w_{j} \);  
3. Calculate the category center \( C_{j}^{(k)} \) in step \( k \) by formula (13);  
4. Calculate \( c_{\text{max}} \);  
5. Update \( T_{ij}^{(k)} \) to \( T_{ij}^{(k+1)} \) by using formula (10), using formula (11) to update \( I_{i}^{(k)} \) to \( I_{i}^{(k+1)} \), and update \( F_{i}^{(k)} \) to \( F_{i}^{(k+1)} \) by equation (12);  
6. If \( |T_{ij}^{(k+1)} - T_{ij}^{(k)}| < \varepsilon \), then stop the iteration; otherwise, turn to 3;  
7. Compare the size of each membership degree in \( [T_{ij}, I_{i}, F_{i}] \), and categorized by maximum membership.

### 2.3 GA-ELM algorithm

#### 2.3.1. ELM algorithm

The steps of the ELM model\(^{4}\): Suppose the sample data set \((x_{j}, t_{j}) \in \mathbb{R}^{n} \times \mathbb{R}^{m}\), where \( x_{j} = [x_{j1}, x_{j2}, ..., x_{jm}]^{T}, t_{j} = [t_{j1}, t_{j2}, ..., t_{jm}]^{T}\). The output function of the network is shown in equation (14).

\[
y(x_{j}) = \sum_{i=1}^{L} \beta_{i}g(w_{i} \cdot x_{j} + b_{i}) = o_{j}, \quad j = 1, ..., N \quad (14)
\]

If the model approaches the training sample with zero error, \( \sum_{i=1}^{L} \|o_{i} - t_{i}\| = 0 \), there are parameter \( \beta_{i}, \omega_{i}, b_{i} \) such that:

\[
\sum_{i=1}^{L} \beta_{i}g(w_{i} \cdot x_{j} + b_{i}) = t_{j}, \quad j = 1, ..., N \quad (15)
\]

The matrix expression of equation (15) is:

\[
H \beta = T \quad (16)
\]

And:

\[
H = \begin{bmatrix}
g(w_{1} \cdot x_{1} + b_{1}) & \cdots & g(w_{L} \cdot x_{1} + b_{L}) \\
\vdots & \ddots & \vdots \\
g(w_{1} \cdot x_{N} + b_{1}) & \cdots & g(w_{L} \cdot x_{N} + b_{L})
\end{bmatrix}_{N \times L}, \quad \beta = \begin{bmatrix}
\beta_{1}^{T} \\
\vdots \\
\beta_{L}^{T}
\end{bmatrix}_{L \times \omega}, \quad T = \begin{bmatrix}
t_{1}^{T} \\
\vdots \\
t_{N}^{T}
\end{bmatrix}_{N \times \omega} \quad (17)
\]
In order to improve the stability and generalization ability of the model, the regularization coefficient $C$ is introduced, and the regularized least square solution of $\beta$ is obtained, namely

$$\hat{\beta} = H^T \left( I/C + HH^T \right)^{-1} T$$  \hspace{1cm} (18)

Therefore, the output function of the ELM model is

$$y(x) = h(x) \hat{\beta} = H \hat{\beta}$$  \hspace{1cm} (19)

2.3.2. GA-ELM algorithm

The algorithm flowchart is shown in Figure 1, and the specific steps are as follows:

1. Randomly generate ELM input parameters, encode these parameters including input weights and thresholds, and then generate initial population. The number of population is the number of samples, and the dimension of the population is the number of parameters that need to be optimized.

$$\theta = [w_{11}, w_{12}, \ldots, w_{n1}, w_{12}, \ldots, w_{n2}, \ldots, w_{nm}, b_{1}, \ldots, b_{m}]$$  \hspace{1cm} (20)

   Where $\theta$ is a certain body in the population; $w_{ij}$ and $b_{j}$ are random numbers initialized to $[-1, 1]$.

2. Decode to obtain the weights and thresholds, and train and predict the ELM network with the weights and thresholds respectively, and use the mean square error of the network as the objective function $V$.

$$V = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$  \hspace{1cm} (21)

   Where $n$ is the number of test samples; $y_i$ is the predicted value; $\hat{y}_i$ is the actual value.

3. Set fitness function and evolution times. The ranking function $[5]$ is used to determine the fitness function, namely $V_{fit} = ranking(V)$.

