



Research on Reduced Dimension Classification Algorithm of Complex Attribute Big Data in Cloud Computing

Wei Song^(✉) and Yue Wang

Software College and Nanyang Institute of Technology, Nanyang, China
songwei2462@163.com, wangyue6651@163.com

Abstract. In order to improve the ability of data retrieval in cloud computing environment, a reduced dimension classification algorithm of complex attribute big data in cloud computing based on deep neural network learning is proposed. The complex attribute big data under cloud computing is constructed by low dimensional feature set, and the complex attribute big data under cloud computing is analyzed by linear programming and fitting using grid clustering method. Big data samples of all complex attributes are sampled and trained to extract the associated features of big data, which is a complex attribute under cloud computing. The feature extraction results of complex attribute big data under cloud computing are inputted into the deep neural network learner for data classification, and the complex attribute big data dimensionality reduction classification under cloud computing is realized by combining big data fusion clustering method. The simulation results show that the accuracy of big data dimension reduction classification for complex attributes in cloud computing is high and the error rate is small.

Keywords: Cloud computing · Complex attribute big data · Dimensionality reduction classification

1 Introduction

With the development of cloud computing technology, the stability of cloud computing database becomes higher and higher. In the process of data transmission using cloud computing, it is affected by the interference of transmission channel and inter-symbol characteristic disturbance, which results in complex attributes of cloud computing [1]. It reduces the output accuracy of cloud computing database, and needs to effectively mine and classify complex attributes under cloud computing, and carries out intelligent dimensionality reduction classification combined with the categories of complex attributes in cloud computing to improve the ability of intelligent dimension reduction classification of cloud computing. In order to ensure the security, stability and transmission efficiency of cloud computing, it is of great significance to study the classification algorithm of cloud computing complex attribute data in order to reduce the dimension of intelligent complex attributes in cloud computing database [2].

The classification of complex attribute big data under cloud computing is realized by data mining and feature extraction, and the associated feature quantity of complex

attribute big data under cloud computing is extracted, and the anti-jamming algorithm is used to filter and detect [3]. Combined with complex attribute data classification, intelligent dimensionality reduction classification is carried out to improve the stability of cloud computing database communication and ensure communication quality. In traditional methods, the classification method of complex attribute big data under cloud computing mainly includes Reverse KNN method. Fuzzy C-means classification method, support vector machine classification method and BP neural network classification method were used to classify data using deep neural network learning, expert system, statistics and other classification models. The classification and prediction ability of complex attributes was improved. In references [4], a complex attribute big data classification algorithm based on ART model and Kohonen prediction in cloud computing was proposed. Big data, a complex attribute to be classified in cloud computing, was used to extract association rule features and big data to mine, and fuzzy clustering method was combined to classify complex attribute data under cloud computing to improve the accuracy of classification. However, the real-time classification of this method was not good. The efficiency of data retrieval in cloud computing environment was not high. In reference [5], a technology of complex attribute big data classification based on fuzzy C-means mathematical classification method was proposed. The distributed storage design of complex attribute large database under cloud computing was carried out by using grid topology. The semantic autocorrelation function analysis method was used to cluster the nearest neighbor points of complex attribute big data in cloud computing. This method had poor anti-interference ability in large-scale cloud computing database data classification [6].

In view of the disadvantages of traditional methods, this paper proposes a reduced dimension classification algorithm for complex attribute big data in cloud computing based on deep neural network learning. Firstly, big data, a complex attribute collected in cloud computing, is constructed with a low-dimensional feature set, and a large database of complex attribute distribution under cloud computing is constructed. Then the grid clustering method is used to fit the complex attribute big data in cloud computing, and the disturbance of data clustering center is analyzed by combining K-means algorithm and nearest neighbor algorithm. The feature extraction results of big data, a complex attribute under cloud computing, are classified. Finally, simulation experiments are carried out to demonstrate the superior performance of the proposed method in improving the ability of dimensionality reduction of big data, a complex attribute in cloud computing.

