POLISH MARITIME RESEARCH Special Issue 2017 S3 (95) 2017 Vol. 24; pp. 182-191
10.1515/pomr-2017-0121

BIG DATA ANALYSIS AND SIMULATION OF DISTRIBUTED MARINE GREEN ENERGY RESOURCES GRID-CONNECTED SYSTEM

Jun Tian
Lirong Huang
Foshan Polytechnic, Foshan 528137, China

ABSTRACT

In order to improve the working stability of distributed marine green energy resources grid-connected system, we need the big data information mining and fusion processing of grid-connected system and the information integration and recognition of distributed marine green energy grid-connected system based on big data analysis method, and improve the output performance of energy grid-connected system. This paper proposed a big data analysis method of distributed marine green energy resources grid-connected system based on closed-loop information fusion and auto correlation characteristic information mining. This method realized the big data closed-loop operation and maintenance management of grid-connected system, and built the big data information collection model of marine green energy resources grid-connected system, and reconstructs the feature space of the collected big data, and constructed the characteristic equation of fuzzy data closed-loop operation and maintenance management in convex spaces, and used the adaptive feature fusion method to achieve the auto correlation characteristics mining of big data operation and maintenance information, and improved the ability of information scheduling and information mining of distributed marine green energy resources grid-connected system. Simulation results show that using this method for the big data analysis of distributed marine green energy resources grid-connected system and using the multidimensional analysis technology of big data can improve the ability of information scheduling and information mining of distributed marine green energy resources grid-connected system, realizing the information optimization scheduling of grid-connected system. The output performance of grid connected system has been improved.

Keywords: Distributed, Ocean, Green energy resources, Grid-connected system, Big data analysis

INTRODUCTION

Green energy resources is also known as clean energy, which is the symbol and synonym of the environmental protection and the good ecosystem. Marine green energy resources are the renewable natural energy resources contained in the oceans, which is renewable and inexhaustible in the era of existence of solar system [1].

The marine green energy resources grid-connected system includes of ocean tidal power generating system, marine wind power generating system and ocean thermal energy conversion. The marine green energy resources is the current important resource. We carr y out the effective integration scheduling and the operation and maintenance management on the distributed marine green energy grid-connected system big data distributed marine green energy grid system, and provide accurate data basis for grid-connected generating system, so as to improve the ability of prediction and judgment of grid-connected generating system. The construction of distributed marine green energy resources grid-connected system and the parallel scheduling and the operation and maintenance management of the grid-connected system in the big data environment can improve the stability of the of green energy resources grid-connected system. The research of distributed marine green energy grid-connected system big data analysis method has important significance [2].

The data of marine green energy grid system big data analysis is mainly used for data fusion and scheduling. There are many big data analysis fusion scheduling and operation and maintenance management algorithms, which can be divided into the fuzzy fusion scheduling and operation and maintenance management algorithm[3], the hierarchic fusion scheduling and operation and maintenance management algorithm, the mesh fusion scheduling and operation and maintenance management algorithm, the support vector machines fusion scheduling and operation and maintenance management algorithm and BP neural network fusion scheduling and operation and maintenance management algorithm[4-8]. Among them, the hierarchic fusion scheduling and operation and maintenance management algorithm takes the attributes category of distributed marine green energy grid-connected system according to the number of as hierarchic grid feature for the fusion scheduling and operation and maintenance management. In the hierarchic fusion scheduling and operation and maintenance management, with the change of category level, the object also changes [9]. Reference [10] proposes the data fusion scheduling and operation and maintenance management based on Naive Bayesian in cloud computing environment for the big data classification of distributed marine green energy grid-connected system. This method extracts semantic relevance and rule characteristics of big data of distributed marine green energy grid-connected system, and carries out the fusion scheduling and operation and maintenance management for the characteristics, and improves the precision of distributed marine green energy grid-connected system large database retrieval. But with the increase of distributed marine green energy grid-connected system large scale database, the accuracy of information fusion scheduling and operation and maintenance management is not good. Reference [11] proposes the method of big data fusion scheduling with operation and maintenance management of distributed marine green energy grid-connected system in cloud model combined fusion scheduling with operation and maintenance management. This method uses multi strategy similarity calculation for the substructure information feature modeling of big database in distributed marine green energy grid-connected system big data information, and realizes the collaborative recommendation fusion scheduling and operation and maintenance management of big data query and access in distributed marine green energy grid-connected system, but this method under similar information interference has low accuracy of big data information fusion of grid-connected system, and has no data dimension reduction, resulting in large computational overhead.

Aiming at the above problems, this paper proposed a big data analysis method of distributed marine green energy resources grid-connected system based on closed-loop information fusion and auto correlation characteristic information mining. This method realized the big data closed-loop operation and maintenance management of grid-connected system, and firstly built the big data information collection model of marine green energy resources grid-connected system, and reconstructs the feature space of the collected big data, and constructed the characteristic equation of fuzzy data closed-loop operation and maintenance management in convex spaces, and then used the adaptive feature fusion method to achieve the auto correlation characteristics mining of big data operation and maintenance information, and improved the ability of information scheduling and information mining of distributed marine green energy resources grid-connected system, finally, we got valid conclusion from the simulation experiment analysis of big data analysis, which show the superior performance of this method in improving the big data analysis and operation and maintenance management ability of distributed marine green energy resources grid-connected system.

BIG DATA STRUCTURE ANALYSIS AND FEATURE SELECTION OF MARINE GREEN ENERGY GRID-CONNECTED SYSTEM

OVERALL FRAMEWORK OF LARGE DATA ANALYSIS OF MARINE GREEN ENERGY GRID-CONNECTED SYSTEM

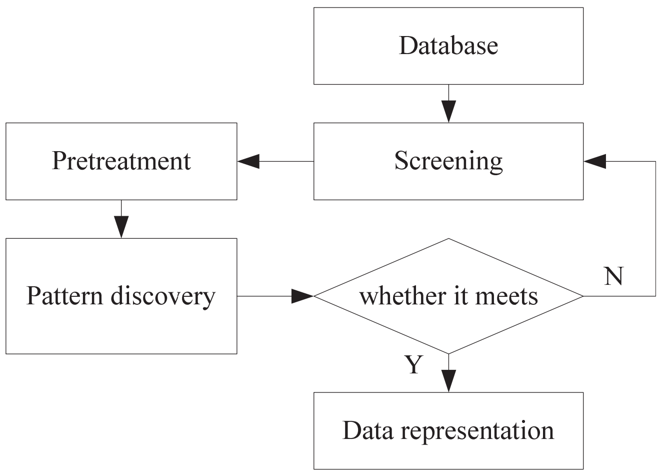


Fig. 1. Big data mining process of marine green energy resources grid-connected system.

