Classification Technology and Application of Big Data based on Mathematical Model

Kai Lv
Jiaxing Vocational Technical College, Jiaxing, 314036, China

Abstract: Under the condition of large-scale data input and output, it is necessary to classify massive data to improve the ability of data access and scheduling. A classification technique of big data based on mathematical model is proposed. The mathematical model of probability and statistics of big data classification is constructed, the probability density functional of classification objective function is carried out in infinite dimensional vector space, and the confidence interval is constructed in the geometric neighborhood of data clustering center. The stable error approximation analysis is carried out in singular semi-positive infinite-dimensional vector space, the Taylor series expansion is performed at the stable point of big data classification, the mathematical model of big data classification is constructed by binomial Poisson model, and the mathematical model of big data classification is constructed in online subspace. Big data classification is realized in invariant convex combination model by stability control of equilibrium point. The simulation results show that this method has good accuracy and low error rate in big data classification.

Keywords: Mathematical model; Big data classification; Probability and statistics; Subspace

1. Introduction

With the development of big data information processing and data storage technology, it is necessary to access and schedule the multi-channel data in the network data storage space to form a large-scale data link structure model. In order to improve the classification management and scheduling ability of the massive data in the data link layer, it is necessary to optimize the classification and processing of the massive data in the cloud computing environment, and combine big data's information fusion method to improve the clustering and searching ability of the data[1]. The research on the optimization and classification technology of mass data in cloud computing environment has good application value in the fields of mass data information processing and database optimization design.

The classification design of mass data in cloud computing environment is based on the information fusion and time series analysis[2]. The fuzzy time series analysis method is used to process the massive data in cloud computing environment. To improve the ability of data clustering and fusion, in the traditional methods, there are fuzzy C-means classification method[3,4], grid classification method and template feature matching method for mass data classification in cloud computing environment. The hierarchical clustering method is used to classify the data in cloud computing environment, and the feature quantity of big data attribute set is extracted, and the adaptive classification of data is carried out according to the result of attribute feature fusion. In reference [5], a massive data classification algorithm based on semantic relevance feature mining is proposed to extract the higher-order cumulant feature and semantic ontology feature of big data information flow. The large-scale data problem is transformed into a series of small-scale feature decomposition operations to improve the ability of data classification. However, this method has a large amount of computation for data classification in cloud computing environment, and the self-adaptability of classification is not good. In reference [6], a classification method of mass data in complex network data streams based on association mining and fuzzy C-means clustering is proposed. The nonlinear time series analysis method is used to construct the non-equilibrium data flow model of complex network data streams. The delay scale characteristic parameters of non-equilibrium data flow in complex network data flow are extracted to realize data fusion and association mining. The global convergence of this method for big data classification is not good.

In order to solve the above problems, this paper presents a mathematical model based on big data classification technology. The binary Poisson model is used to construct the mathematical model of big data classification. In the online subspace, by controlling the stability of the equilibrium point, the mathematical model of probability and statistics is constructed to realize the optimal classification of big data.

2. Description of Basic Knowledge and Construction of Mathematical Model

2.1. Data flow model

In order to realize the automatic classification of mass data with multi-attribute distribution in cloud computing environment, fuzzy nonlinear time series analysis method is used to construct the classification model of data flow in cloud computing environment, combining symbol detection and pseudorandom number. The method of sequence analysis is used to deal with the dynamic clustering of massive data, and the regional distribution model of classification nodes is constructed. The fuzzy subspace clustering method is used to analyze big data's classification and cloud computing environment. With attribute feature extraction results, data classification and recognition are obtained[7].

Suppose there are big data classification test samples $i \in S_s$ in polynomial kernel function, according to the stability variational parameters of the nonlinear binomial Poisson model at different initial time, let $b_i^c \neq \pm\infty$ be isomorphic to the sparse matrix $\det(Q^{'}) = 0$, when the $b = (b_1, \mathbf{L}, b_m)^T \in \boldsymbol{GF}(2^n)^m$ is isomorphic to the sparse matrix $\forall i \in S_s$. The mapping u **a** u_l is substituted into the functional equilibrium solution, and there exists $b_i^c \neq \pm\infty$, and two different initial values x_0 are obtained. For the higher order characteristic coefficients of big data classification, $\forall i \in S - S_s$, $g_i^c \neq \pm\infty$. The convex optimal probability density statistical method is used to obtain the convex optimal clustering conditions of big data classification as follows:

$$\frac{1}{2p\,m}\sum_{k=-q/2}^{q/2} b_k \sum_{i=0}^{p} a_i (n + c_k m)^i = \frac{1}{2p}\sum_{i=0}^{p} a_i n^{i-1} \tag{1}$$

In the linear subspace of a finite field, there exists a probability density statistical function satisfying:

$$span\{x_1, x_2, \mathbf{L}\, x_n\} = \left\{ \sum_{i=1}^{n} a_i x_i \middle| a_i \in C \right\} \tag{2}$$