2 Data Preprocessing

2.1 Large Database Construction with Complex Attribute Distribution

In order to realize the dimensionality reduction classification of complex attribute big data in cloud computing, the fuzzy rough clustering method is used to construct cloud computing distributed large database model, and the nearest neighbor priority distributed information mining method is used to mine complex attribute big data in cloud computing. The adaptive association rule scheduling method is used to detect and filter

complex attribute data in cloud computing, and a distributed large database model of complex attribute big data under cloud computing is constructed by integrating correlation detection method. The data set is vectorized and the frequent itemsets of complex attribute data under cloud computing are calculated under the uncertain data frequent itemsets pattern. The fusion analysis method of expected frequent term (EFI) and probabilistic frequent item (PFI) is adopted [7]. The scheduling set function of big data, a complex attribute distribution under cloud computing, is obtained as follows:

$$R_d^i(t+1) = \min\{R_s, \max\{0, R_d^i(t) + \beta(n_i - |N_i(t)|)\}\} \quad (1)$$

$$N(t) = \{j : \|x_j(t) - x_i(t)\| < R_d^i; l_i(t) < l_j(t)\} \quad (2)$$

Where, $x_j(t)$ represents the classification information entropy in data set D, describes the sample subset of the $l_j(t)$ cluster center, and t represents the sample set of the generation learning in the process of complex attribute data classification. The output label attributes of complex attribute data in the cluster center are calculated under cloud computing. The statistical characteristic quantity of complex attribute data under cloud computing is analyzed by using split information detection method. The storage sample database model of complex attribute data is obtained by using scalar sequence analysis method:

$$AVG_X = \frac{1}{m \times n} \sum_{x=1}^n \sum_{y=1}^m |G_X(x, y)| \quad (3)$$

Where, m, n are the category number and sampling node of sample samples of complex attribute data in cloud computing respectively. Let p_i be the uncertain database and S be the classification element of complex attribute data under cloud computing. The statistical distribution probability of massive cloud computing data sampling is H , and the distribution width of complex attribute data under cloud computing is obtained by mining frequent itemsets:

$$\text{sgn}(z_R^2(k) - R_{MDMMA_R}) = \text{sgn}(z_R^2(k) - \hat{e}_R^2(k)) \quad (4)$$

$$\text{sgn}(z_I^2(k) - R_{MDMMA_I}) = \text{sgn}(z_I^2(k) - \hat{e}_I^2(k)) \quad (5)$$

According to the above analysis, the distributed storage design of complex attribute large database under cloud computing is carried out by using grid topology, and the vector quantization feature coding model of complex attribute big data under cloud computing is constructed. The feature distribution gradient map of complex attribute big data under massive cloud computing is extracted, and the complex attribute data classification is carried out by combining sample statistical average analysis and depth neural network learning algorithm [8].

2.2 Data Sample Regression Analysis and Fusion Processing

On the basis of constructing the large database model of complex attribute distribution under cloud computing, a small amount of sample class data is taken as test set, and the complex attribute big data under cloud computing is analyzed by linear programming fitting method with grid clustering method. In the fuzzy grid clustering center, the expected support degree $esup(D)$ of the data element t is greater than the threshold H , then the attribute element of the complex attribute big data classification under cloud computing is called a frequent term, that is, The classified attribute elements of all complex attribute data that satisfy the constraints satisfy:

$$esup^t(D) > \theta \tag{6}$$

The clustering result of complex attribute big data in cloud computing is modulated adaptively. If the element t of complex attribute category satisfies the finite scheduling mode, it is called probability frequent term.

$$\sum_{\omega \in PW, C^i(\omega) \geq minsup} P[\omega] > \delta \tag{7}$$

According to the sequence of information gain ratio values from large to small, the method of frequent item mining is used to analyze the threshold value of δ , when the probability of complex attributes is maximum, and then randomly find a point to repeat the above steps [9]. Considering the probability that an element is a frequent item throughout the possible instance space, the cluster center point summary output is as follows:

$$x_i(k+1) = x_i(k) + s \left(\frac{x_j(k) - x_i(k)}{\|x_j(k) - x_i(k)\|} \right) \tag{8}$$

Where: $\|\vec{x}\|$ denotes the norm of \vec{x} . Thus, the statistical regression analysis and sample test of complex attribute data are realized.

3 Dimensionality Reduction Classification of Complex Attribute Big Data in Cloud Computing

3.1 Cloud Computing Complex Attribute Big Data Grid Clustering Method

In this paper, we propose a dimensionality reduction classification algorithm for complex attribute big data in cloud computing based on deep neural network learning.