In order to realize the integrated feature extraction and data fusion scheduling of big data of distributed marine green energy resources grid-connected system, we achieve the big data operation and maintenance management and pattern recognition of marine green energy resources grid-connected system through big data analysis method. Firstly, we build big data information acquisition model of marine green energy resources grid-connected system, and carry out the feature space reconstruction of collected big data, and combines with feature extraction methods for data mining, achieving the selective preference and data mining of big data feature in marine green energy resources grid-connected system. Data mining is used to analyze big data of grid-connected system. Data mining (DM) is the process of finding target data from massive data [12]. The massive data has a lot of interference options and the fuzziness and randomness. The location of target data is unknown, which is hidden in massive data, and

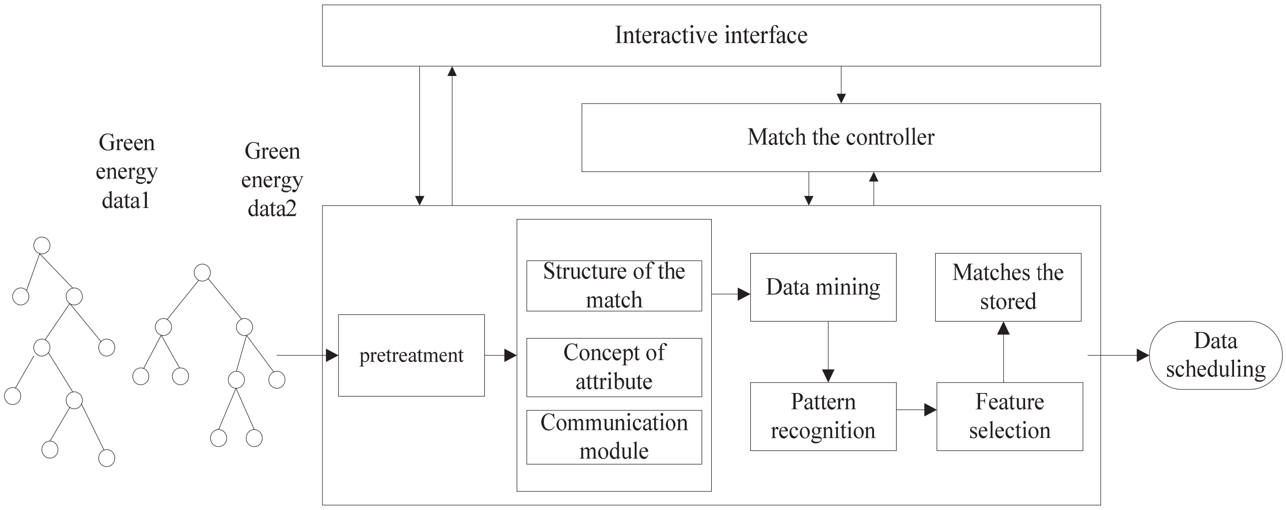


Fig. 2. Big data analysis model of marine green energy resources grid-connected system.

the class is unknown. In the big data analysis of green energy resources grid-connected system, we need to perform the data screening, the data preprocessing, the pattern extraction, the pattern analysis and the data application, to achieve the big data mining of marine green energy resources grid-connected system [13]. According to the above analysis, we build the big data analysis model of marine green energy resources grid-connected system, as shown in Fig. 2.

BIG DATA INFORMATION COLLECTION MODEL

The Takens embedding theorem is used to reconstruct the phase space and the feature, and a clustering model of big data distribution of distributed marine green energy resources grid-connected system is obtained. [14-16].

Takens theorem: let M be the d-dimensional compact manifold. F is a smooth vector field. h is a smoothing function on M , and that is $M \to R^{2d+1}$, which shows that $\Phi(z) = (h(z), h(\varphi_1(z)), \cdots, h(\varphi_{2d}(z)))^T$ is an embedded vector.

For big data sampling time series $\{x(t_0 + i\Delta t)\}$, $i = 0,1,\cdots, N-1$ of distributed marine green energy resources grid-connected system, its phase space reconstruction trajectory is:

$$X = [x(t_0), x(t_0+\Delta t), \cdots, x(t_0+(K-1)\Delta t)]$$
$$= \begin{bmatrix} x(t_0) & x(t_0+\Delta t) & \cdots & x(t_0+(K-1)\Delta t) \\ x(t_0+J\Delta t) & x(t_0+(J+1)\Delta t) & \cdots & x(t_0+(K-1)\Delta t+J\Delta t) \\ \vdots & \cdots & \ddots & \cdots \\ x(t_0+(m-1)J\Delta t) & x(t_0+(1+(m-1)J)\Delta t) & \cdots & x(t_0+(N-1)\Delta t) \end{bmatrix} \tag{1}$$

Among them, $x(t)$ represents the state vector of the embedding space. J is the reconstruction delay, m is the embedding dimension, Δt is the time interval of sampling, $K = N-(m-1)J$, and τ represents the delay parameter, $\tau = J\Delta t$. $\tau_w = (m-1)\tau = (m-1)J\Delta t$ is time window. We can see that in reconstructing phase space, we only select any two parameters in m , τ and τ_w , another parameter can be obtained directly by $\tau_w = (m-1)\tau$.

