Feature Detection Algorithm of Chaotic Data in Distributed Networks

Ning Pan*
Department of Information Development and Management, Hubei University, Wuhan, 430062, China
*Corresponding author e-mail:panda20087@Hotmail.com

Abstract. The chaotic data in the distributed network is affected by the disturbance information of the network nodes, which leads to the poor convergence of the data feature detection. A new algorithm for the chaotic data feature detection in the distributed network is proposed based on fuzzy C-means clustering. According to the association rule attributes of the chaotic data in the distributed network, the principal component analysis is carried out, and the semantic attributes and fuzzy clustering features of the chaotic data in the distributed network are extracted. The joint association rule analysis method is used to deal with the interference filtering of the chaotic data in the distributed network, and the classified fuzzy set of the chaotic data in the distributed network is constructed based on the sensor fusion tracking measure method. In the fuzzy data set, the mixed weighting and adaptive block matching of chaotic data in distributed network are carried out, and the association rule feature quantity of chaotic data in distributed network is extracted. The extracted features are input into the fuzzy C-means clustering classifier for data classification and recognition, and the optimized feature detection of the chaotic data in the distributed network is completed. The simulation results show that the proposed method has good data convergence, strong convergence and anti-interference in the process of feature detection in distributed networks.

Keywords: Fuzzy C-Means Clustering, Chaotic Data In Distributed Networks, Feature Detection, Feature Extraction

1. Introduction
Database technology and system have become an important support for the construction of information society infrastructure, and the problem of data quality has gradually gained the attention and attention of people. Recent statistics show that American companies spend billions of dollars a year on dirty data. Chaotic data feature detection in distributed networks is a dense and complex process. It is estimated that about 30% of the time in a typical data warehousing project is used to detect chaotic data features in distributed networks. The
consistency of data is an important topic in the problem of feature detection of chaotic data in distributed networks, so it also plays a key role in data quality\cite{1}. A key technical problem in data management is the potential conflict of information sources. These conflicts will be reflected at different levels: conflicts of relational patterns, conflicts of data representation, conflicts of data values. The inconsistency of data is often used to describe these conflicts. Scrambling data detection in distributed networks is one of the core issues in data quality and feature detection of chaotic data in distributed networks\cite{2}.

At present, most of the researches on feature detection of chaotic data in distributed networks focus on the problem of merging and removing similarity data, or the problem of detecting domain differences and structural conflicts of data\cite{3}. The consistency of data is expressed by making constraint conditions, and the inconsistency of data is detected when the constraint condition is violated. Most of the existing work is based on traditional dependency constraints, such as functional dependencies or full dependencies. Traditional dependency constraints are mainly formed or generated for the design of relational patterns, and are often not sufficient to cover the semantic relationships contained in data. In reference \cite{4}, the traditional functional dependency constraints are extended, in which the conditional function dependency (Conditional Functional Dependencies, CFDs) is proposed to describe the inconsistent constraints. In the traditional centralized database, given a CFDs constraint set, only a fixed number of SQL queries can automatically find out the set of tuples in the database that have violated the constraint condition in polynomial time\cite{4}. This SQL technique is often used to detect inconsistency under eCFDs constraints, in which eCFDs is an effective extension of CFDs to support logical disjunction and logic negation\cite{5}. However, this SQL technology can not solve the inconsistency detection of distributed data, and this topic is far more challenging than the centralized data domain. In addition, in references \cite{6}, it extend the distributed data based on CFD, it defined the partition of data, normalize the distributed detection problem into optimization problems, and propose some effective detection algorithms. In order to improve the capability of feature detection of chaotic data in distributed networks, this paper proposes an algorithm for feature detection of chaotic data in distributed networks based on fuzzy C-means clustering. According to the association rule attributes of the chaotic data in the distributed network, the principal component analysis is carried out, and the semantic attributes and fuzzy clustering features of the chaotic data in the distributed network are extracted\cite{7}. The joint association rule analysis method is used to deal with the interference filtering of the chaotic data in the distributed network, and the classified fuzzy set of the chaotic data in the distributed network is constructed based on the sensor fusion tracking measure method. In the fuzzy data set, the mixed weighting and adaptive block matching of chaotic data in distributed network are carried out, and the association rule feature quantity of chaotic data in distributed network is extracted. The extracted features are input into the fuzzy C-means clustering classifier for data classification and recognition, and the optimized feature detection of the chaotic data in the distributed network is completed. Finally, the simulation results show that the proposed method can improve the capability of feature detection of chaotic data in distributed networks.
2. Feature Analysis and Extraction of Chaotic Data in Distributed Networks

2.1. Data Principal Component Analysis

In order to detect the optimal feature of the chaotic data in the distributed network, firstly, the similarity information of the chaotic data in the distributed network is analyzed, and the decision tree model of C classification is adopted. Decompose the similarity of chaotic data in distributed network [8], as shown in figure 1.