4. Solve the local optimal fitness value $V_{fit}$, solve each individual’s $V_{fit}$ successively, and then select the local optimal individual.

5. Solve the global optimal fitness value $V_{fit}$. After solving the local optimal fitness value in each round, use crossover and mutation to evolve individuals with better fitness values and recalculate the fitness function. Insert the individual of the child population into the parent population, and replace the individual with the smallest fitness value in the parent population to obtain the new population. Repeated evolution, each time evolution, the number of evolution $m$ increases by one, if $m$ is less than the total number of evolution $N$, then return (4); If $m$ is equal to the total number of evolution $N$, then end the cycle, the resulting $V_{fit}$ is the global optimal adaptation value.

6. The individual individuals of the population corresponding to the global optimal adaptation value are decoded to obtain optimized input weights and thresholds, and the optimized parameters are used to establish the GA-ELM model.

3. Prediction model based on Neutrosophic clustering and GA-KELM

3.1. Training sample clustering

Using NCM method to perform cluster analysis on NWP data, the cluster number $c$ and membership matrix $[T_{ij}, I_{ij}, E_{ij}]$ ($i \in [1,N], j \in [1,c]$) are obtained. Using the evaluation index of cluster validity as the criterion, determine the optimal number of NWP data sample clusters. Suppose $V(U, V, c)$ is the smallest when the number of clusters is $c$. At this time, NWP data has been divided into $c$ typical weather types, and the category center $c_j$ and sample data of each typical weather type are determined.

The evaluation indexes of clustering effectiveness are as follows:
In the formula: $U$ is the membership matrix, $V$ is the cluster center, $c$ is the number of clusters, $x_i$ is the position of the $i$-th point, and $v_k$ is the position of the $k$-th cluster center.

The smaller $X$ is, the more compact each cluster is, and the more independent each cluster is, which means that the cluster analysis results are more reasonable.

3.2. Predicted sample classification
To classify the prediction set, the classification method uses the Gaussian index method\(^6\). The specific steps are as follows:
First, calculate the Euclidean distance $d$ between the $s$-th feature of the prediction sample and the $s$-th feature of the clustering center of the dataset.

$$d_s(a,b) = \sqrt{(X_{as} - X_{bs})^2}$$  \hspace{1cm} (23)

$a$ —— prediction sample; $b$ —— clustering center; $X_{as}$ —— $s$ feature of the prediction sample; $X_{bs}$ —— $s$ feature of the cluster center.

Next, calculate the Gaussian index of the $s$th feature, see equation (24) and equation (25).

$$g_s(a,b) = \exp \left[-\left(d_s(a,b)/\sqrt{2}\times\sigma_s\right)^2\right]$$  \hspace{1cm} (24)

$$\sigma_s = \sigma \times (s_{\text{max}} - s_{\text{min}})$$  \hspace{1cm} (25)

$\sigma_s$ —— deflection point; $s_{\text{max}}$ —— maximum value of the $s$-th feature; $s_{\text{min}}$ —— minimum value of the $s$-th feature; $\sigma$ —— a constant within the range of $[0,1]$.

Finally, suppose that the weight of the $s$ feature of the sample is $w_s \in [0,1]$ and $\sum w_s = 1$. Use equation (26) to find the similarity of the predicted sample and the cluster set:

$$SIM(a,b) = \sum w_s g_s(a,b)$$  \hspace{1cm} (26)

$m$ —— number of features in the cluster set. The larger $SIM(a,b)$ indicates that the predicted samples are more similar to the cluster center, and the predicted samples are classified into the cluster set with the highest similarity.

3.3. Model building
First, reprocess the NWP data, and then the Neutrosophic clustering analysis is performed to obtain $c$ typical weather types. The GA-ELM models were trained for each of the $c$ typical weather types to obtain the $c$-group GA-ELM model. The forecast data are classified by Gaussian index method. The forecast data of different categories are predicted by corresponding GA-ELM models to obtain the forecast value. The specific block diagram is shown in Figure 2.

