test results, and can manage the data resources safely and effectively, so as to inspect, modify and improve the system itself in time.

3). Introduce data mining techniques to build the athlete nerve type group traits evaluation model and test system, involved in sports training, competition command and athlete selection, broaden the field of data mining technology and services, and facilitate the pace of China's sports into the information age.

References

1. U.M.Fayyad, Editors.(1996) Advances in GPiatetsky-Shaperio, P.Smyth and R.Uthurusamy. *Knowledge Discovery and Data Mining*. AAAI/MIT.
2. Rajendra K Srivastava, Tasadduq A Shervani, Lianrey Fahey (1998) *Market-based assets and shareholder value: A framework for analysis. Journal of Marketing*, 62(1), p.p.2-18.
3. James T Rothe, Michael G Harvey, Candice E Jtickson (1997) *The marketing audit: Five decades later. Journal of Marketing Theory and Practice*, 5(3), p.p.1-16.
4. Bruce W.(1997) The market orientation construct: Measurement and scaling issues. *Journal of Marketing Theory and Practice*, 5(3), p.p.31-54.
5. Bonoma Thomas V and Clark Brace H.(1988) *Marketing Performance Evaluation*. Harvard Business School Prcsa.
6. Kumar, Nirmalya, Stern, Louis W, Achrol, Ravi S. Assessing Reseller Performance from the Perspective of the Supplier. *Journal of Marketing Research*, 29(2), p.p. 238-253.
7. Chen Dong, Fengyuan Liu, Haiyan Wang et al.(2010) Application Research of Mobile GIS in orestry Informatization . *The 5th International Conference on Computer Science & Education*, p.p. 351-355.
8. Arindam Dasgupta. A Framework for Mobile Based Geospatial Public Health management. *Proceedings of the 2010 IEEE Students' Technology Symposium*, p.p. 82-87.

A Time Series is Approximated Parameter Control Algorithm

Rongli Zhang, Bo Han, Aijun Liu

School of Mathematics and Computer Applications, Shangluo University, Shangluo 726000, Shaanxi, China

Abstract

In order to improve the efficiency of time series data mining, to better serve the community activities; the use of effective methods of data simulation study proposed an approximate representation of the decomposition technique based on discrete wavelet transform key multi-scale; the results showed that: compared to the existing approximate representation, it retains the original scale representation of the approximate coefficients, get rid of the user precise control of the operating parameters to achieve the key point sequence is approximated approximate wavelet coefficients of each scale; this method both retain the main characteristics of time series data sequences, but also greatly to achieve an effective reduction of the dimension of time series data, experimental simulation study tested it with the traditional method on the approximation of quality and adaptability has obvious advantages and operability .

Keywords: TIME SERIES DATA, APPROXIMATE LATE, WAVELET TRANSFORM

1. Introduction

Information society is highly developed today, people through the Internet, mobile phone terminals, cloud cluster services, Big Data era of social finance, trade, transportation, enterprises generated a massive time-series data in social activities, such time real value chronological order of collection of computer data analysis, we called time series[1]. With the continuation of the business object and the transaction process subsequent shift time, record the relationship will become of these time series of increasingly large mass of data, the structure has become increasingly complex, data analysis directly with the original time sequence data mining will have a high price and poor performance[2,3].To solve the domestic and foreign researchers and improve the efficiency of time series data mining, a common practice is to replace currently used raw time data series methods based on approximated that the main form of retention time data sequence, ignoring the tiny details of the time series compression, dig out the useful time-series data.

However, they generally use research methods, how to efficiently represent the time series, time series data mining to improve efficiency and better serve the community, has its advantages and disadvantages, are summarized as follows.

The main advantages are: the length of this time series now widely used approximate representation of time series data represented far less than the length of the original sequence, thereby reducing the storage cost time series, time series mining to improve the efficiency and reduce the time details of the redundancy of the sequence, which greatly enhanced the performance data research and analysis [4].

The main disadvantage is that: These approximate methods are arguments, the basic input parameters need to be set correctly, if difficult to determine the parameters to be entered or inappropriate, at this time of the algorithm cannot achieve the best results, dimension reduction will be lost basic information data, coupled with the disturbing factors human action, techniques and instrumentation sensitivity, etc., are often difficult to obtain valid parameter control, the reduction cannot be maximized and close to the original data[5].