![Figure 1. Decision tree model of chaotic data similarity decomposition in distributed networks](image-url)

According to the decision tree model shown in figure 1, the mixed feature recognition and data classification of chaotic data in distributed network are carried out, and the mixed attribute fuzzy classification model of chaotic data in distributed network is constructed. According to the association rule attribute of data, the principal component analysis is carried out, and the segmented attribute set of fuzzy information is decomposed by singular value (SVD):

\[
X = UDV^T
\]  

(1)

Where, the feature matrix of ontology mapping is \( U \in R^{m \times n}, \ V \in R^{M \times M}, \) denotes the closed frequent items matrix of chaotic data in distributed network, and \( U^T = U^{-1}, \ V^T = V^{-1}, \) \( D \in R^{n \times M}, \) satisfies \( D = \sum \ 0 \), in ontology mapping. The weighted value of distributed feature quantity of chaotic data in distributed network is as follows. The distributed matrix \( \sum = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n}) \), \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \), the semantic concept set of chaotic data in distributed network is obtained. The non-zero eigenvalue is taken as the training subset, and the data flow model is reconstructed. The mixed similarity feature analysis method is used to analyze the distributed network. The phase space reconstruction of the chaotic data in the complex is carried out\(^9\), and the average mutual information characteristic expression of the output after the phase space reconstruction is obtained as follows:
\[ I(Q,S) = \sum_{i} \sum_{j} p_{ij}(s_i, q_j) \log \left( \frac{p_{ij}(s_i, q_j)}{p_i(s_i)} \right) \] (2)

The information distribution model in a cluster of chaotic data in a distributed network is defined as the closed frequent term of fuzzy information obtained by \([s, q] = [x(t), x(t+\tau)]\). At the fuzzy centroid, a random sequence sample of \(s\) representing the chaotic data flow in a sampled distributed network is used[10]. Assume that the initial symbol for each chaotic data classification attribute value is the data object and cluster center distribution model of the chaotic data in the \(C_0 = C_{N/2} = 0\), \(C_{N-x} = C^*\), distributed network is shown as follows:

\[
P_s = \frac{p_i}{(4\pi)^{\frac{d}{2}}} \left[ 1 + \alpha^2 + 2\alpha \cos \left( \frac{4\pi d^2}{\lambda} \right) \right] \] (3)

According to the difference of different attributes in clustering, the feature recognition of abnormal data is carried out, and the probability density function of accurate recognition is obtained as follows:

\[
P_s = p_{2D}^k (1 - p_{2D})^{N-1-k} \sum_{j=0}^{N-k} \lambda_j = \lambda_s \] (4)

The degree of dissimilarity between the centers of chaotic data clusters in distributed networks is calculated as follows:

\[
\text{DisSim}(A, B) = 1 - \left| \frac{\text{SameDis}(A) - \text{SameDis}(B)}{\text{Dis}(A) + \text{Dis}(B)} \right| \] (5)

In the above formula, \(\text{Dis}(A)\) denotes extended loss in clustering process, and \(\text{Dis}(B)\) represents attribute data set.

2.2. Attribute Feature Extraction of Chaotic Data in Distributed Networks

On the basis of principal component analysis of chaotic data in distributed network, semantic attributes and fuzzy clustering features of chaotic data in distributed network are extracted, and \(X\) is assumed to be a distributed network with \(m\) attributes. Chaotic data set in the network is:

\[
y(k) = s_1(k) + n_1(k), \quad \varphi(k) = s_2(k) + n_2(k) \] (6)

\[
s_1(k) = AA_1 e^{i(\Omega t + \phi_1)}, \quad s_2(k) = AA_2 e^{i(\Omega t + \phi_2)} \] (7)