Deep neural network (DNN) is a kind of software developed by Microsoft, whose main purpose is to imitate the way of human thinking, with fast classification ability and high accuracy. Its main structure is shown in Fig. 1:

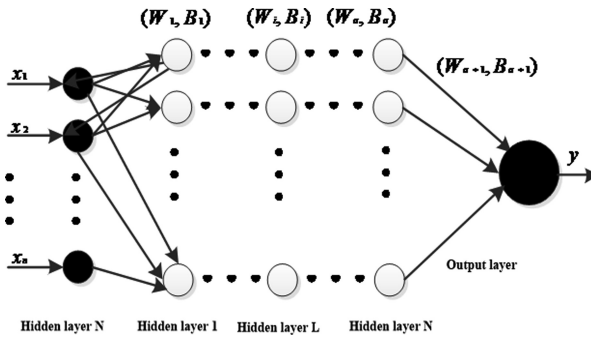


Fig. 1. Deep neural network structure

As can be seen from Fig. 1, the deep neural network consists of three parts: input layer, multi-hidden layer and input layer. Compared with traditional neural network, deep neural network (DNN) has many hidden structures. X represents the network input, including the column vector value of dimension m , and (W, B) represents the matrix formed by the weight and threshold between the hidden layers. The deep neural network USES the vector value obtained from the previous layer of each hidden layer to carry out the nonlinear transformation of function activation, and then transfers the obtained value to the next neuron, which iterates successively, and finally transfers it to the network output y . Compared with traditional neural network, DNN has significantly higher depth of multiple hidden layers, which can make up for the shortcomings of traditional neural network.

The correlation layer of the deep neural network contains a storage module that is favorable to the historical information of the storage layer. The storage unit reserves the historical information of the current time point as the input value of the hidden layer of the first layer in the next time period. The prediction model based on the deep neural network makes the internal structure and state of the network at the best level, and the final output is not only related to the current time period, but also related to the historical information within the time period, which promotes the prediction model to have better dynamic memory ability.

In order to calculate the probability of frequent items, the complex attribute big data under cloud computing is analyzed by linear programming using grid clustering method. This paper introduces the concept of complex attribute occurrence probability of complex attribute big data under cloud computing and clustering frequency distribution of complex attribute data. t is used to represent the probability of different frequency of element in large database of complex attribute under cloud computing, which is called $sup^t(D)$. Then the fuzzy iterative inequality of complex attribute big data's complex attribute grid clustering in cloud computing can be transformed into:

$$\sum_{i=minsup}^{num^t(D)} sup^t(D) > \delta \tag{9}$$

Wherein, $num^t(D)$ is the maximum number of cluster analysis elements of complex attribute big data in the complex attribute big data sample distribution database under cloud computing. For the calculation of $sup^t(D)$, big data's reduced dimension classification global search method is used to carry out the dynamic programming of complex attribute data classification. The calculation formula is expressed as follows:

$$P_{ij}^t = \begin{cases} P_{i-1,j-1}^t \times p_i + P_{i-1,j}^t \times (1 - p_i), & v_i = t \\ P_{i-1,j}^t, & v_i \neq t \end{cases} \quad (10)$$

The probability of all possible instances in the current complex attribute big data's dimensionality reduction classification window is calculated. The fuzzy learning iteration of big data's dimension reduction classification is as follows:

$$r_d^i(k + 1) = \min\{r_s, \max\{0, r_d^i(k) + \beta(n_i - |N_i(k)|)\}\} \quad (11)$$

Wherein: β denotes the associated feature quantity of global search in big data's reduced dimension classification, and the first part of the complex attribute big data sample set indicates that element t appears on the H element. That is, in the first $i - 1$ complex attributes big data reduced dimension classification attribute element $j - 1$ only appears the statistical probability of S times, taking a small amount of sample class data as test set, the fuzzy random number analysis of complex attribute big data under cloud computing is carried out by grid clustering method, and the fuzzy random number of complex attribute big data under cloud computing is analyzed by grid clustering method. The cloud computing under complex attributes big data grid clustering is realized [10] (Fig. 2).

3.2 Deep Neural Network Learning and the Optimization of Big Data's Dimensionality Reduction Classification Steps for Complex Attribute

The feature extraction results of complex attribute big data under cloud computing are inputted into the deep neural network learner for data classification, and the complex attribute big data dimension reduction classification under cloud computing is realized by combining big data fusion clustering method. Set the following algorithm steps based on the above ideas:

Input: under the cloud computing complex attribute big data uncertain data flow DS, cloud computing complex attribute big data association sample threshold δ , statistical distribution probability threshold $minsup$, cloud computing complex attribute big data sampling window length W ;

Output: frequent item sets for support vector machine learning D

1. Parameter of initialization machine and classification coefficient of complex attribute data: $SWF = null, D = null, P_{ij} = 0, sup^{ki}(\omega) = 0$.
2. for X_{ij} , a random point is found and the central point of all clusters of the complex attribute big data's dimension reduction classification is obtained.
3. Calculate the probability of cluster crossing P_{ij} ;

