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].

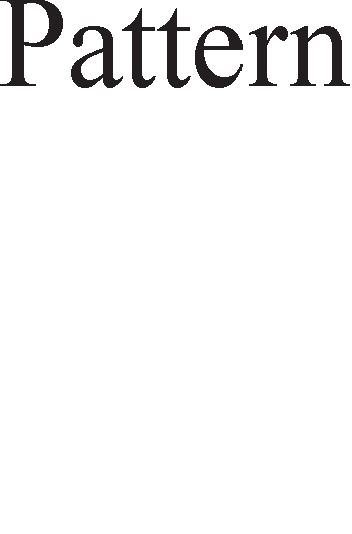
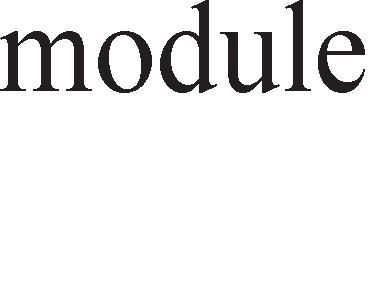
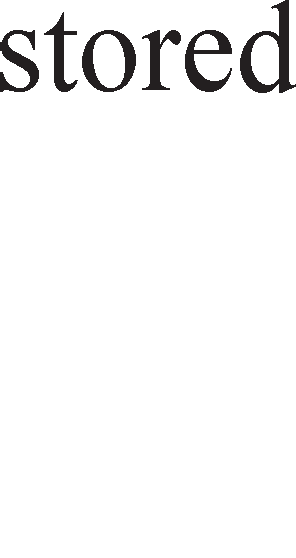
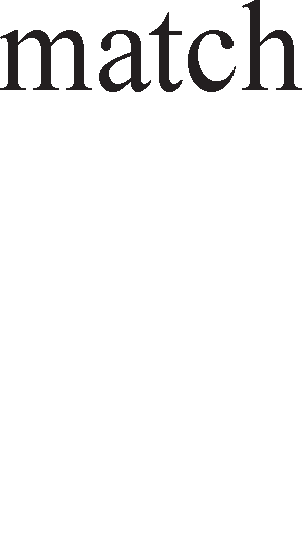
# BIG DATA STRUCTURE ANALYSIS AND FEATURE SELECTION OF MARINE GREEN ENERGY GRID-CONNECTED SYSTEM

## 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

two parameters in *m* ,** and* w* , another parameter can be obtained directly by * w*  (*m* 1)** .

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:

grid-connected system, as shown in Fig. 2.

*B* ( *A*)  min{*w*(** )  *w*( *AT*  ** )}

**(2)**

*l * 0

## BIG DATA INFORMATION COLLECTION MODEL

In formulas, *AT* represents the transpose of matrix *A* .

Simultaneous equation solving gets the solution of the

The Takens embedding theorem is used to reconstruct the

equations: *c*  *c*  0.675 , *c*  *c*  21/ 3 *c*  0.85 , for

1 1 2 2 1

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*  *R*2*d* 1 , which shows that

*s*, *s*  0 , there is the mapping *T* :*U*  *U* . The characteristic solution *u*(*t*) *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

(*z*)  (*h*(*z*), *h*(**1(*z*)),, *h*(**2*d* (*z*))) is an embedded vector.

*T*

in convex space:

For big data sampling time series {*x*(*t*0  *i**t*)} , *i*  0,1,, *N* 1of distributed marine green energy resources grid-connected system, its phase space reconstruction

max

*xa* ,*b* ,*d* , *p*

  *xa*,*b*,*d* , *pVp a*A *b*B *d*D *p*P

**(3)**

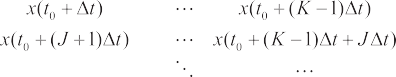
trajectory is:





**(1)**

s.t.