Combining the objective cost minimization method with adaptive optimization[11-12], the semantic attribute features and fuzzy clustering features are obtained as follows:

\[
R_{\beta} X = \{ E \in U / R \mid c(E, X) \leq \beta \} \] (8)
The data object \( m_{i,j} (1 \leq i \leq n, 1 \leq j \leq k) \) is decomposed with mixed features. The clustering feature coefficient of abnormal data is \( \{\lambda_i : 1 \leq i \leq S\} \), criterion \( \{R_j : 1 \leq j \leq L\} \). According to the difference between attribute values of chaotic data classification attribute in distributed network, the training function \( f \) and base \( d_{\gamma_0} \) are obtained. The set of fuzzy concepts is:

\[
\lambda^n(d_{\gamma_0}) = \int_{-\infty}^{\infty} f(t)d_{\gamma_0}'(t)dt
\]

The joint association rule analysis method is used to deal with the interference filtering of the chaotic data in the distributed network\(^{[13]}\), and the autocorrelation feature block function of the chaotic data in the distributed network under the common attribute value is obtained:

\[
S_b = \sum_{i=1}^{n} p(\omega_i)(u_i - u)(u_j - u)^T
\]

\[
S_w = \sum_{i=1}^{n} p(\omega_i) E \left[ \frac{(u_i - u)(u_j - u)^T}{\omega_i} \right]
\]

\[
S = S_b + S_w
\]

The joint association rule model \( X(t) \) of chaotic data in distributed network is normalized and a new clustering modal function is obtained:

\[
X'(t) = X(t) / ||X(t)||
\]

Where, \( ||X(t)|| \) denotes the modulus of \( X(t) \), and the ability of classification and recognition of data feature detection is improved by processing the interference filtering of chaotic data in distributed network\(^{[14]}\).

3. Data Feature Detection Algorithm Optimization

3.1. Mixed Weighted Block Matching of Chaotic Data in Distributed Networks

On the basis of principal component analysis based on association rule attribute of chaotic data in distributed network, semantic attribute feature, fuzzy clustering feature and data fusion processing of chaotic data in distributed network are extracted, and distribution is carried out. An improved feature detection algorithm for chaotic data in distributed networks based on fuzzy C-means clustering is proposed in this paper. Under the condition of mixed attributes, a new algorithm for feature detection of chaotic data in distributed networks is proposed. The weighted filter function for feature detection of chaotic data is:
By using the method of correlation variable analysis, we obtain the heterogeneity metric 
\((u,u_j) \in C_{i}(K,H^{i}_{+} \times H^{i}_{-1})\), and fuzzy weighting method of the chaotic data in the distributed network, and obtain the numerical attribute weighting variables of the center of the chaotic data cluster in the distributed network:

\[
X(u) = \sqrt{\frac{1-j \cot \alpha}{2\pi}} \int_{-\infty}^{\infty} x(t) \exp \left\{ \frac{j^2 + \alpha^2}{2} \cot \alpha - j \cot \alpha \right\} dt
\]

(15)

The joint association rule analysis method is used to deal with the interference filtering of the chaotic data in the distributed network\(^{[15]}\), and the classified fuzzy set of the chaotic data in the distributed network is constructed based on the sensor fusion tracking method:

\[
\alpha_{\text{sensor}} = \alpha_i \sum \text{Density}_i + \alpha_z \frac{\text{AP}_{\text{init}}}{\text{AP}_{\text{init}}}
\]

(17)

The fuzzy fusion weighting coefficients of chaotic data in distributed networks satisfy:

\[
\begin{cases}
\alpha_i + \alpha_z = 1, \alpha_i, \alpha_z \in [0,1] \\
\alpha_z = \frac{\max(\text{AP}) - \min(\text{AP})}{\text{AP}_{\text{init}}}
\end{cases}
\]

(18)

Under the finite data set, the mixed weighted block matching set of chaotic data in distributed network is obtained as follows:

\[
x_{id}(t+1) = w_{x_{id}}(t) + c_{T1} \left[ x_{id}(t) \right] + c_{T2} \left[ x_{id}(t) \right] + c_{T3} \left[ x_{id}(t) \right]
\]

(19)

Thus, this method is used to match the chaotic data in distributed network.

3.2. Fuzzy C-means Clustering Classification for Data Feature Detection

On the basis of mixed weighting and adaptive block matching of chaotic data in distributed network in fuzzy data set, the association rule feature quantity of chaotic data in distributed network is extracted, and the extracted feature quantity is input to fuzzy data. In the mean value clustering classifier, the data classification recognition is carried out, and the one-dimensional data vector \(X_n\) is obtained. The extracted feature quantity is expressed as follows:

\[
X_n = \sum_{n=0}^{N-1} C_n \cdot e^{j2\pi k/n} \quad k = 0,1,\cdots,N-1
\]

(20)
The fuzzy C-means clustering classification is used for adaptive training for the weak association rule feature quantity of group anomaly net data \((x_i, x_{i+1}, \ldots, x_{i+m-1}, y_i), i = 1, 2, \ldots, n\).