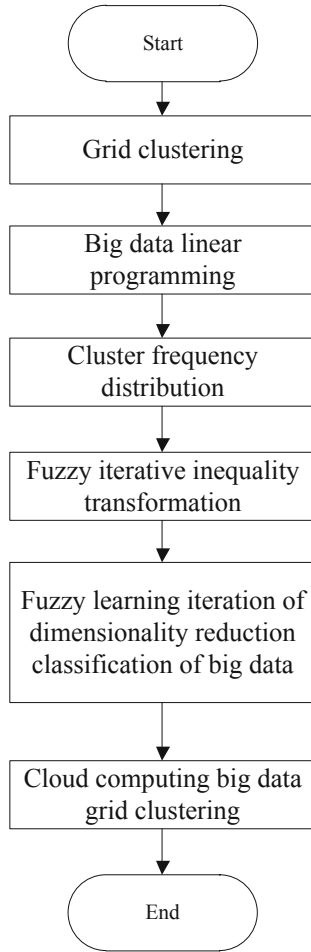


Fig. 2. Algorithm flow

4. if (the current window is not full), the nearest neighbor priority absorption method is used to reconstruct complex attribute features.
5. Update the sample of complex attribute big data under cloud computing in the current window and calculate the probability distribution value of complex attribute category $sup^{ki}(\omega)$;
6. From the complex attribute big data sample set which exceeds the frequency threshold in the data set, the statistical characteristic quantity is obtained by combining the cumulative probability distribution method $Q = \sum_{i=minsup}^{num^l(D)} sup^l(D)$;
7. if $Q \geq \delta$
8. Big data samples of complex attributes under cloud computing based on deep neural network learning are added to the set of frequent items D ;

9. Sample regression analysis to store complex attribute data in window set SWF ;
10. Find out-of-date sample elements and delete;
11. All the complex attribute big data samples are sampled and trained to update the probability distribution value of the window $sup^{ki}(\omega)$.

4 Simulation Experiment and Result Analysis

In order to test the performance of this method in the reduced dimension classification of complex attribute big data under cloud computing, the simulation experiment is carried out. The experiment adopts Matlab 7 and C joint simulation design, and the sample size of big data, a complex attribute of cloud computing database, is 1000 Mbit. The training sample set is 1024, the time width of data sampling is 10 s, and the time domain waveform of data sampling is shown in Fig. 3.

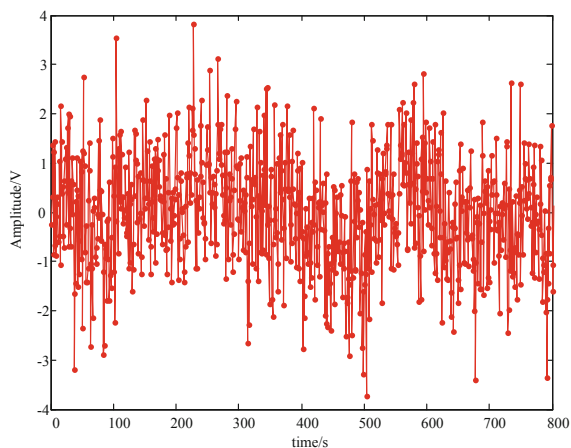


Fig. 3. Sample of complex attribute big data research objects under cloud computing

Set $W = 1000$, $minsup = 100$, set the frequency threshold of complex attribute big data collection under cloud computing: $minsup = 2$, when the probability threshold of complex attribute big data distribution is $\delta = 0.3$, the mining results of association rule feature of complex attribute category element 4 and frequent item can be found in Table 1.

According to the association rule mining results of complex attribute big data under cloud computing, the complex attribute big data dimension reduction classification under cloud computing is carried out, and the classification probability is calculated as shown in Table 2.

Table 1. Mining results of complex attribute big data association rules in cloud computing

Instance	P	Instance	P
{}	0.0653	{D ₂ }	0.0435
{D ₁ }	0.0454	{D ₂ , D ₃ }	0.0675
{D ₁ , D ₂ }	0.0664	{D ₂ , D ₄ }	0.0654
{D ₁ , D ₃ }	0.2434	{D ₂ , D ₃ , D ₄ }	0.0376
{D ₁ , D ₄ }	0.0564	{D ₃ }	0.1564
{D ₁ , D ₂ , D ₄ }	0.0433	{D ₃ , D ₄ }	0.0646
{D ₁ , D ₃ , D ₄ }	0.0456	{D ₁ , D ₂ , D ₃ , D ₄ }	0.0356
{D ₁ , D ₂ , D ₃ }	0.1564	{D ₄ }	0.0686

Table 2. Reduced dimension classification probability of complex attribute big data in cloud computing

Frequency	Complex attribute probability distribution
0	0.765
1	0.873
2	0.297
3	0.554

The results of Tables 1 and 2 show that the method proposed in this paper can effectively realize the dimensionality reduction classification of complex attributes big data under cloud computing, and the accurate probability of dimension reduction classification detection for complex attributes big data is high. The accuracy of data classification of sample sets D1 and D2 is tested, and the comparison results are shown in Fig. 4.

