Research on Network Aware Data Fusion Algorithm Based on Fuzzy Time Series

Jia Chen, Yingxue Cai, Song Hu, Sibo Huang, Hui Hu, Zhaoquan Cai*
Huizhou University, Huizhou, 516000, China

Abstract: In the environment of big data, the data mining and fusion of network awareness is affected by the interference between the small disturbances of the data, which leads to the poor clustering characteristics of the information fusion. A network aware data fusion method based on improved fuzzy time series analysis algorithm is proposed. The data stream of web address is reorganized in the distributed structure, and the phase space reconstruction technique is used to deal with the autocorrelation feature matching of the fuzzy time series attractor load. Combined with association rule fuzzy pairing method, fuzzy time series analysis is used to realize the optimal clustering of network perceptual data. According to the result of data clustering, the accurate fusion of network perceptual data is realized, and the feature compression of fused high-dimensional network perceptual data is carried out, so as to reduce computational overhead. The simulation results show that the proposed method has good clustering ability, it has strong ability to resist inter-class disturbance and strong ability of network perceptual data fusion.

Keywords: Big data; Fuzzy time series; Fusion; Feature clustering

1. Introduction

In large-scale data, effective data differentiation and mining are carried out, and how to use and manage large data is a key problem that people need to solve. In large data environment, the accurate analysis and fusion of network perception data are carried out, data information is quickly excavated, data is integrated, and information sharing is realized[1]. Accurate link transmission has become a key research topic in future information processing and large data analysis. It has great significance to improve the classification and recognition ability of network perceived data through the research of network perception data association fusion, and it is of great significance to the computer aided classification ability of network perceived data.

At present, the algorithm of feature fusion for network aware data information data is being actively carried out, and some research results have been obtained. In the traditional method, the fusion algorithm for network perception data is mainly taken based on the genetic algorithm for network aware data information data fusion, and the network sensing data based on the simulated quenching algorithm[2]. Fusion algorithm, network aware data information data fusion algorithm based on node supervision and homomorphism estimation, the most common method is particle swarm optimization (PSO) method, it is used for network sensing data information data fusion algorithm. In reference [3], it first proposed a SVM logic regulation network model, it used energy management and tracking differential. The time delay error compensates for the fusion processing of the network perception data, but the algorithm is affected by the noise and the precision is not high. In reference [4], it proposed a fusion algorithm based on the non-stationary time-varying signal analysis. The algorithm has a better accuracy in the gene expression under the better condition of the prior information, but the algorithm is more accurate. It cannot effectively solve the problem of feature partition and fusion of genetic data of network perception data, and it is not good to classify and diagnose network perception data.

In order to solve the above problems, this paper proposes a network perceptual data fusion method based on improved fuzzy time series analysis algorithm. The data stream of web address is reorganized in the distributed structure, and the phase space reconstruction technique is used to deal with the autocorrelation feature matching of the fuzzy time series attractor load. Combined with association rule fuzzy pairing method, fuzzy time series analysis is carried out to realize the optimal clustering of network perceptual data. According to the result of data clustering, the accurate fusion of network perceptual data is realized, and the fused high-dimensional network perceptual data is obtained. Feature compression is carried out to reduce computational overhead. Finally, the simulation results show the superior performance of this method.

2. Reconfiguration of Distributed Structure and Phase Space Reconstruction of Network-aware Data

2.1. Network aware data distributed structure reorganization

In order to realize the fusion of network perceptual data in big data, the association rules are mined by fuzzy time series analysis method, and the characteristic quantity of network perceptual data is fused, then the temporal order of big data network perceptual data is put forward[5]. The first step of sequence fuzzy time series analysis is to reconstruct the phase space and observe the time series of big data's information flow, which is to be mined, so that $\{x_1, x_2, \cdots x_N\}$ is a group of nonstationary wideband time series, so that $x(n)$ can move forward in the dimensional distributed feature space. Based on the structural mapping of line network aware data, big data's distributed recombination structure is obtained as follows:

$$\boldsymbol{X}(n) = \{x(n), x(n+t), \cdots, x(n+(m-1)t)\} \quad n = 1, 2, \cdots, N \tag{1}$$