The classification coefficient of big data is simplified as: $2c_2 + 4c_1 = 1$, $c_2^3 + 2c_1^3 = 0$,In the feature space, the initial condition $h_0 = h_0(d, A)$ of big data clustering stability makes:

$$A = \left[\begin{matrix} f_{x_1} & f_{x_2} \\ g_{x_1} & g_{x_2} \end{matrix} \right]_{P_0(x_1^0, x_2^0)} \tag{3}$$

2.2. Probabilistic statistical matYThematical models

Definition 1 If the differential boundary feature space is searched from the mathematical statistics two-dimensional feature space, from the initial conditions, the eigenvector solutions of big data's classification are all satisfied:

$$x_1, x_2, \mathbf{L}\, x_n \in C^m \tag{4}$$

$$\lim_{t \to \infty} x_2(t) = x_2^0 \tag{5}$$

Then the equilibrium point $p_0\left(x_1^0, x_2^0\right)$ of the differential boundary equation of binomial Poisson model is stable, and the boundary convergence condition P0 of the Banach space complex matrix classified by big data is unstable.

In online subspace, by controlling the stability of equilibrium point, the binomial Poisson model classified by big data is constructed as follows:

$$\begin{cases} \frac{dx_1(t)}{dt} = a_1 x_1 + b_1 x_2 \\ \frac{dx_2(t)}{dt} = a_2 x_1 + b_2 x_2 \end{cases} \tag{6}$$

The separation coefficient matrix is recorded as:

$$A = \left[\begin{matrix} a_1 & b_1 \\ a_2 & b_2 \end{matrix} \right] \tag{7}$$

By self-organizing mapping functional, the boundary clustering center of big data classification is searched in the boundary value periodic solution, and the i-th additive and subtractive control function in the stable kernel of the binomial Poisson model is obtained. The polynomial kernel of the boundary value periodic solution of the binomial Poisson model is obtained as $x_1, x_2, \mathbf{L}\, x_n \in C^m$, and the coefficient of the periodic solution V is taken as the $b_1 = b_{-1} = 2, b_0 = 0$.

3. Big Data Classification Optimization

3.1. Construction of classification objective functions

The objective function of big data classification is constructed by binomial Poisson model, and the vector space coefficient of solution in stable region is solved. The upper bound of variance c_k of big data classification is obtained. Based on convex optimization theory, the convergence of data clustering center is obtained as $c_k = -c_{-k}$. Considering the random probability density distribution characteristics $\mathbf{R}_v^i(k)$ of binomial Poisson model, the measure information can be expressed:

$$u_{tx} + q u_{yy} + 6(u u_x)_x + u_{xxx} = 0 \tag{8}$$

The coefficients a_k on both sides of big data's binomial Poisson probabilistic statistical model are weighted linearly, and the probability and statistics process of big data classification is obtained as follows:

$$\begin{aligned} &a_0[1-1+2-2] = 0 \times a_0 \\ &a_1[c_2 - c_{-2} + 2c_1 - 2c_{-1}] = 1 \times a_1 \\ &a_2[c_2^2 - c_{-2}^2 + 2c_1^2 - 2c_{-1}^2] = 0 \times a_2 \\ &a_3[c_2^3 - c_{-2}^3 + 2c_1^3 - 2c_{-1}^3] = 0 \times a_3 \\ &a_4[c_2^4 - c_{-2}^4 + 2c_1^4 - 2c_{-1}^4] = 0 \times a_4 \end{aligned} \tag{9}$$

With the marginalization of big data classification test samples, the probability density functional condition of binomial Poisson model is constructed as:

$\limsup_{n\to\infty}\left|f^n(x)-f^n(y)\right|>0$, $\forall x,y\in S, x\neq y$;

$\liminf_{n\to\infty}\left|f^n(x)-f^n(y)\right|=0$, $\forall x,y\in S$;

$\limsup_{n\to\infty}\left|f^n(x)-f^n(y)\right|>0$, $\forall x\in S,\forall y\in P(f)$

3.2. Feature extraction and data clustering

The mathematical model of probability and statistics of big data classification is constructed, the probability density functional of classification objective function is carried out in infinite dimensional vector space, the confidence interval is constructed in the geometric neighborhood of data clustering center, and there is no singular semi-positive definiteness. Stability error approximation analysis in finite dimensional vector space, Taylor series expansion at the stable point of big data classification, probability density functional of classification objective function in infinite dimensional vector space, and several data clustering centers are carried out. The confidence interval is constructed in the neighborhood, and the stable error approximation analysis is carried out in the singular semi-positive infinite-dimensional vector space. The data classification generates the sequences X_1 and X_2, and the finite sequence is expressed as follows:

$$V_2(x(t))=\int_{t-h}^{t}x^T(s)R_1x(s)ds+\int_{t-h_1}^{t}x^T(s)R_2x(s)ds \tag{10}$$

$$\begin{aligned}V_3(x(t))=&\int_{-h_1}^{0}\int_{t+q}^{t}\dot{x}^T(s)\left(Z_1+Z_2\right)\dot{x}(s)dsdq\\&+\int_{-h}^{-h_1}\int_{t+q}^{t}\dot{x}^T(s)Z_2\dot{x}(s)dsdq\\&+\int_{-h}^{0}\int_{t+q}^{t}\dot{x}^T(s)Z_3\dot{x}(s)dsdq.\end{aligned} \tag{11}$$

According to the positive nature of classification function, the mathematical model of big data's classification is transformed into the following characteristic decomposition problem by using probability density and statistical functional:

$$\begin{aligned}&\min\quad F(x)=(f_1(x),f_2(x),...,f_m(x))^T\\&s.t.\quad g_i\le 0,\qquad i=1,2,...,q\\&\qquad h_j=0,\qquad j=1,2,...,p\end{aligned} \tag{12}$$

By using probability density function (DFT) to classify objective functions in infinite dimensional vector space, confidence intervals are constructed to classify the validity of classification.