In the Laplace convex optimization space [17], we constructs the high-order linear differential equation combination model of fuzzy data closed loop operation and maintenance management:

$$B_l(A) = \min_{\beta \neq 0}\{w(\beta) + w(A^T \cdot \beta)\} \tag{2}$$

In formulas, A^T represents the transpose of matrix A . Simultaneous equation solving gets the solution of the equations: $c_1 = -c_{-1} = 0.675$, $c_2 = -c_{-2} = -2^{1/3}c_1 = -0.85$, for $s, \tilde{s} \geq 0$, there is the mapping $T: U \to U$. The characteristic solution $\forall u(t) \in U$ in variable kernel convex space solution obtains the objective function of big data clustering through constructing the expectation function of fuzzy data distributed distributed marine green energy resources grid-connected system closed loop operation and maintenance management in convex space:

$$\max_{x_{a,b,d,p}} \quad \sum_{a\in \mathrm{A}}\sum_{b\in \mathrm{B}}\sum_{d\in \mathrm{D}}\sum_{p\in \mathrm{P}} x_{a,b,d,p}V_p \tag{3}$$

$$\text{s.t.} \quad \sum_{a\in \mathrm{A}}\sum_{d\in \mathrm{D}}\sum_{p\in \mathrm{P}} x_{a,b,d,p}R_p^{bw} \leq K_b^{bw}(\mathrm{S}), b \in \mathrm{B} \tag{4}$$

$$\sum_{a\in \mathrm{A}}\sum_{b\in \mathrm{B}}\sum_{p\in \mathrm{P}} x_{a,b,d,p}R_p^{cp} \leq K_d^{cp}(\mathrm{S}), d \in \mathrm{D} \tag{5}$$

$$\sum_{b\in \mathrm{B}}\sum_{d\in \mathrm{D}} x_{a,b,d,p} \leq \mathrm{D}_{a,p}, a \in \mathrm{A}, p \in \mathrm{P} \tag{6}$$

$$\sum_{d\in \mathrm{D}}\sum_{p\in \mathrm{P}} x_{a,b,d,p} \leq M\alpha_{a,b}, a \in \mathrm{A}, b \in \mathrm{B} \tag{7}$$

$$\sum_{b\in \mathrm{B}} x_{a,b,d,p} \leq M\beta_{a,d,p}, a \in \mathrm{A}, d \in \mathrm{D}, p \in \mathrm{P} \tag{8}$$

$$x_{a,b,d,p} \geq 0, a \in \mathrm{A}, b \in \mathrm{B}, d \in \mathrm{D}, p \in \mathrm{P} \tag{9}$$

According to the above data acquisition model, we carry out the big data analysis. The data analysis involves many techniques and methods, this paper uses fuzzy neural network classification decision method to classify the data, and builds the high dimensional feature space distribution structure model of distributed marine green energy resources grid-connected system big data.

BIG DATA CLASSIFICATION AND HIGH DIMENSIONAL INFORMATION REORGANIZATION

We extract the feature vector of correlative dimension of distributed marine green energy resources grid-connected system big data in the phase space, and carry out the ensemble feature selection of data. In the reconstructed distributed marine green energy resources grid-connected system big data distribution phase space, and uses the nonlinear time series analysis method for the feature selection of correlative dimension [18]. For distributed marine green energy resources grid-connected system big data sequence $x_1, x_2, \cdots, x_n, \cdots$, we set the total number of points as N, the sampling time span of sequential $\{x_i\}$ is $j\tau$, and its autocorrelation function is:

$$R_{xx}(j\tau) = \frac{1}{N}\sum_{i=0}^{N-1} x_i x_{i+j\tau} \tag{10}$$

Thus, we can fix j and make the correlation function about time τ (taking $\tau = 1, 2, \cdots$), and the finite data set X is divided into c class according to the time delay of the big data embedding, among them 1<c<n , and the fuzzy clustering central matrixis defined as :

$$V = \{v_{ij} \,|\, i = 1, 2, \cdots, c,\ j = 1, 2, \cdots, s\} \tag{11}$$

V_i is the i -th vector of the cluster center (the i -th cluster center vector). The fuzzy partition matrix is represented as:

$$U = \{\mu_{ik} \,|\, i = 1, 2, \cdots, c, k = 1, 2, \cdots, n\} \tag{12}$$

To construct the ontology model which reflects the associated feature of distributed marine green energy resources grid-connected system big data [19], the relative weight of distributed marine green energy resources grid-connected system big data is $\omega = ((\omega_1, a_1'),\ (\omega_2, a_2),\ldots,\ (\omega_n, a_n))^T$, $\omega_j \in [0,1]$ and we use the average mutual information method to obtain the objective function that the data association is the features extraction:

$$J_m(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} \mu_{ik}{}^m (d_{ik})^2 \tag{13}$$

In the formula, m is the weight index. $(d_{ik})^2$ is the measuring distance between the sample x_k and V_i, which is expressed by Euclidean distance:

$$(d_{ik})^2 = \|x_k - V_i\|^2 \tag{14}$$

and

$$\sum_{i=1}^{c} \mu_{ik} = 1, k = 1, 2, \cdots, n \tag{15}$$

Combining with the decision function of association mapping, the differential control coefficient of feature extraction of correlation dimension of distributed marine green energy resources grid-connected system big data is obtained by using Lagrange theorem:

$$\mu_{ik} = 1 \Big/ \sum_{j=1}^{c} \left(d_{ik} \big/ d_{jk} \right)^{\frac{2}{m-1}} \tag{16}$$

$$V_i = \sum_{k=1}^{m} (\mu_{ik})^m x_k \Big/ \sum_{k=1}^{n} (\mu_{ik})^m \tag{17}$$

The maximal linearly independent subset of distributed marine green energy resources grid-connected system big data closed loop operation and maintenance management is obtained, for an arbitrary set of data geometry i, s_i^* is the least square fitting vector, the feature solution $\forall u(t) \in U$ in kernel convex spaces satisfies the maximum Lyapunove functional, through construction of the fuzzy feature subset of the massive information, we get the iterative acceleration formula of big data classification of marine green energy resources grid-connected system:

$$(1-\omega)x_i^{(k)} + \omega x_i^{-(k+1)} = x_i^{(k+1)}, i = 1, 2, \cdots, n \tag{18}$$

In clustering sample sets, we add or delete sample A or B, and for fuzzy data sample set A, B, the convergence control function of fuzzy data closed loop operation and maintenance management in massive information is obtained under the control of attenuation constant T_1 and T_2:

$$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \frac{\omega}{a_{n\,i}}\left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} \right) \tag{19}$$
$$i = 1, 2, \cdots, n$$
$$k = 1, 2, \cdots, n$$

We use the support vector machine model to obtain the initial clustering center of big data clustering. The initial value the data fusion center has been given, we adjust the parameter c of clustering attributes classification and the fuzziness index m, and carry out the feature compression processing, reducing the dimension of feature selection, so as to reduce the computational overhead of data fusion scheduling and operation and maintenance management [20]. According to the data classification results, we use the following five steps for high dimensional information reorganization:

(1) we firstly select a k value to determine the total number of recombination clusters of high dimensional information. If the data set is m, order $A_j(L)$ as the center of the cluster, among them, $j = 1, 2, \ldots, k$. the distance to the cluster center is calculated;

(2) select k examples in the data set, and initialise cluster centers $F(x_i, A_j(L))$, $i = 1, 2, ..., m$, $j = 1, 2, ..., k$;
(3) use the simple Euler distance to assign the remaining clusters to the nearest cluster center, if satisfied:

$$D(x_i, A_j(L)) = \min\{D(x_i, A_j(L))\} \tag{20}$$

Thus $x_i \in \omega_k$;
(4) in the massive information storage space, supposing a sample $i \in S_s$, let $\beta_i^c \neq \pm\infty$, when the stationary vector set $\det(Q^{'}) = 0$ of the standard data set, through adaptive weighting of clustering center of massive data, when $\forall i \in S_s$, we can get $\beta_i^c \neq \pm\infty$. In convex optimization space, we carry out the self-adaptive adjustment of data clustering center, we can get: $\forall i \in S - S_s$, $\gamma_i^c \neq \pm\infty$ in the sample updating process of set S_s , we use the example of each cluster to calculate the average value as the average value of new cluster:

$$C(l) = \sum_{j=1}^{k} \sum_{k=1}^{n_j} (\| x_k^j - A_j(L) \|)^2 \tag{21}$$

(5) if the average value is equal to the average value of the last iteration, $\| C(l) - C(l-1) \| < \xi$,the program is stopped, otherwise return to third step, let $l = l + 1$,and the new cluster heart is calculated:

$$A_j(L+1) = {}^{1}\!/_{n_j} \sum_{i=1}^{k} X_i^j \tag{22}$$

BIG DATA CLOSED LOOP OPERATION AND MAINTENANCE MANAGEMENT OF DISTRIBUTED MARINE GREEN ENERGY RESOURCES GRID-CONNECTED SYSTEM

FUZZY DATA CLOSED-LOOP OPERATION AND MAINTENANCE INFORMATION FUSION

This paper proposes a big data analysis method of distributed marine green energy resources grid-connected system based on and closed loop information fusion and auto correlation feature information mining, realizing the big data closed-loop operation and maintenance management of grid-connected system [21]. In the convex space, we build the characteristic equation of fuzzy data closed loop operation and maintenance management, and the global fluctuation combination stable solution of high order linear differential equation of data closed loop operation and maintenance management of closed loop system satisfies $\forall i \in S_s$, $\beta_i^c \neq \pm\infty$, and $\forall i \in S - S_s$, $\gamma_i^c \neq \pm\infty$. For $\forall i \in S_s$, the the dual periodic solitary wave solution of data closed-loop operation and maintenance management is:

$$\begin{aligned} \beta_i^c &= -\sum_{k \in S_s} R_{ik} Q_{kc} - R_{i1} y_c \\ &= -\frac{1}{\det\left(Q^{'}\right)} \left(\sum_{k \in S_s} (-1)^{i+k} \det\left(Q^{'}_{\setminus ki}\right) Q_{kc} + y_c (-1)^{i+1} \det\left(Q^{'}_{i1}\right) \right) \end{aligned} \tag{23}$$

To build a set of homogeneous equations for $\dot{H}_x^s$ -global strong stable functional, when the extremal functional formulas of polynomial kernel function and Gaussian kernel function satisfy $d = 4, s_c = \frac{3}{2}$, maximal linearly independent set expressions of fuzzy data closed-loop operation and maintenance management respectively are :

$$K(x_i, x_j) = < x_i, x_j > \tag{24}$$

$$K(x_i, x_j) = (< x_i, x_j > +1)^d \tag{25}$$

$$K(x_i, x_j) = \exp(\left\| x_i - x_j \right\|^2 / 2\sigma^2) \tag{26}$$

For distributed marine green energy resources grid-connected system big data sequence $\{x_n\}_{n=1}^{N}$ of single variable, in the reconstruction of the m -dimensional phase space dimension, the distance form phase point x_j to x_i , besides x_i itself is less than the x_j points of r which is expressed as:

$$Q = \sum_{j \neq i} H\left(r - \left\| x_i - x_j \right\|\right) \tag{27}$$

$H(.)$ represents the Heavside function. That is:

$$H(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases} \tag{28}$$

We set the sampling point number of one-dimensional distributed marine green energy resources grid-connected system big data is n . The number of vector points in the reconstructed phase space is $N = n - (m-1)\tau$, and we calculate the number of related phase point pair in the phase points. The ratio of all possible pairs of $N(N-1)/2$ matches is called association integral:

$$C_m(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} H\left(r - \left\| x_i - x_j \right\|\right) \tag{29}$$

When the amount of data is $N \to \infty$ and the distance r is small, that is $r \to 0$, if the correlation integral $C_m(r)$ is obeyed by the exponential law, the correlation dimension features of the big data of distributed marine green energy resources grid-connected system is obtained as follows:

$$D = \frac{\ln C_m(r)}{\ln r} \tag{30}$$

In the calculation, the range $r_{\min}$ and $r_{\max}$ of r value is usually given, according to a certain growth rate to change

r, and then start to increase from 1, increase 1 every time, gradually increase it to $m_{\max}$, we uses the adaptive feature fusion method to achieve the self correlation feature mining of big data operation and maintenance information. The MOLAP method can be used for data scheduling of distributed marine green energy resources grid-connected system, which is faster than other traditional analytical techniques and can be predicted [22].