3.4. Evaluation index
In order to verify the prediction effect of the model, this paper uses Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) as evaluation indicators to verify the prediction effect\(^7\).

$$MAE = \frac{1}{m} \sum_{i=1}^{m}|e_i|$$  \hspace{1cm} (27)
\[ MAPE = \frac{1}{m} \sum_{t=1}^{m} \left| \frac{e_t}{R_t} \right| \]  
(28)

\[ RMSE = \sqrt{\frac{1}{m} \sum_{t=1}^{m} (e_t)^2} \]  
(29)

\( e_t \) —— predicted absolute error value, \( e_t = R_t - Y_t \) (where \( R_t \) —— the t-th wind power measured value; \( Y_t \) —— the t-th wind power predicted value); \( m \) —— the number of predicted time points.

**Figure 1. Algorithmic Flow Chart of GA-ELM Network**

**Figure 2. Prediction Model Block Diagram Based on Nutrosophic Clustering and GA-ELM**

### 4. Simulation Experiment

#### 4.1. Determination of cluster centers and numbers

This paper takes the January 2017 NWP data of the 14MW wind farm in Northeast China as an example. The NWP data of the first 30 days and a total of 2880 sets of training data were subjected to Neutrosophic cluster analysis, and the data of the 31st day was used as prediction data for prediction, the time interval is 15 minutes.

The NCM operating parameters are as follows: \( m = 2.2, \delta = 1.4, \varepsilon = 10^{-5}, w_1 = 0.4, w_2 = 0.3, w_3 = 0.3 \). The FCM operating parameters are as follows: \( m = 2, \varepsilon = 10^{-5} \). Table 1 shows the effectiveness function values of different clustering numbers in the cluster analysis of the NWP data of wind measurement by the two methods of FCM and NCM.

| Number of clusters | 4   | 5   | 6   | 7   | 8   |
|--------------------|-----|-----|-----|-----|-----|
| V-FCM              | 0.1034 | 0.0853 | 0.0742 | 0.1823 | 0.1211 |
| V-NCM              | 0.0372 | 0.0405 | 0.0267 | 0.0295 | 0.0447 |

It can be seen from the table that the FCM method takes the minimum cluster validity when the number of clusters is 6, and the NCM method also takes the minimum cluster validity value when the number of clusters is 6, which shows that the best wind data can be divided into 6 typical weather types. Among
them, the clustering validity function value of the NCM method is smaller than that of the FCM method in the number of different clusters, which shows that the clustering effect of the NCM method on the wind measurement data is better than the FCM method. The cluster centers of 6 typical weather types are shown in Table 2.

Table 2. Clustering Centers Obtained by NCM Algorithms

| category | Clustering center |
|----------|------------------|
| 1        | (0.6033,0.3732,0.4235,0.5402,0.5311) |
| 2        | (0.6289,0.2690,0.7134,0.1799,0.1929) |
| 3        | (0.8001,0.5991,0.5112,0.2337,0.4213) |
| 4        | (0.7195,0.2735,0.5033,0.3921,0.3538) |
| 5        | (0.9022,0.3977,0.0511,0.9029,0.9455) |
| 6        | (0.5998,0.5965,0.8564,0.1024,0.0724) |

The six typical weather types are recorded as cluster sets $D_1, D_2, D_3, D_4, D_5, D_6$. The number of samples in each cluster set is shown in Table 3.

Six types of training sets $D_1, D_2, D_3, D_4, D_5, D_6$ were substituted into GA-ELM models to establish $M_1, M_2, M_3, M_4, M_5, M_6$ models.

4.2. Forecast data classification results
The predicted samples are classified by Gaussian index method, and the classification results are shown in Table 3.

4.3. Forecast results analysis
The prediction results are shown in Figure 3. The fitting degree of the prediction curve of the GA-ELM model is better than that of the ELM model, which shows that the genetic algorithm optimizes the input weights and thresholds of the ELM algorithm to improve the prediction accuracy. The fitting curve of GA-ELM model based on FCM clustering is better than that of GA-ELM model. This shows that the clustering analysis of the training data will separate the uncorrelated data to avoid the influence of the data with greater differences on the establishment of the model, so that the established model presents a better mapping relationship. Obviously, the NCM-GA-ELM model fits the power curve well.