  *x Rbw* *K bw* S, *b*  B

*a*A *d*D *p*P

*a*,*b*,*d* , *p p b*

 *x Rcp* *Kcp* S, *d*  D

*a*,*b*,*d* , *p p d*

*a*A *b*B *p*P

 *xa*,*b*,*d* , *p* D*a*, *p* , *a*  A, *p*  P

*b*B *d*D

 *xa*,*b*,*d* , *p* *Ma*,*b* , *a*  A, *b*  B

**(4)**

**(5)**

**(6)**

Among them, *x*(*t*) represents the state vector of the embedding space. *J* is the reconstruction delay, *m* is the

embedding dimension, *t* is the time interval of sampling,

*d*D *p*P

 *xa*,*b*,*d* , *p*  *M a*,*d* , *p* , *a*  A, *d*  D, *p*  P

**(7)**

*K*  *N*  (*m* 1)*J* , and ** represents the delay parameter,

*b*B

**(8)**

**  *J* *t* . * w*  (*m* 1)**  (*m* 1)*J* *t* is time window. We can see that in reconstructing phase space, we only select any



*xa*,*b*,*d* , *p*

 0, *a*  A, *b*  B, *d*  D, *p*  P

**(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

*c*

 *ik*  1,*k*  1, 2,, *n*

*i* 1

**(15)**

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.

## BIGDATACLASSIFICATION ANDHIGHDIMENSIONAL

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:

## INFORMATION REORGANIZATION

We extract the feature vector of correlative dimension of

*ik*  1 *c*

( *ik*



*d*

*d*

*j* 1 *jk*

2

)*m*1

**(16)**

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

*m*

(**

*Vi*  *k* 1

*m*

*ik k*

) *x*

*n*

(*ik* )

*k* 1

*m*

**(17)**

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 ,, *xn* ,, we set the total number of points as *N* , the sampling time span of sequential {*xi* } is *j* , and its autocorrelation function is:

1 *N* 1

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* is the least square fitting vector, the feature solution *u*(*t*) *U*

in kernel convex spaces satisfies the maximum Lyapunove

*i*

*Rxx* ( *j* )   *xi xi*  *j*

*N*

*i* 0

**(10)**

functional, through construction of the fuzzy feature subset of the massive information, we get the iterative acceleration

Thus, we can fix *j* and make the correlation function about time** (taking**  1, 2, ), and the finite data set *X* is divided into c class according to the time delay of the big

formula of big data classification of marine green energy resources grid-connected system:

(1 **)*x* (*k* )  * x* (*k* 1)  *x* (*k* 1) , *i*  1, 2,, *n*

data embedding, among them 1<c<n , and the fuzzy clustering central matrixis defined as :

*i i i*

**(18)**

*V*  {*vij i*  1, 2,, *c*, *j*  1, 2,, *s*}

**(11)**

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

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

*U*  {*ik i*  1, 2,, *c*, *k*  1, 2,, *n*}

management in massive information is obtained under the control of attenuation constant *T*1 and *T*2 :

**(12)**

*x* (*k* 1)  (1 **)*x* (*k* )  **

 *b*  *i* 1 *a x* (*k* 1) 

*n a x* (*k* ) 

*i i a*  *i*

 *i j j*

 *i j j* 

Toconstructtheontology modelwhichreflectstheassociated feature of distributed marine green energy resources grid- connectedsystem big data [19], the relative weightof distributed marine green energy resources grid-connected system big

*i*  1, 2,, *n k*  1, 2,, *n*

*n i* 

*j* 1

*j* *i* 1

 **(19)**

data is **  ((** , *a* ＇), (** , *a* ),, (** , *a* ))*T* ，** 0,1 and

We use the support vector machine model to obtain the

1 1 2 2 n n j

we use the average mutual information method to obtain the objective function that the data association is the features extraction:

*n c*

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,

*J* (*U* ,*V* )   * m* (*d* )2

**(13)**

reducing the dimension of feature selection, so as to reduce

*m ik ik*

*k* 1 *i* 1

In the formula, m is the weight index.