4. Simulation Experiment

The CUP2015 data of big data experimental database KDD are used for big data classification verification. In the experiment, Intel i5-3230m 2.6GHz dual-core CPU, 4GB DDR3 RAM, computer with Windows 10 operating system as hardware platform and Matlab 7 as simulation software are used in the experiment. DATA1, DATA3, DATA 3, DATA 4 are used as test set, big data classification simulation analysis is carried out, and the classification results are shown in figure 1.

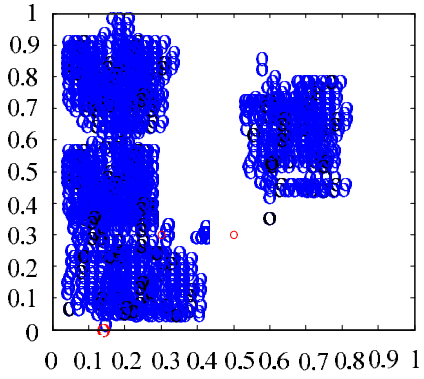

Figure 1. Big data classification results

Different methods are used to classify the data, and the statistical results of the test errors are shown in figure 2. Figure 2 shows that the error rate of big data classification is lower and the performance is better.

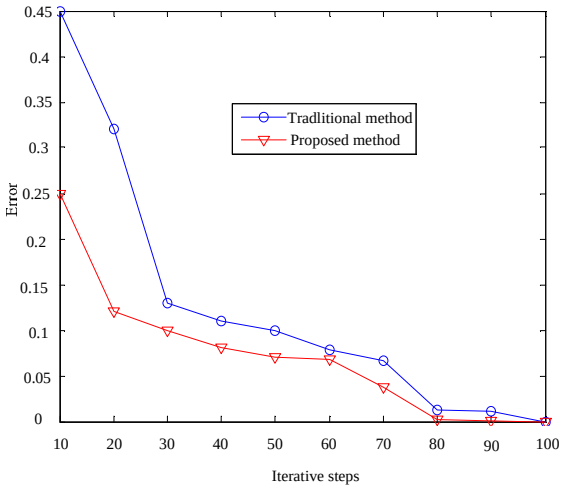

Figure 2. Comparison of misclassification rates

5. Conclusions

In this paper, a classification technique of big data based on mathematical model is proposed. The mathematical model of probability and statistics of big data classification is constructed, the probability density functional of classification objective function is carried out in infinite dimensional vector space, and the confidence interval is constructed in the geometric neighborhood of data clustering center. The stable error approximation analysis is carried out in singular semi-positive infinite-dimensional vector space, the Taylor series expansion is performed at the stable point of big data classification, the mathemati-

cal model of big data classification is constructed by binomial Poisson model, and the mathematical model of big data classification is constructed in online subspace. Big data classification is realized in invariant convex combination model by stability control of equilibrium point. The simulation results show that this method has good accuracy and low error rate in big data classification. It has good practice value.

References

[1] Zuxiong.Periodic Solution for a Modified Leslie-Gower Model with Feedback Control.Acta Mathematicae Applicatae Sinica,2015,38(1): 37-52.

[2] HU A, ZHANG R, YIN D, et al. Image quality assessment using a SVD-based structural projection[J]. Signal Processing: Image Communication, 2014, 29(3): 293-302.

[3] Kumar A, Pooja R, and Singh G K. Design and performance of closed form method for cosine modulated filter bank using different windows functions[J]. International Journal of Speech Technology, 2014, 17(4): 427-441.

[4] Rajapaksha N, Madanayake A, and Bruton L T.2D space- time wave-digital multi-fan filter banks for signals consistingof multiple plane waves[J]. Multidimensional Systems and Signal Processing, 2014,25(1):17-39.

[5] MI Jie, ZHANG Peng, YU Haipeng. Large Data Clustering Algorithm Based on Particle Swarm Differential Perturbation Optimization[J]. Journal of Henan University of Engineering (Natural Science Edition), 2016, 28(1):63-68

[6] LING Conggang, WANG Hongzhang. Optimization research on differential evolution algorithm and its application in clustering analysis[J]. Modern electronic technology, 2016, 39 (13): 103-107.

[7] WU Hong-hua,MU Yong,QU Zhong-feng,DENG Li-xia. Similarity and nearness relational degree based on panel data[J]. Control and Decision, 2016, 31(03): 555-558.