In the closed-loop operation and maintenance management of fuzzy data, it is divided into data ETL layer, data storage layer, data analysis layer and application layer. [23]

SELF CORRELATION FEATURES MINING OF BIG DATA OPERATION AND MAINTENANCE INFORMATION

We reduce the dimension for extracted integrated feature of correlation dimension of distributed marine green energy resources grid-connected system big data, and realize the autocorrelation feature mining of the big data operation and maintenance information, and use the K-L feature compression method for reducing the dimension [24-26], realizing the selective preference control of ensemble features of distributed marine green energy resources grid-connected system big data. The steps can be summarized as follows:

(1) calculate the intra-class dispersion matrix $\hat{S}_w$ of l -dimensional feature vector $\vec{X}(l,n_i)$ of distributed marine green energy resources grid-connected system big data, and find its l eigenvalues $\lambda_1,\lambda_2,\cdots,\lambda_l$ and eigenvectors matrix $Y=[y_1,y_2,\cdots,y_l]$. $\hat{S}_w$ is:

$$\hat{S}_w=\sum_{i=1}^{c}p_i\frac{1}{n_i}\sum_{k=1}^{n_i}\left[\left(\vec{X}_k^{(i)}-\vec{m}_i\right)\left(\vec{X}_k^{(i)}-\vec{m}_i\right)^T\right] \tag{31}$$

Among them, p_i represents the prior probability of i -th categories of data fusion scheduling and operation and maintenance management attribute. n_i represents the number of samples of i -th categories of data fusion scheduling and operation and maintenance management attribute. $\vec{m}_i$ represents the data represent the mean vector of n_i -th sample feature vector set $\left\{\vec{X}_k^{(i)},k=1,2,\cdots,n_i\right\}$ of i -th categories;

(2) compute the between-class scatter matrix of fusion scheduling and operation and maintenance management of distributed marine green energy resources grid-connected system :

$$S_b=\sum_{i=1}^{c}p_i\left(\vec{m}_i-\vec{m}\right)\left(\vec{m}_i-\vec{m}\right)^T \tag{32}$$

Among them, $\vec{m}=\sum_{i=1}^{c}p_i\vec{m}_i$ is the ensemble average;

(3) calculate the average distance between classes:

$$J\left(\vec{X}_j\right)=\frac{y_j^TS_by_j}{\lambda_j} \tag{33}$$

And put them in descending order:

$$J\left(\vec{X}_1\right)\geq J\left(\vec{X}_2\right)\geq\cdots\geq J\left(\vec{X}_l\right);$$

(4) if it is reduced to d dimension, then we take top d larger eigenvectors y_j corresponding to $J\left(\vec{X}_j\right)$, $j=1,2,\cdots,d$ and the transformation matrix is generated:

$$W=\left[y_1,y_2,\cdots,y_d\right] \tag{34}$$

(5) considering the point which is satisfied $|i-j|\geq\omega$, the dimension reduction output results of the extracted feature vector of correlation dimension is obtained by K-L transform:

$$\vec{X}^*=W^T\vec{X} \tag{35}$$

Through the above processing, the correlation dimension feature vector of the big data of the distributed marine green energy resources grid-connected system is reduced to d dimension form l dimension, thus reducing the complexity and memory overhead of the fusion scheduling and the operation and maintenance management calculation.

Based on the principle of singular positive semidefinite, the quadratic programming model of support vector machines fuzzy data closed-loop operation and maintenance management is established. Suppose the matrix Q is positive semidefinite matrix:

$$\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)\neq 0 \tag{36}$$

The homogeneous solution of fuzzy data closed loop operation and maintenance management support vector machine model satisfies:

$$\alpha^TQ\alpha=\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_jQ_{ij}\geq 0 \tag{37}$$

Suppose there is n samples in the information data set S_s, then:

$$Q^{'}=\begin{bmatrix}0 & y_1 & \cdots & y_n\\ y_1 & Q_{11} & \cdots & Q_{1n}\\ \vdots & \vdots & \ddots & \vdots\\ y_n & Q_{n1} & \cdots & Q_{nn}\end{bmatrix}\overset{def}{=}\begin{bmatrix}0 & y^T\\ y & Q\end{bmatrix} \tag{38}$$

According to the above hypothesis, we can see that the distribution matrix Q of periodic point is positive definite, and obtain the inverse matrix Q^{-1} of Q, Q^{-1} is also the positive definite matrix at the same time. Because:

$$\begin{bmatrix}0 & -y^TQ^{-1}\\ 0 & I_n\end{bmatrix}\begin{bmatrix}0 & y^T\\ y & Q\end{bmatrix}=\begin{bmatrix}-y^TQ^{-1} & 0\\ y & Q\end{bmatrix} \tag{39}$$

The fuzzy degree point set of massive fuzzy data, closed-loop operation and maintenance management satisfies:

$$\det\left(Q^{'}\right)=\det\left(Q\right).\left(-y^TQ^{-1}\right)\neq 0 \tag{`40}$$

For the utility value of a single distributed marine green energy resources grid-connected system, the utility value model formula is:

$$\mathrm{Y}\left(\text{Utility value}\right)=\sum f\left(x_i\right)(i\text{represents }1\ldots\ldots n\text{-th Electric energy meter after NotionalPooling})$$

$$f(x_i) = j(x_i) + s(x_i)$$

Among them, the solution result of cluster centers of closed loop operation and maintenance management is:

$$a_{ii}x_i^{(k+1)} = (1-\omega)a_{ii}x_i^{(k)} + \omega\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right) \tag{41}$$

For selected feature vector, we use the random forest fusion scheduling and operation and maintenance management method for feature fusion scheduling and operation and maintenance management, to get a homogeneous vector group that the clustering center of distributed marine green energy resources grid-connected system big data is the quadratic programming [25].

$$(D-\omega L)x^{(k+1)} = \left[(1\text{-}\omega)D+\omega U\right]x^{(k)} + \omega b \tag{42}$$

$$x^{(k+1)} = (D-\omega L)^{-1}\left[(1\text{-}\omega)D+\omega U\right]x^{(k)} + \omega(D-\omega L)^{-1}b \tag{43}$$

Let:

$$\begin{cases} L_\omega = (D-\omega L)^{-1}\left[(1\text{-}\omega)D+\omega U\right] \\ f_\omega = (D-\omega L)^{-1}b \end{cases} \tag{44}$$

Thus, the convergence process of the fuzzy directional clustering of the test data satisfies:

$$x^{(k+1)} = L_\omega x^{(k)} + f \tag{45}$$

There is the Local optimal solution of big data fusion in grid connected system:

$$\begin{aligned} &\frac{x_k-\alpha}{x_{k-1}-\alpha} \approx \frac{x_{k+1}-\alpha}{x_k-\alpha} \\ &(x_k-\alpha)^2 \approx (x_{k+1}-\alpha)(x_{k-1}-\alpha) \\ &{x_k}^2 - 2ax_k - \alpha^2 \approx x_{k+1}x_{k-1} - \alpha(x_{k-1}+x_{k+1}) + \alpha^2 \\ &{x_k}^2 - x_{k+1}x_{k-1} \approx \alpha(2x_{k-1} - x_{k+1} - x_{k-1}) \end{aligned} \tag{46}$$