(*dik*

)2 is the

the computational overhead of data fusion scheduling and operation and maintenance management [20]. According to

measuring distance between the sample is expressed by Euclidean distance:

*xk* and *Vi* , which

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

(*d* )2  *x* *V* 2

**(14)**

of recombination clusters of high dimensional information.

*ik k i*

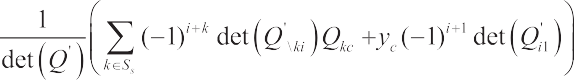
If the data set is *m* , order *Aj* (*L*) as the center of the cluster,

and

among them, *j*  1, 2,..., *k* . the distance to the cluster center is calculated;

1. select k examples in the data set, and initialise cluster centers *F* (*xi* , *Aj* (*L*)) , *i*  1, 2,..., *m* , *j*  1, 2,..., *k* ;



1. use the simple Euler distance to assign the remaining

clusters to the nearest cluster center, if satisfied:

*D*(*xi* , *Aj* (*L*))  min{*D*(*xi* , *Aj* (*L*))}

Thus *xi*  *k* ;

**(20)**

**(23)**

 *s*

1. in the massive information storage space, supposing

To build a set of homogeneous equations for *Hx* -global

a sample *i*  *Ss* , let **   , when the stationary vector set det(*Q*')  0 of the standard data set, through adaptive weighting of clustering center of massive data, when

*i*

*c*

strong stable functional, when the extremal functional

formulas of polynomial kernel function and Gaussian kernel

3

*i*  *Ss* , we can get **   . In convex optimization

*c*

*i*

function satisfy *d*  4, *sc*  2 , maximal linearly independent

space, we carry out the self-adaptive adjustment of data

clustering center, we can get: *i*  *S*  *Ss* , * c*   in the sample updating process of set *Ss* , we use the example of

*i*

set expressions of fuzzy data closed-loop operation and maintenance management respectively are :

each cluster to calculate the average value as the average

value of new cluster:

*K* (*xi* , *x j* )  *xi* , *x j* 

**(24)**

*k nj*

*C*(*l*)  (|| *x j*  *A* (*L*) ||)2

*k j*

*j* 1 *k* 1

**(21)**

*K* (*xi* , *x j* )  ( *xi* , *x j*

 1)*d*

**(25)**

1. if the average value is equal to the average value of the last iteration, || *C*(*l*)  *C*(*l* 1) ||﹤** ,the program is stopped, otherwise return to third step, let *l*=*l* 1 ,and the new

cluster heart is calculated:

*K* (*xi* , *x j* )  exp( *xi*  *x j*

2 / 2** 2 )

**(26)**

*k* For distributed marine green energy resources grid-

*A* (*L* 1)   *X j*

1

*nj*

*j i*

*n n*1

**(22)**

connected system big data sequence {*x* }*N* of single variable,

*i* 1

# BIG DATA CLOSED LOOP OPERATION

in the reconstruction of the *m* -dimensional phase space dimension, the distance form phase point *x j* to *xi* , besides *xi* itself is less than the *x j* points of *r* which is expressed as:

# AND MAINTENANCE MANAGEMENT OF

*Q*   *H* *r* 

*j* *i*

*xi*  *x j* 

**(27)**

# DISTRIBUTED MARINE GREEN ENERGY RESOURCES GRID-CONNECTED SYSTEM

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

*H* (*x*)  0,

1,



*x*  0

*x*  0

**(28)**

## 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

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)** , 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:

data closed-loop operation and maintenance management of

grid-connected system [21]. In the convex space, we build the

*Cm* (*r*) 