The training formula is expressed as follows:

\[
\begin{align*}
net_{i1}(k) &= r_i(k) \\
net_{i2}(k) &= y_i(k)
\end{align*}
\] (21)

In the input layer of fuzzy C-means clustering, according to the numerical properties, the neuron states are obtained as follows:

\[u_{sii}(k) = net_{si}(k)\] (22)

According to the adaptive correction of neuron output measurement error, the output error of chaotic data feature detection in distributed network is obtained as follows:

\[
\begin{align*}
e_1 &= \varphi_a - \varphi_{ad} \\
e_2 &= \dot{\varphi}_a - \dot{\varphi}_{ad}
\end{align*}
\] (23)

Finally, continuous integral method is used to classify the chaotic data in the distributed network, and the classification model is presented as follows:

\[
\begin{align*}
\sigma_1(\varphi_a, \dot{\varphi}_a) &= \frac{1}{1 + e^{-(\alpha_1\varphi_a + \alpha_2\dot{\varphi}_a)}} \\
\sigma_2(\varphi_a, \dot{\varphi}_a) &= \frac{1}{1 + e^{-(\alpha_1\varphi_a + \alpha_2\dot{\varphi}_a)dt}} \\
\sigma_3(\varphi_a, \dot{\varphi}_a) &= \frac{1}{1 + e^{-d(\alpha_1\varphi_a + \alpha_2\dot{\varphi}_a)}}
\end{align*}
\] (24)

Finally, the error of the feature detection output of distributed network chaotic data converges to:

\[\lambda_i = \lim_{N \to \infty} \frac{1}{N} \sum_{j=0}^{N-1} \log(R_{ij}) = 0\] (25)

It can be seen that the method in this paper is used to detect the chaotic data in the distributed network. The feature detection process is stable convergence and the error tends to zero.

4. Simulation Experiment and Performance Analysis

In order to test the application performance of this method in realizing the optimized feature detection of chaotic data in distributed network, the simulation experiment is carried out. The experiment adopts Matlab design, the test data set is selected as Zoo dataset, and the initial scale of data set is obtained. The data packet size is 2000, the ratio of the main sidelobe height of the associated feature is 30 dB, the interference intensity of the chaotic data feature detection in the distributed network is -10 dB, and the initial sampling interval of the chaotic
The data in the distributed network is 0.01. The real parameter is set and the simulation experiment of feature detection of chaotic data in distributed network is carried out. Firstly, the original sampling of network transmission data is carried out, and the sampled time domain waveform of network transmission data is obtained as shown in figure 2.

Figure 2. Original sampled time domain waveform of chaotic data in distributed network

Taking the data collected in figure 2 as the research object, the chaotic data features in the distributed network are detected, the semantic attributes and fuzzy clustering features of the chaotic data in the distributed network are extracted, and the points are extracted according to the results of information fusion. The association rule feature quantity of chaotic data in distributed network is obtained, and the result of feature extraction is shown in figure 3. The output results of feature extraction of chaotic data in distributed networks using traditional spectral analysis method are also given in figure 3.

Figure 3. Feature extraction of chaotic data in distributed networks

Figure 3 shows that the proposed method for feature extraction of chaotic data in distributed network has a good capability of resisting sidelobe interference and the weak
association rule feature intensity of outputting chaotic data in distributed network is strong. The feature detection of chaotic data packets in distributed network with 10 classification attributes is carried out, and the accuracy of feature detection is tested. The result is shown in Table 1. The analysis of Table 1 shows that the method of this paper is used to carry out the chaotic data in distributed network. The accuracy of feature detection is 13.6% and 24.3% higher than that of the traditional spectral analysis method and K-means method and Wavelet analysis method.