We take the cluster centers obtained from the support vector machine model as the neighborhood data set, and the training template set is constructed. The quadratic programming of convex combination method is carried out for Lyapunove functional:

$$\alpha \approx \frac{x_{k+1}x_{k-1} - {x_k}^2}{x_{k-1} - 2x_k + x_{k+1}} = x_{k+1} - \frac{(x_{k+1}-x_k)^2}{x_{k-1} - 2x_k + x_{k+1}} \tag{47}$$

Among them:

$$\overline{x_k} = x_{k+1} - \frac{(x_{k+1}-x_k)^2}{x_{k-1} - 2x_k + x_{k+1}} \tag{48}$$

The expected output of fuzzy K means clustering is:

$$E_i((s_j^*,(s_i^*)_{i\in N\setminus\{j\}}) = \sum\nolimits_{j=1}^{m}\sum\nolimits_{k=1}^{l} p(x_{j_1})\cdots p(x_{j_w})u_{k_i} \tag{49}$$

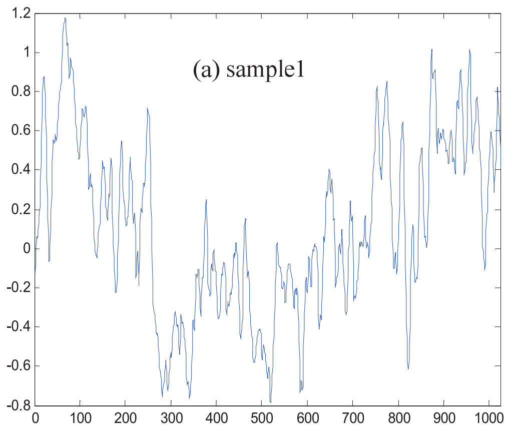
Through the above processing, we select a certain basic function to return the redundant data to the set, and fuzzy data closed-loop operation and maintenance management is realized. The method of adaptive feature fusion is used to realize the self correlation feature mining of big data operation and maintenance information.

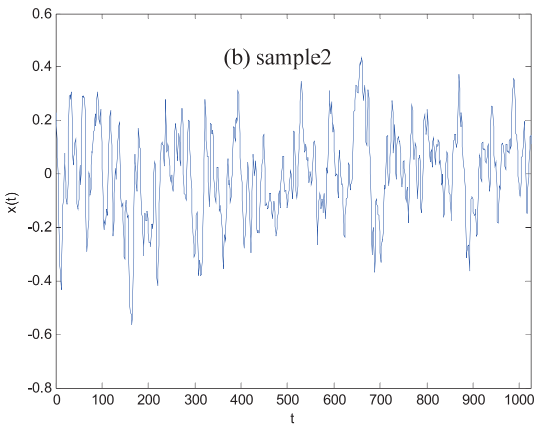
4. Simulation experiment and result analysis

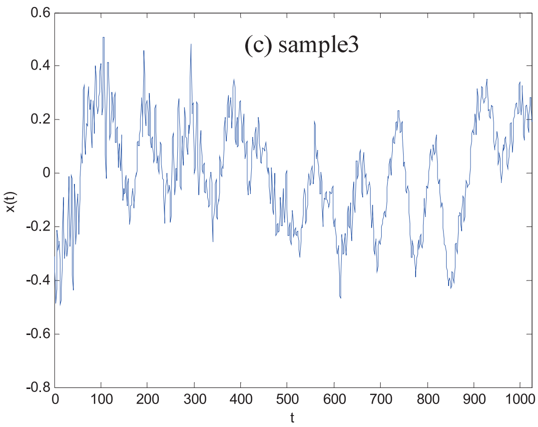
In order to test the application performance of this method in the implementation of the information fusion and the data scheduling and other big data analysis of distributed marine green energy resources grid-connected system, we carry out the simulation experiment. The experimental hardware platform uses Intel i5-3230M 2.6GHz dual core CPU, RAM4GB DDR3, operating system is Windows 7, and editing software of algorithm is VC++ and Matlab. The experimental database of distributed marine green energy resources grid-connected system is CUP2016 grid-connected database of KDD. We carry out the big data feature information sampling of distributed marine green energy resources grid-connected system in the database. The normalized initial sampling frequency is $f_1 = 0.8$ Hz. Termination frequency is $f_2 = 0.15$ Hz. The training tree of random forest is 10, and the maximum step number of iterations is $NP = 30$. The embedding dimension of phase space reconstruction is m=4. Embedding time-delay parameter is τ=11, sampling interval is 0.25 s. The sampling points are 1000 points. There is a 250Hz intra-class attribute difference unbalanced data data frequency component between the 400~600 sampling points. We take the frequency domain characteristic parameters of big data distribution $\sigma = 10$, $b = 8/3$, $r = 28$, and select the initial value $[x_0, y_0, z_0] = [2, 2, 20]$, integration step $h = 0.01$, and use four-order Rung-Kutta method to solve the characteristic equation of big data distribution for the optimum selection control of clustering information fusion center feature. When carrying out the recursive analysis, we take the parameters of phase space reconstruction m =3, τ =8, the threshold ε =0.2std (x), among them, STD (x) represents the standard deviation of the time series, $l_{\min}$ =3. According to the simulation environment and parameter settings, the big data raw data sampling of the distributed marine green energy resources grid-connected system is carried out, and the time domain waveform of data sampling is shown in Fig. 3.

Taking the above sampling data as the research object, we extract the correlation dimension integrated feature of data, and directly input the selected feature quantity to the big data information fusion scheduling and operation and maintenance manager for data fusion scheduling and operation and maintenance management, to achieve big data classification recognition of grid-connected system, and the data classification results are shown in Fig. 4.