2 ∑*N* ∑*N H* *r* 

*xi*  *x j* 

**(29)**

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 *i*  *S* , * c*  

*N* (*N* 1) *i* 1 *j* *i* 1

When the amount of data is *N*  and the distance *r* is small, that is *r*  0 , if the correlation integral *Cm* (*r*) is obeyed by the exponential law, the correlation dimension features

of the big data of distributed marine green energy resources

*s i*

, and *i*  *S*  *S* , * c*   . For *i*  *S* , the the dual periodic grid-connected system is obtained as follows:

*s i s*

solitary wave solution of data closed-loop operation and maintenance management is:

*D*  ln *Cm*

(*r*)

**(30)**

ln *r*

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,

1. if it is red*y*uced to *d* dimension, then we take top *d* larger

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

eigenvectors *j* corresponding to *J*  *X j*  , the transformation matrix is generated:

*W*   *y*1, *y*2 ,, *yd* 

*j*  1, 2,, *d* and

**(34)**

is faster than other traditional analytical techniques and can be predicted [22].

In the closed-loop operation and maintenance management

1. considering the point which is satisfied *i*  *j*  ** , the

dimension reduction output results of the extracted feature vector of correlation dimension is obtained by K-L transform:

of fuzzy data, it is divided into data ETL layer, data storage layer, data analysis layer and application layer. [23]

\*

*X*



 *WT X*

**(35)**

## 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:

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:

**  (** ,** ,,** )  0

1. calculate the intra-class dispersion matrix

*S*ˆ*w* of *l*

1 2 *n*

**(36)**

-dimensional feature vector *X* (*l*, *ni* ) of distributed marine green energy resources grid-connected system big data, and find its *l* eigenvalues **1, **2 ,, *l* and eigenvectors matrix

*Y*  [ *y*1, *y*2 ,, *yl* ] . *S*ˆ*w* is:

1

*ni*

 (*i* )

 (*i* )



*T* 

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

* TQ* 

*n n*

* Q*

 0

*S*ˆ*w*   *pi*

*c*

*X k*

* *mi* *X k*
* *mi*  

**(31)**

 *i j ij*

**(37)**

*i* 1

Among them,

*ni k* 1  

*pi* represents the prior probability of *i*

*i* 1 *j* 1

Suppose there is *n* samples in the information data set

-th categories of data fusion scheduling and operation and maintenance management attribute. *ni* represents the number

*Ss* , then:

 0 *y*1  *yn* 

of samples of *i* -th categories of data fusion scheduling and

 *y Q*  *Q*  *def* 0

*yT* 

**(38)**

operation and maintenance management attribute. *m*

*Q*'  

1 11 1*n*    

*i*     

  *y Q* 

represents the data rep(*i* )resent the mean vector of *ni* -th sample

 *y Q*

 *Q* 

feature vector set*X k* , *k*  1, 2,, *ni*  of *i* -th categories;

 *n n*1

*nn* 

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

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:

*c*     *T*

*S* 

0  *yTQ*1  0 *yT*   *yTQ*1 0 

*b*  *pi* *mi*  *m**mi*  *m*

**(32)**

      

**(39)**

*i* 1





*c*

0 *In*

  *y Q*  

*y Q*

Among them, *m*   *pi mi* is the ensemble average;

*i* 1

1. calculate the average distance between classes:

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



*yT S y*

' *T* 1

*J*  *X j*   *j b j*

**

*j*

**(33)**

det *Q*   det *Q*. *y Q*   0

**`(40)**

And put them in descending order:

   ;

*J*  *X* 1   *J*  *X* 2     *J*  *X l* 

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

YUtility value   *f*  *xi* (*i*represents 1*n*-th Electric energy meter after NotionalPooling)

*f* (*xi* )  *j*  *xi*   *s*(*xi* )

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

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

*a x* (*k*1)  (1 **)*a x* (*k*)  **  *b*  *i* 1 *a x* (*k*1) 

*n a x* (*k* ) 

realize the self correlation feature mining of big data operation

*ii i ii i*

 *i* 

 *j* 1

*i j j*



*j* *i* 1

*i j j*





**(41)**

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

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].

