The Method of Time Granularity Determination on Time Series Based on Structural Similarity Measure Algorithm

Gao Xuedong, Chen Hailan
Donlinks School of Economics and Management
University of Science and Technology Beijing
Beijing, China
gaoxuedong@manage.ustb.deu.cn, s20141163@xs.ustb.edu.cn

ABSTRACT

With the constant progress of science and technology, data size is increasing in the areas of social life and industrial production. And people are gradually aware of the potential value of data, causing a flood of big data and data mining. Data in real life is mostly related to time, called time series. At present, data analysis and data mining for time series has become the research focus in the field of data mining. Dimension reduction, similarity matching, cluster analysis and etc. for time series of large scales can dig out the useful information so that effectively guide the business decision and production application.

According to the method of granular computing [1], information granularity affects the complexity and the validity of problems. In general, as the granularity decreases, the result becomes more accurate, while the computational complexity rises sharply. However it is likely to cover the nature of problems, so the size of granularity produces a great impact on the results of problems. Time granularity is the information granularity of time series, so time granularity is important for time series researches. In year 1996 to 1998, Claudio Bettini et al [2, 3] firstly proposed the concept that time interval can be regarded as time granularity. In year 2014, Xu Jianfeng et al [4] proposed the basic model on time granularity determination of multi-granulation time series, and used it to cluster analysis. But the model stays the level that time granularity is made by men, and there is no scientific calculation method is put forward. Granular computing theory [5] mainly transform time granularity according to the different problems, so as to simplify the problem solving. In this paper, the method for determining the optimal time granularity is different from traditional granular computing theory, it is based on basic granularity to determine the optimal time granularity. The wavelet analysis method is widely applied in the field of data mining, such as time series trend information extraction, time series similarity matching, time series data dimension reduction, etc. [6]. The wavelet analysis method is firstly applied in time series pattern recognition in paper [7]. Based on it, Zhang Haiqin et al [8] put forward the method of time series similarity pattern matching based on the wavelet transform. In year 1999, Zbigniew R et al [9] used Haar wavelet transform in time series similarity expression.

The research of time granularity determination on time series is an important problem in data mining research. With the rapid growth of data quantity, it is necessary to obtain valuable information quickly and accurately by time series data mining. In order to solve the problem, in this paper, a method of time granularity on time series is proposed based on structural similarity measure algorithm.

In this paper, the preliminary researches are carried on in the following three aspects.

(1) Fluctuation point recognition

Obviously there is an essential relationship between the size of time granularity and the fluctuation of time series, so the fluctuation point recognition of time series is particularly important. In this paper, it is proposed the definition of fluctuation point and its recognition method, we can extract the key information of a curve, and for further analysis.

(2) Haar wavelet transform

Conducted multi-scale decomposition of time series data by Haar wavelet transform, each decomposition time series data is decomposed into scale signal contains low frequency component and noise signal contains high frequency component and noise. Each Haar wavelet decomposition, the time granularity of Haar wavelet decomposition time series is 2^i ,

and i is the number of Haar wavelet decomposition. As time granularity of Haar wavelet is fixed, therefore, in this paper, Haar wavelet is selected as the data dimension reduction method, and ready to measure the structural similarity.

(3) Structural similarity measure method

Structural similarity measure method is proposed in this paper, and it is mainly considered that the position similarity of corresponding elements in two time series, rather than their distance. This similarity measure method plays an important role for time granularity calculation.

Finally, the process of the optimal time granularity determination is as follows:

Step 1: Identify the fluctuation points of original time series to get fluctuation point sequence;

Step 2: For the original time series to conduct multilayer Haar wavelet decomposition, the decomposed layer is determined by the length of original time series, and identifies its turning points to get Haar wavelet decomposition sequence;

Step 3: Calculate the similarity of fluctuation point sequence and Haar wavelet decomposition sequence based on structural similarity measure method;

Step 4: The similarity of fluctuation point sequence and wavelet decomposition sequence constitutes a similarity sequence. Then obtain the turning point of similarity sequence, it is the optimal time granularity that the turning point correspond to the time granularity.

In this paper, the method is proposed to determine the time granularity of time series. It solves the problem that data mining relevant work is difficult to put into effect because of big data, and its computational complexity is $O(n^2)$. Before data mining, we can select a subsequence varies stable to analyze by applying the proposed algorithm to determine the optimal time granularity, then use the time granularity to do data mining research for the original time series.

Keywords: time granularity, fluctuation point recognition, Haar wavelet transform, structural similarity measure

References

[1]. Li Daoguo, Miao Duoqian, Zhang Dongxing, and Zhang Hongyun, "Granularity computing research review," Computer Science, Vol. 9, pp. 1-12, 2005.

[2]. Claudio Bettini, X. Sean Wang, and Sushil Jajodia, "Testing complex temporal relationships involving multiple granularities and its application to data mining," in Fifteenth ACM Sigact-Sigart Symposium on Principles of Database Systems, Montreal, Canada, June 3-5, 1996.

[3]. Claudio Bettini, X. Sean Wang, Sushil Jajodia, and Jia-Ling Lin, "Discovering frequent event patterns with multiple granularities in time sequences," IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No. 6, pp. 222-237, 1998.

[4]. Xu Jianfeng, Zhang Yuanjian, Zhou Duanning, Li Dan, and Li Yu, "Uncertain multi-granulation time series modeling based on granular computing and the clustering practice," Journal of Nanjing University(Natural Sciences), Vol. 1, pp. 86-94, 2014.

[5]. Fang Gang, The research of dynamic mining technology based on granular computing, CN: University of Electronic Science and Technology of China, 2015.

[6]. Tian Zhengxiong, Study on Time Series Data Mining by Wavelet Transform, CN: Tianjin University, 2008.'

[7]. Kin-pong Chan, Ada Wai-chee Fu, "Efficient time series matching by wavelets," in IEEE International Conference on Data Engineering, 1999.

[8]. Zhang Haiqin and Cai Qingsheng, "Time series similar pattern matching based on wavelet transform," Chinese Journal of Computers, Vol. 3, pp. 373-377, 2003.

[9]. Zbigniew R, Struzik, Arno Slebes, "The Haar wavelet transform in the time series similarity paradigm," in Principles of Data Mining & Knowledge Discovery, 3 Rd European Conference Prague, Czech Republic. 1999.