Table 4. Evaluation Index of Wind Farm Prediction in January

| Method of prediction | MAE/MW | RMSE/MW | MAPE/% |
|----------------------|--------|---------|--------|
| ELM                  | 1.7950 | 2.0444  | 23.79  |
It can be seen from Table 4 that the MAE and RMSE values of the FCM-GA-ELM and NCM-GA-ELM methods are significantly lower than those of the ELM and GA-ELM methods. The clustering analysis of the training data can improve the accuracy of the prediction model. The MAPE value of the NCM-GA-ELM method is lower than that of the FCM-GA-ELM method by 1.58%, and the MSE and RMSE values are reduced by 0.0546MW and 0.0594MW, respectively. This shows that the neutrosophic clustering algorithm can reduce the impact of boundary data and abnormal data on the clustering center, thereby making the clustering center more accurate and improving the final prediction accuracy.

In order to verify the universality of the NCM-GA-ELM method, this paper uses the NWP data of the wind field in March and August for verification. As shown in Figures 4 and 5, in March and August, the power curve fit of the NCM-GA-ELM method is better than the FCM-GA-ELM, GA-ELM and ELM methods, which shows that the method in this paper is suitable for NWP data in different seasons has certain applicability. As shown in Table 5, in the March prediction results, the MAPE value of the NCM-GA-ELM method is reduced by 3.54% compared with the FCM-GA-ELM method, and the MSE and RMSE values are reduced by 0.1090MW and 0.1298MW, respectively. In August, the MAPE value of the NCM-GA-ELM method was 3.90% lower than that of the FCM-GA-ELM method, and the MSE and RMSE values were reduced by 0.2293MW and 0.2908MW, respectively. This shows that the neutrosophic clustering method has certain universality in processing wind power NWP data.
Table 5. Evaluation Index of Wind Farm Prediction in March and August

| Time | Method of prediction | MAE/MW  | RMSE/MW | MAPE/% |
|------|----------------------|---------|---------|--------|
| March| ELM                  | 0.5866  | 0.6930  | 21.12  |
|      | GA-ELM               | 0.2639  | 0.3271  | 8.89   |
|      | FCM-GA-ELM           | 0.1287  | 0.1525  | 4.27   |
|      | NCM-GA-ELM           | 0.0197  | 0.0227  | 0.73   |
| August| ELM                 | 0.9523  | 1.1602  | 20.05  |
|       | GA-ELM               | 0.7196  | 0.7940  | 16.18  |
|       | FCM-GA-ELM           | 0.3622  | 0.4611  | 6.69   |
|       | NCM-GA-ELM           | 0.1329  | 0.1603  | 2.79   |

5. Conclusions
This paper proposes a short-term wind power forecasting method based on neutrosophic clustering and GA-ELM. Aiming at the regularity and difference of NWP data, the NWP data is clustered and analyzed, and the NCM method is used to solve the problems of cluster center shift, edge data points and the membership of abnormal data points in the traditional FCM method. In this paper, the genetic algorithm is used to optimize the ELM network, which makes the model more accurate and more stable. The example of NWP data of a 14MW wind farm in Northeast China shows that this method has good prediction effect. This method applies neutrosophic theory to the field of wind power forecasting. It is an extension of NWP data clustering in wind farms. It has certain engineering value and provides an effective way for short-term wind power forecasting.

References
[1] Y. Mao, Z. Guanqiang, Real-time prediction method of wind power based on decision tree theory, Electric Measurement and Instrumentation, vol. 55, pp. 11, 2017.
[2] W. Chengliang, Z. Haiyan, Portrait of power customer electricity consumption behavior pattern based on fuzzy clustering, Electric Measurement & Instrumentation, 2018.
[3] Y. Guo, A. Sengur, Necm: neutrosophic evidential c-means clustering algorithm, Neural Computing & Applications, vol. 26, issue 3, pp. 561-571, 2015.
[4] G. B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Transactions on Systems Man & Cybernetics Part B, vol. 42, issue 2, pp. 513-529, 2012.
[5] Z. Jian, C. Wei, Day-to-day electricity price prediction based on GA-ELM neural network, Journal of Shanghai Electric Power University, vol. 34, issue 1, pp. 90-94, 2018.
[6] Z. Kangning, L. Xiangyang, Y. Kai, Gauss—Prediction Model and Application of Case Reasoning Method, Operations Research and Management, 2011.
[7] Z. Nianfang, L. Shanming, Short-term wind power prediction model based on neighborhood KNN algorithm, Electric Measurement & Instrumentation, vol. 16, pp. 26-30, 2017.