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

 *D*  *L* *x*(*k* 1)  1-**  *D*+*U*  *x*(*k* )  *b*

*x*(*k* 1)   *D*  *L*1 1-**  *D*+*U*  *x*(*k* )  **  *D*  *L*1 *b*

Let:

**(42)**

**(43)**

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** =11 , sampling interval

*L*   *D*  *L*1 1-**  *D*+*U* 

is 0.25 s. The sampling points are 1000 points. There is a

 **  



 *f*

  *D*  *L*1 *b*

**(44)**

250Hz intra-class attribute difference unbalanced data data frequency component between the 400~600 sampling points.

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

We take the frequency domain characteristic parameters of big data distribution **  10 , *b*  8 / 3 , *r*  28 , and select the initial value [*x*0 , *y*0 , *z*0 ]  [2, 2, 20] , integration step

*x*(*k* 1)  *L x*(*k* )  *f*

**(45)**

*h*  0.01 , and use four-order Rung-Kutta method to solve

the characteristic equation of big data distribution for the

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

optimum selection control of clustering information fusion center feature. When carrying out the recursive analysis,

*x*  ** *x*  **

we take the parameters of phase space reconstruction *m* =3,

*k*  *k* 1

** =8, the threshold ** =0.2std ( *x* ), among them, STD ( *x* )

*xk* 1  **

*xk*  **

represents the standard deviation of the time series, *l* =3.

(*x*  ** )2  (*x*

 ** )(*x*

 ** )

min

*k k* 1

*k* 1

**(46)**

According to the simulation environment and parameter

*x* 2  2*ax*  ** 2  *x x*

 ** (*x*

 *x* )  ** 2

settings, the big data raw data sampling of the distributed

*k k k* 1 *k* 1

*k* 1

*k* 1

*x* 2  *x x*  ** (2*x*  *x*

* *x* )

marine green energy resources grid-connected system is

*k k* 1

*k* 1

*k* 1

*k* 1

*k* 1

carried out, and the time domain waveform of data sampling

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:

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

*x x*  *x* 2

(*x*  *x* )2

and maintenance manager for data fusion scheduling and

**  *k* 1 *k* 1 *k*  *x*

*k* 1 *k*

**(47)**

operation and maintenance management, to achieve big data

*x*  2*x*  *x*

*k* 1 *x*

* 2*x*  *x*

classification recognition of grid-connected system, and the

*k* 1

Among them:

*k k* 1

*k* 1

(*x*  *x* )2

*k k* 1

data classification results are shown in Fig. 4.

The analysis results of Fig. 4 show that using this method

*x*  *x*

*k* 1 *k*

**(48)**

for big data classification can realize data fusion scheduling

*k k* 1 *x*

*k* 1

*k k* 1

* 2*x*  *x*

and the operation and maintenance management, and can

The expected output of fuzzy K means clustering is:

effectively realize the classification and recognition of big data, and improves the adaptive analysis and processing ability

*E* ((*s*,(*s*) )  *m* *l p*(*x* ) *p*(*x* )*u*

*i j i i**N* \{ *j*}

*j* 1

*k* 1

*j*1

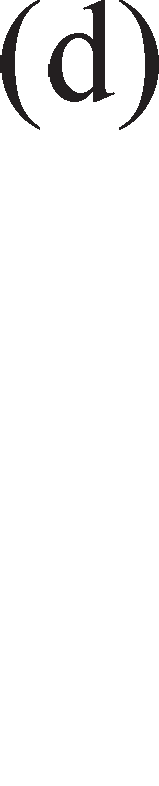
*jw*

*ki*

**(49)**

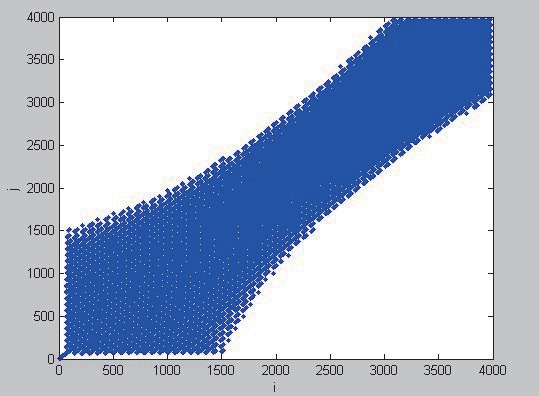
of big data of distributed marine green energy resources