Table 1 Accuracy test for feature detection

| SNR/d | Proposed method | K-means | Wavelet analysis |
|-------|-----------------|---------|-----------------|
|       |                 |         |                 |
| -12   | 95.7            | 65.8    | 72.5            |
| -8    | 96.9            | 75.3    | 84.7            |
| -4    | 99.8            | 84.8    | 89.7            |
| 0     | 100             | 89.9    | 96.4            |
| 4     | 100             | 97.9    | 98.7            |

5. Conclusions

In this paper, a new algorithm for the chaotic data feature detection in the distributed network is proposed based on fuzzy C-means clustering. According to the association rule attributes of the chaotic data in the distributed network, the principal component analysis is carried out, and the semantic attributes and fuzzy clustering features of the chaotic data in the distributed network are extracted. The joint association rule analysis method is used to deal with the interference filtering of the chaotic data in the distributed network, and the classified fuzzy set of the chaotic data in the distributed network is constructed based on the sensor fusion tracking measure method. In the fuzzy data set, the mixed weighting and adaptive block matching of chaotic data in distributed network are carried out, and the association rule feature quantity of chaotic data in distributed network is extracted. The extracted features are input into the fuzzy C-means clustering classifier for data classification and recognition, and the optimized feature detection of the chaotic data in the distributed network is completed. The simulation results show that the proposed method has good data convergence, strong convergence and anti-interference in the process of feature detection in distributed networks. This method has a good application value in the distributed network chaotic data feature detection.

References

[1] FLETCHER T D, ANDRIEU H, HAMEL P. Understanding, management and modelling of urban hydrology and its consequences for receiving waters:a state of the art[J]. Advances in Water Resources, 2013, 51(1):261-279.

[2] XIONG Xinglong, YANG Lixiang, MA Yuzhao, ZHUANG Zibo. Alerting algorithm of low-level wind shear based on fuzzy C-means[J]. Journal of Computer Applications, 2018, 38(3): 655-660.

[3] ZHENG J F, ZHANG J, ZHU K Y, et al. Gust front statistical characteristics and automatic identification algorithm for CINRAD[J]. Acta Meteorologica Sinica, 2014, 28(4):607-623.
[4] HWANG Y, YU T Y, LAKSHMANAN V, et al. Neuro-fuzzy gust front detection algorithm with S-band polarimetric radar[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(3):1618-1628.

[5] LIU D, WU Q Y, HAN W H, et al. User identification across multiple websites based on username features[J]. Chinese Journal of Computers, 2015, 38(10):2028-2040.

[6] ROEDLER R, KERGL D, RODOSEK G D. Profile matching across online social networks based on Geo-tags[C]//NaBIC 2015:Proceedings of the 7th World Congress on Nature and Biologically Inspired Computing, AISC 419. Berlin:Springer-Verlag, 2016:417-428.

[7] ZHANG F Z, LIU S, LI Z H, et al. Collaborative filtering recommendation algorithm incorporating user's reviews and contextual information[J]. Journal of Chinese Computer Systems, 2014, 35(2):228-232.

[8] ZHOU Wei, LUO Jianjun, JIN Kai, WANG Kai. Particle swarm and differential evolution fusion algorithm based on fuzzy Gauss learning strategy[J]. Journal of Computer Applications, 2017, 37(9): 2536-2540.

[9] PATEL H. Accelerated PSO swarm search feature selection with SVM for data stream mining big data[J]. International Journal of Research and Engineering, 2016, 3(9):15761-15765.

[10] KHABBAZIBASMENJ A, HASSANIEN A, VOROBYOV S, et al. Efficient transmit beamspace design for search-free based DOA estimation in MIMO radar[J]. IEEE Transactions on Signal Processing, 2014, 62(6): 1490-1500.

[11] SAMMARTINO P F, BAKER C J, and GRIFFITHS H D. Frequency diverse MIMO techniques for radar[J]. IEEE Transactions on Aerospace and Electronic Systems, 2013, 49(1): 201-222.

[12] WANG W Q and SHAO H. Range-angle localization of targets by a double-pulse frequency diverse array radar[J]. IEEE Journal of Selected Topics in Signal Processing, 2014, 8(1): 106-114.

[13] WANG W Q and SO H C. Transmit subaperturing for range and angle estimation in frequency diverse array radar[J]. IEEE Transactions on Signal Processing, 2014, 62(8): 2000-2011.

[14] WANG Youcheng, ZHANG Feng, FANG Guangyou, et al. A novel ultrawideband exponentially tapered slot antenna of combined electric-magnetic type[J]. IEEE Antennas and Wireless Propagation Letters, 2016, 15: 1226-1229.

[15] SHI En, LI Qian, GU Daquan, ZHAO Zhangming. Weather radar echo extrapolation method based on convolutional neural networks. Journal of Computer Applications, 2018, 38(3): 661-665.