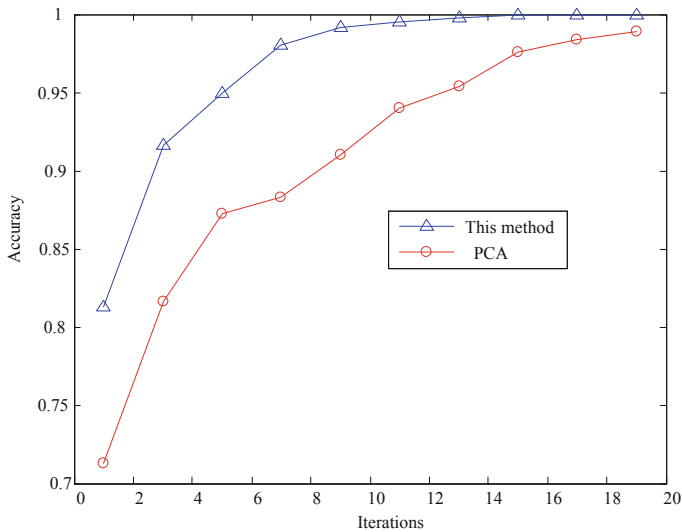


Fig. 4. Comparison of accuracy of big data dimension reduction classification for complex attributes in cloud computing

The analysis of Fig. 4 shows that the accuracy of big data dimension reduction classification of complex attributes under cloud computing is high and the error rate is low.

5 Conclusions

In this paper, a reduced dimension classification algorithm of complex attribute big data in cloud computing based on deep neural network learning is proposed. The complex attribute big data under cloud computing is constructed by low dimensional feature set, and the complex attribute big data under cloud computing is analyzed by linear programming and fitting using grid clustering method. Big data samples of all complex attributes are sampled and trained to extract the associated features of big data, which is a complex attribute under cloud computing. The feature extraction results of complex attribute big data under cloud computing are inputted into the deep neural network learner for data classification, and the complex attribute big data dimensionality reduction classification under cloud computing is realized by combining big data fusion clustering method. The simulation results show that the accuracy of big data dimension reduction classification for complex attributes in cloud computing is high and the error rate is small. The method presented in this paper has a good application value in the reduced dimension classification of big data in cloud computing.

References

1. Han, D., Chen X., Lei, Y., et al.: Real-time data analysis system based on Spark Streaming and its application. *J. Comput. Appl.* **37**(5), 1263–1269 (2017)
2. Sun, D.W., Zhang, G.Y., Zheng, W.M.: Big data stream computing, technologies and instances. *J. Softw.* **25**(4), 839–862 (2014)
3. He, H., Tan, Y.: Automatic pattern recognition of ECG signals using entropy-based adaptive dimensionality reduction and clustering. *Appl. Soft Comput.* **55**, 238–252 (2017)
4. Yuelong, Z., Xiaoxiao, Z., Jimin, W.: Time series motif discovery algorithm based on subsequence full join and maximum clique. *J. Comput. Appl.* **39**(2), 414–420 (2019)
5. Tu, G., Yang, X., Zhou, T.: Efficient identity-based multi-identity fully homomorphic encryption scheme. *J. Comput. Appl.* **39**(3), 750–755 (2019)
6. Wang, Z., Huang, M., et al.: Integrated algorithm based on density peaks and density-based clustering. *J. Comput. Appl.* **39**(2), 398–402 (2019)
7. Farnadi, G., Bach, S.H., Moens, M.F., et al.: Soft quantification in statistical relational learning. *Mach. Learn.* **106**(12), 1971–1991 (2017)
8. Tu, B., Chuai, R., Xu, H.: Outlier detection based on K-mean distance outlier factor for gait signal. *Inf. Control* **48**(1), 16–21 (2019)
9. Wei, X.S., Luo, J.H., Wu, J.: Selective convolutional descriptor aggregation for fine-grained image retrieval. *IEEE Trans. Image Process.* **26**(6), 2868–2881 (2017)
10. Liu, Q., Guan, W., Li, S., Wang, F.: Indoor WiFi-PDR fusion location algorithm based on extended Kalman filter. *Comput. Eng.* **45**(4), 66–71 (2019)