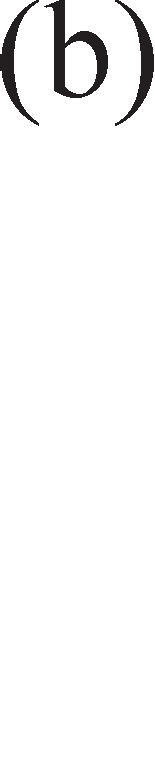
grid-connected system. In order to quantitatively analyze the application performance of the proposed method for big data classification processing, Table 1 lists the performance comparison of distributed marine green energy resources grid-connected system big data analysis in different methods [27-30]. 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].



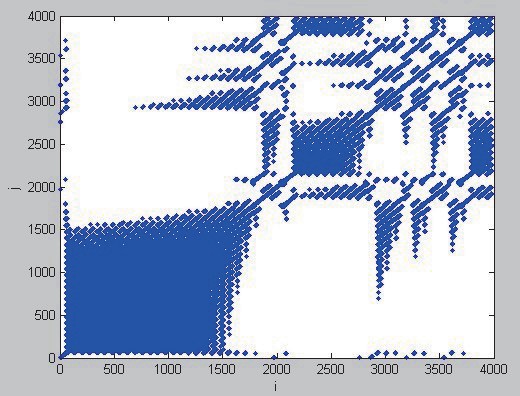
*Fig. 3. Time domain waveform of big data sampling of distributed marine green energy resources grid-connected system.*



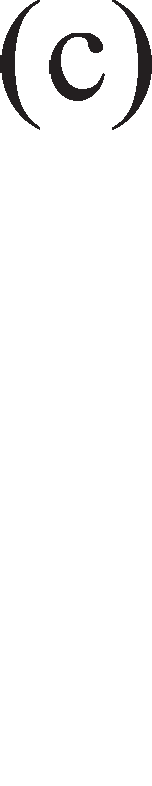
(a)sample1

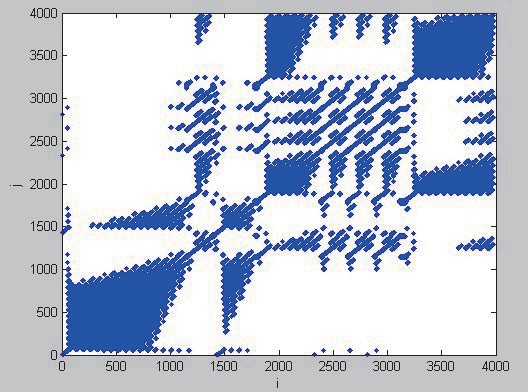


(b)sample2

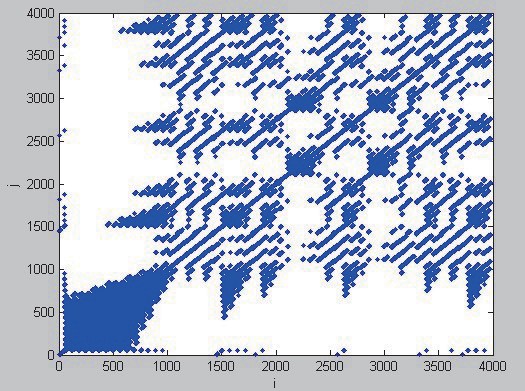
 

(c) sample3





(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 |