Where, t denotes big data's sampling time delay in high-dimensional space. Adaptive fuzzy time series training method is used for feature fusion, and phase trajectory evolution analysis in high-dimensional feature space is carried out to obtain a large number of reconstructed features[6]. According to the path of network perceived data distribution

$$X = [s_1, s_2, \cdots s_K] = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_{1+t} & x_{2+t} & \cdots & x_{K+t} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1+(m-1)t} & x_{2+(m-1)t} & \cdots & x_{M+(m-1)t} \end{bmatrix} \tag{2}$$

Where, $K = N-(m-1)t$, big data searches for embedding dimension of feature space, t is delay, m is layer number of semantic ontology attribute of information, $s_i = (x_i, x_{i+t}, \cdots, x_{i+(m-1)t})^T$ is called feature vector set of phase space. In the network aware data distributed structure reorganization model[]7, the fuzzy set search method is used to determine and delay the time. The high order statistic analysis and feature fusion of network perceptual data are carried out in high dimensional phase space.

2.2. Network aware data phase space reconstruction

On the basis of reorganizing the distributed structure of the high-dimensional data flow, the phase space of the network perceptual data is reconstructed. The Lorenz fuzzy time series attractor is used as the training test set to fuse the network perceptual data of the network. Combined adaptive learning training, the detection probability of network perceptual data fusion is $P_q(q_j)$, $j = 1, 2, \cdots, n$. Under the training of fuzzy time series analysis, the phase space reconstruction model of network aware data is constructed as follows:

$$H(S) = -\sum_{i=1}^{n} P_s(s_i) \log_2 P_s(s_i) \tag{3}$$

$$H(Q) = -\sum_{i=1}^{n} P_q(q_j) \log_2 P_q(q_j) \tag{4}$$

Where, $P_s(s_i)$ represents the semantic concept set of network-aware data. The probability of appearing in the region s_i of fuzzy time series analysis, the average mutual information satisfying the clustering condition of network perceptual data in phase space S is obtained as follows:

$$I(Q,S) = H(Q) - H(Q|S) \tag{5}$$

Where

$$H(Q|s_i) = -\sum_j \left[{p_{sq}(s_i, q_j)} \Big/ {p_s(s_i)} \right] \log_2 \left[{p_{sq}(s_i, q_j)} \Big/ {p_s(s_i)} \right] \tag{6}$$

It represents the classification attribute set which satisfies the condition P , and adopts the method of average mutual information fusion

3. Optimization of Network Aware Data Fusion Algorithm

3.1. Improved fuzzy time series analysis algorithm

On the basis of the preprocessing of distributed structure reorganization and phase space reconstruction for high-dimensional data information flow, the optimal design of network aware data fusion algorithm for network aware data is carried out. In this paper, an improved fuzzy-time algorithm based on improved fuzzy time is proposed[8]. Network aware data fusion method based on sequence analysis algorithm. Using Lorenz fuzzy time series attractor as the training test set, the fuzzy attribute partition of network aware data is carried out, and the expression of Lorenz fuzzy time series attractor is given as:

$$\begin{cases} dx/dt = -s\,x + s\,y \\ dy/dt = -xz + rx - y \\ dz/dt = xy - bz \end{cases} \tag{7}$$

All the variables are dimensionless. and x , y , z denote the sampling time of fuzzy data. The clustering center Euclidean distance of fuzzy time series analysis of network perceptual data is obtained as:

$$\begin{aligned} R_{mn} &= \left\| \boldsymbol{X}_{h(n)} - \boldsymbol{X}_n \right\|_2^{(m)} = \min_{j=N_0,\cdots,N,\, j\neq n} \left\| \boldsymbol{X}_n - \boldsymbol{X}_j \right\|_2 \\ &= \sqrt{\sum_{l=0}^{m-1} (x_{h(n)+lt} - x_{n+lt})^2} \end{aligned} \tag{8}$$

With adaptive learning in the reconstructed phase space, when the embedding dimension m of phase space is increased to $m+1$, the clustering center optimization value of fuzzy time series analysis of network perceptual data fusion is obtained as follows:

$$R_{(m+1)n} = \left\| \boldsymbol{X}_{h(n)} - \boldsymbol{X}_n \right\|_2^{(m+1)} = \sqrt{\sum_{l=0}^{m} (x_{h(n)+lt} - x_{n+lt})^2} = \sqrt{\left[\left\| \boldsymbol{X}_{h(n)} - \boldsymbol{X}_n \right\|_2^{(m)} \right]^2 + \left[x_{h(n)+mt} - x_{n+mt} \right]^2} \tag{9}$$

If $R_{(m+1)n}$ is much larger than R_{mn}, it can be used as the center of fuzzy clustering.