The analysis results of Fig. 4 show that using this method for big data classification can realize data fusion scheduling and the operation and maintenance management, and can effectively realize the classification and recognition of big data, and improves the adaptive analysis and processing ability of big data of distributed marine green energy resources grid-connected system. In order to quantitatively analyze the application performance of the proposed method for big data classification processing, Table 1 lists the performance comparison of distributed marine green energy resources grid-connected system big data analysis in different methods [27-30]. From the analysis, the proposed method classifies distributed marine green energy resources grid-connected system big data, which reduces the misclassification rate, and shortens the time of analysis processing of big data, and improves the ability of parallel scheduling of big data, and ameliorates the output performance of the grid-connected system [31,32].







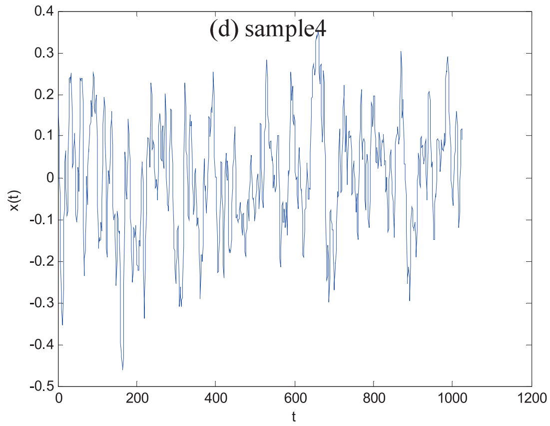
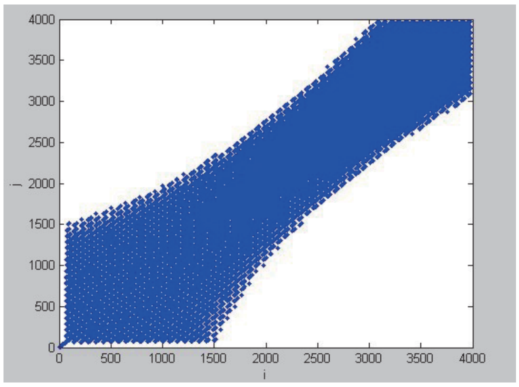
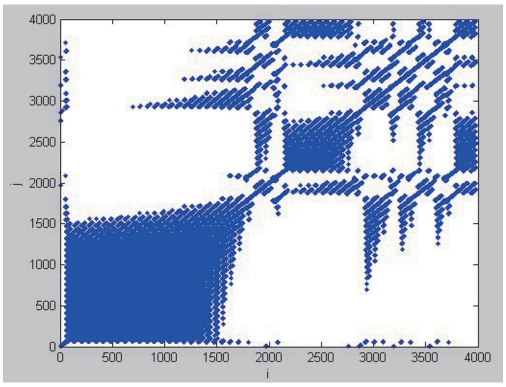


Fig. 3. Time domain waveform of big data sampling of distributed marine green energy resources grid-connected system.

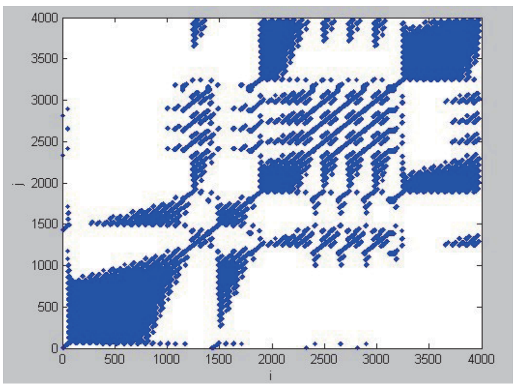
(a)sample1



(b)sample2



(c) sample3



(d)sample4

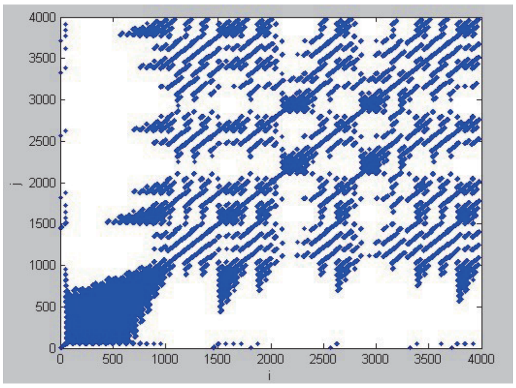


Fig. 4. Big data classification results of distributed marine green energy resources grid-connected system.

Table 1. Performance comparison of big data analysis and processing.

Test sample	Proposed method		Traditional method	
	Misclassification rate /%	Computing time /s	Misclassification rate /%	Computing time /s
1	0.14	1.15	1.32	16.32
2	0.11	1.32	2.21	10.21
3	0.02	2.12	3.21	12.33
4	0.03	1.09	5.43	9.09

CONCLUSIONS

This paper studies the data mining, classification and identification in big data analysis of distributed marine green energy resources grid-connected system, and carries out the information integration and recognition of distributed marine green energy resources grid-connected system based on big data analysis method, improving the output performance of energy resources grid-connected system. This paper proposed a big data analysis method of distributed marine green energy resources grid-connected system based on closed-loop information fusion and auto correlation characteristic information mining. This method realized the big data closed-loop operation and maintenance management of grid-connected system, and built the big data information collection model of marine green energy resources grid-connected system, and reconstructs the feature space of the collected big data, and constructed the characteristic equation of fuzzy data closed-loop operation and maintenance management, and used the adaptive feature fusion method to achieve the auto correlation characteristics mining of big data operation and maintenance information, and improved the ability of information scheduling and information mining of distributed marine green energy resources grid-connected system. The research results show that using this method for the big data analysis of distributed marine green energy resources grid-connected system and using the multidimensional analysis technology of big data can improve the ability of information scheduling and information mining of distributed marine green energy resources grid-connected system, realizing the information optimization scheduling of grid-connected system, and it has positive benefits on improving the output performance of grid-connected system.