3.2. Information fusion and feature compression

Assuming that the distributed time series $\{X_n\}, n = 1, 2, \cdots, N$ of network perceptual data represent the feature distribution set of the network perceptual data to be partitioned, the feature distribution $X_N = X_n + h$ of fuzzy time series analysis is obtained under the fuzzy time series analysis, where h is the observation noise. The phase space reconstruction technique is used to deal with the load feature quantity of fuzzy time series attractor of network perceptual data, and the feature matching output is obtained as:

In the closed frequent item region of fuzzy time series analysis, the average mutual information characteristic quantity of network perceptual data is fused as follows:

$$R_1 = \{X_1, X_2, X_3, \cdots, X_d\}^T \tag{10}$$

The set of association rule vectors for network-aware data is:

$$R_1^T R_1 = \{X_1, X_2, \cdots, X_m\}\{X_1, X_2, \cdots, X_m\}^T \tag{11}$$

The network sensing data is decomposed into single frequency signal by SVD decomposition, as:

$$X = USV^T = U \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} V^T \tag{12}$$

Where, $X \in R^{m\times n}$, $U \in R^{m\times m}$, $V \in R^{n\times n}$, In the feature mining search of the whole correlated network perceptual data, the optimal solution of the associated feature vector of the network perception data is described as the diagonal vector matrix:

$$\Sigma_r = diag(s_1, s_2, ..., s_r) \in R^{r\times r} \tag{13}$$

In order to realize the partition feature fusion of network perceptual data information data, the information entropy feature is scheduled in high dimensional space. Phase space reconstruction technique is used to deal with the feature quantity of fuzzy time series attractor load by autocorrelation feature matching. Fuzzy time series analysis based on association rule fuzzy pairing method is used to realize the optimal clustering of network perceptual data and the accurate fusion of network perceptual data is realized according to the result of data clustering[9]

4. Simulation Experiment Analysis

In order to test the performance of this algorithm in implementing network aware data association fusion and computer-aided diagnosis, simulation experiments are carried out. In the simulation experiment, the algorithm is programmed by Matlab. The hardware environment of the simulation experiment is Intel Core3-530 1G memory, the operating system is Windows 10, Kinect driver, the PrimeSense official version OpenNI framework, the development tool of data fusion algorithm is Microsoft Visual Studio 2010, the delay value of network perceptual data collection is 20ms. the fusion parameters f and j are used to represent the weight distribution of these two prediction results. The parameters are chosen as: 1/ 3 and 2/ of data symbol, respectively. The network perceiving data sample data has the fixed global iteration number of 500 times, the variable dimension is 30 times, and the algorithm runs 30 times independently. According to the above simulation environment and parameter setting, the experiment is carried out, and the network aware data test set and fusion result are obtained as shown in figure 1.

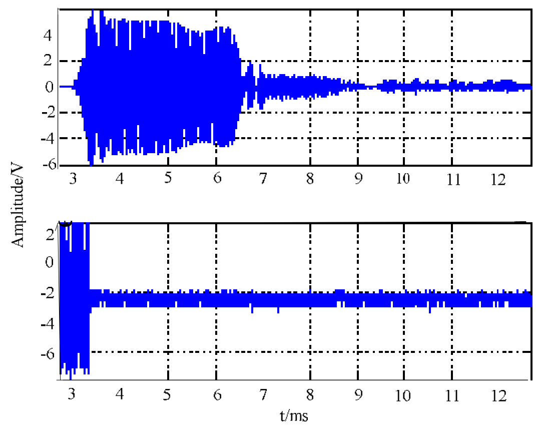

Figure 1. Network aware data set and fusion results

It can be seen from figure 1 that the accuracy of network perceptual data fusion based on this method is high, and the accuracy of different methods is tested, and the comparison results are shown in figure 2.