REFERENCES

1. M. Ceci, D. Malerba, 2007. Classifying Web documents in a hierarchy of categories:a comprehensive study. *Journal of Intelligent Information System*, 28(1): 37-78.
2. H. X. Wang, S. Y. Wang, X. Wang, et al. 2014. Analysis of LFM signals and improvement of IFM system. *Acta Armamentarii,* 35(8): 1193-1199.
3. M. A. Govoni, H. Li, J. A. Kosinski, 2013. Range-Doppler resolution of the linear-FM noise radar waveform. *IEEE Transactions on Aerospace and Electronic Systems*, 49(1): 658-664.
4. J. H. Lu, X. Han, J. X. Li, 2016. Consensus-based distributed fusion estimator with communication bandwidth constraints. *Control and Decision*, 31(12): 2155-2162.
5. M. Huang, L. T. Wang, H. C. Zhang, 2016. Face Recognition Based on Gabor Wavelet Transform and K-L Gaussian Riemannian Manifold Discriminant. *Computer Engineering*, 42(9): 208-213.
6. H. Hao, 2013. Multi component LFM signal detection and parameter estimation based on EEMD-FRFT. *Optik-International Journal for Light and Electron Optics*, 124(23): 6093-6096.
7. M. A. Govoni, H. Li, J. A. Kosinski, 2013. Range-Doppler resolution of the linear-FM noise radar waveform. *IEEE Transactions on Aerospace and Electronic Systems*, 49(1):658-664.
8. Y. Y. Fu, M. Zhang, D. G., Feng et al. 2014. Attribute privacy preservation in social networks based on node anatomy. *Journal of Software*, 25(4): 768-780.
9. D. G. Feng, M. Zhang, H. Li, 2014. Big data security and privacy protection. *Chinese Journal of Computers*, 37(1): 246-258.
10. J. Song, G. Y. Xu, R. P. Yao, 2016. Anonymized data privacy protection method based on differential privacy. *Journal of Computer Applications*, 36(10): 2753-2757.

11. C. Hazay, K. Nissim, 2012. Efficient set operations in the presence of malicious adversaries. *Journal of Cryptology*, 25(3):383-433.

12. Y. Pan, Y. Tang, H. Liu, 2012. Access control in very loosely structured data model using relational databases. *Acta Electronica Sinica,* 240(3): 600-606.

13. A. Rahman, H. Muhammah, L. Sungyoung, et al. 2015. Rough set-based approaches for discretization: a compact reviews. *Artificial Intelligence Review*, 44(2): 235-263.

14. Y. H. Qian, H. Zhang, Y. Sang, et al. 2014. Multigranulation decision-theoretic rough sets. *International Journal of Approximate Reasoning*, 55(1): 225-237.

15. F. Xu, J. J. Ma, 2015. Improvement of Threshold RSA Signature Scheme Based on Chinese Remainder Theorem. *JEIT*, 37(10): 2495-2500.

16. P. Curt, R. J. Thomas, S. Deming, 2012. A high-fidelity harmonic drive model. *ASME J of Dynamic Systems, Measurement, and Control*, 134(1): 457-461.

17. S. Ali, R. Ali, A. Iftikhar, 2017. Physico-chemical and microbiological assessment of some freshwater aquifers and associated diseases in district ghizer, gilgit-baltistan, Pakistan. Acta Scientifica Malaysia, 1(1): 08-12.

18. Y. Pan, C. A. Yuan, W. J. Li, M. H. Cheng, 2016. Access Control Method for Supporting Update Operations in Dataspace. *JEIT*, 38(8): 1935-1941.

19. M. J. Guo, Y. Huang, Z. Xie, 2013. A WebGIS Model Optimization Strategy under Multi-core Environment. Computer Engineering, 39(8): 15-19.

20. Y. Huang, J. Paisley, Q. Lin, et al. 2014. Bayesian nonparametric dictionary learning for compressed sensing MRI. *IEEE Transactions on Image Processing*, 23(12): 5007-5019.

21. L. Shen, G. Sun, Q. Huang, et al. 2015. Multi-level discriminative dictionary learning with application to large scale image classification. *IEEE Transactions on Image Processing*, 24(10): 3109-3123.

22. L. Shen, G. Sun, Q. Huang, et al. 2015. Multi-level discriminative dictionary learning with application to large scale image classification. *IEEE Transactions on Image Processing*, 24(10): 3109-3123.

23. J. J. Thiagarajan, K. N. Ramamurthy, A. Spanias, 2015. Learning stable multilevel dictionaries for space representations. *IEEE Transactions on Neural Networks & Learning Systems*, 26(9): 1913-1926.

24. J. Z. Jiang, F. Zhou, 2013. Iterative design of two-dimensional critically sampled MDFT modulated filter banks. *Signal Processing*, 93(11): 3124-3132.

25. N. Rajapakaha, A. Madanayake, Lt. 2014. Bruton, 2D space-time wave-digital multi-fan filter banks for sig[10]Gao, W. and W. Wang, The fifth geometric-arithmetic index of bridge graph and carbon nanocones. Journal of Difference Equations and Applications, 2017. 23(1-2SI): p. 100-109.

26. Gao, W., et al., Distance learning techniques for ontology similarity measuring and ontology mapping. Cluster Computing-The Journal of Networks Software Tools and Applications, 2017. 20(2SI): p. 959-968.nals consisting of multiple plane waves. *Multidimensional Systems and Signal Processing*, 25(1): 17-39.

27. H. Mahboubi, K. Moezzi, A. G. Aghdam, et al. 2014. Distributed deployment algorithms for improved coverage in a network of wireless mobile sensors. *IEEE Transactions on Industrial Informatics*, 10(1): 163-174.

28. N.S.A. Sukor, N. Jarani, S.F.M. Fisal, 2017. Analysis of Passengers' Access and Egress Characteristics to The Train Station. Engineering Heritage Journal, 1(2): 01-04.

29. S.C.A. Mana, M.M. Hanafiah, A.J.K. Chowdhury, 2017. Environmental characteristics of clay and clay-based minerals. Geology, Ecology, and Landscapes, 1(3): 155-161.

30. M. Bahmani, A. Noorzad, J. Hamedi, F. Sali, 2017. The role of bacillus pasteurii on the change of parameters of sands according to temperatur compresion and wind erosion resistance. Journal CleanWAS, 1(2): 1-5.

31. N.A. Rahman, Z. Tarmudi, M. Rossdy, F.A. Muhiddin, 2017. Flood Mitigation Measres Using Intuitionistic Fuzzy Dematel Method. Malaysian Journal Geosciences, 1(2):01-05.

32. R. Roslee, N. Simon, 2017. Rock Slopes Kinametic Analysis Along the Bundu Tuhan To Kundasang Highway, Sabah, Malaysia. Geological Behavior, 1(2):01–04.

CONTACT WITH THE AUTHOR

Jun Tian

e-mail: 276154989@qq.com

Foshan Polytechnic
Foshan 528137
CHINA