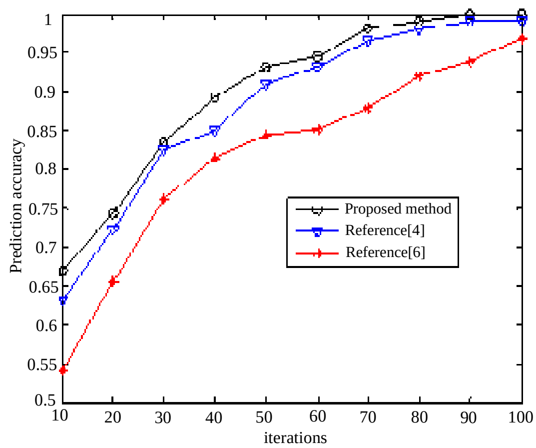

Figure 2. Accuracy comparison

Figure 2 shows that the proposed method is accurate and real-time. On the basis of this proposed method, the fuzzy time series analysis of network perceptual data information has better attribute clustering, higher convergence among classes, stronger ability to resist inter-class disturbance, and better ability of network perceptual data fusion.

5. Conclusions

In this paper, a network aware data fusion method based on improved fuzzy time series analysis algorithm is proposed. The data stream of web address is reorganized in the distributed structure, and the phase space reconstruction technique is used to deal with the autocorrelation feature matching of the fuzzy time series attractor load. Combined with association rule fuzzy pairing method, fuzzy time series analysis is used to realize the optimal clustering of network perceptual data. According to the result of data clustering, the accurate fusion of network perceptual data is realized, and the feature compression of fused high-dimensional network perceptual data is carried out, so as to reduce computational overhead. The simulation results show that the proposed method has good clustering ability, it has strong ability to resist inter-class disturbance and strong ability of network perceptual data fusion, it has good application value in practice.

4. Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61772225);
the Foundation for Distinguished Young Talents in Higher Education of Guangdong (No. 2015KQNCX153); Science and Technology Program of Huizhou (No. 2015B010002002, No. 2016X0431046, No. 2016X0434049, No. 2016X0432047, No. 2017c0406022, No. 2017c0407023, No. 2017c0414030).

References

[1] XUE You-min,YAN Yu-ping,GU Jia-ling,et al.A comparison between two KNN based feature selection algorithms[J].SAMSON,2016,(01):19-22.

[2] YANG Z M, QIAO L Y, PENG X Y. Research on datamining method for imbalanced dataset based on i mproved SMOTE[J]. Acta Electronica Sinica, 2007, 35(S2):22-26.

[3] GU Q, YUAN L, XIONG Q J, et al. A comparative study of cost-sensitive learning algorithm based on imbalanced data sets[J]. Microelectronics and Computer, 2011, 28(8):146-149.

[4] MAO W T, TIAN Y Y, WANG J W, et al. Granular extreme learning machine for sequential imbalanced data[J]. Control and Decision, 2016, 31(12):2147-2154.

[5] GU Q, YUAN L, NING B, et al. A noval classification algorithm for imbalanced datasets based on hybrid resampling strategy[J]. Computer Engineering and Science, 2012, 34(10):128-134.

[6] GUO Huaping, ZHOU Jun, WU Chang'an, FAN Ming. k-nearest neighbor classification method for class-imbalanced problem. Journal of Computer Applications, 2018, 38(4): 955-959.

[7] DURAO F, DOLOG P. Improving tag-based recommendation with the collaborative value of wiki pages for knowledge sharing[J]. Journal of Ambient Intelligence Humanized Computing, 2014, 5(1):21-38.

[8] Lu Xing hua, Fan Tai-lin, Xie Zhen-han .Embedded System Design of Multi Mode Intelligent Control Based on ARM [J]. Computer & Digital Engineering, 2016,44(4): 667-670,674.

[9] CUN Yong-jun, ZHNAG Yong-hua. Linux System Dual Threshold Scheduling Algorithm Based on Characteristic Scale Equilibrium[J]. Computer Science, 2015,42(6):181-184.