In summary, we propose a new time series approximation symbolic representation, based on the key points of the discrete wavelet transform technology, multi-scale decomposition approximate representation, compared to the existing approximate representation, it retains the original approximate representation of the scale factor, and does not require the user precise control of parameters, but also achieve a critical point sequence is approximated wavelet approximation coefficients for each scale, which not only retain the main characteristics of time series data sequences, but also greatly to achieve an effective reduction of time-series data dimension. Through research and experimental results show that this method is feasible and effective, it has some practical value.

2. Related research

Time series is set in real time order. X represents the time series with, $X = <x_1, x_2, ..., x_n>$, where x_i is a real number, the X value represents the time series at time i, n represents the length of the time series. Currently, the approximate representation of the time series of the main frequency domain representation, notation and piecewise linear representation. Frequencies domains representations of a time series will be seen as a signal of the time domain, using orthogonal transform, the signal is converted to the frequency domain, and then ignore those small amplitude time series form a small frequency, so to obtain a high level representation of the time series. Frequency-domain sequence can be transformed through the inverse transform to the time domain to obtain an approximate representation of the time series[6]. Notation is a discrete method is through discrete methods to map real-valued time series or time series waveform over time to a limited symbol table, the time sequence is represented as an ordered set of finite symbols, that string[7].

At present, more commonly used time series approximation method has discrete Fourier transform DFT, discrete wavelet transform, singular value decomposition, landmark model, piecewise linear representation PLR , slide polymerization approximate PAA, adaptive gathered nearly constant, and a symbol of polymerization approximately equal Wait. These approximate representation method has its own advantages and can achieve a certain degree of time-series dimension reduction[8,9]. However, most of these approximation methods are required for multiple input parameters are set correctly. As for the method parameters, the biggest threat is not appropriate parameter settings so that the algorithm cannot achieve optimal performance that cannot balance the degree of sequence compression and retention of key information so that key information is missing or sequence length has not been reduced. Especially when it is difficult to determine the appropriate value of the parameter for the user, the occurrence of such threats is often unavoidable [10]. Therefore, the ideal time sequence should approximate representation can be achieved not only time-series dimension reduction, retained the main features of the sequence, with input parameters must also be little need to pre-set, preferably no arguments.

3. Based on an approximate representation of the time series model building

3.1 Related concept definition

Before modeling, we first define the definition proposed approximate representation of the required basic concepts and terminology, and on the issues to be studied conducted a formal definition. In order to approximate the later chapters make clearer formulation proposed method, we advance to the symbol used in the method are given a definition.

Definition 1: Define the following parameters and symbols:

Y_t: time series to be processed;

n: Dimension of time series, which is the length of time series;

k: Decomposition of scale, time value is $Log_2 n$;

cA_i: Myopia wavelet coefficients on the first i-layer;

cD_i: The i-th layer wavelet detail coefficients;

MM_i: The key point of the i-layer sequence wavelet approximation coefficients;

S_i: Symbolic key sequence on the i-th layer;

W_i: Symbolic encoding of the i layer;

S: Source data.

Definition 2: basic concepts are defined as follows:

(1) high-dimensional time series: time series Y_t is set according to the real value of the chronological order of $\{v_1, v_2, \ldots, v_n\}$, where, v_n is the observation time point t_i on, n is the number of time series, also known as the dimension of time series. Due to the time sequence has massive resistance, so n is often very large, high-dimensional time sequence of the time series is studied in this chapter[11,12].

(2) wavelet coefficients: the discrete wavelet transform, Y_t original signal is decomposed into low frequency and high frequency information cD and cA two parts, namely decomposition has a relation $Y_t = cA + cD$. The main feature information cA contain low frequency information signal, can be seen as an approximation of the original signal, so called wavelet approximation coefficients[13]. cD capture high-frequency information is filtered original signal details, it can be regarded as noise. Such wavelet approximation coefficients cA contains not only the main information of the original signal, but also filters out noise in the original signal[14].

(3) Key points: For the time series Y_t, its key points, including the start point, end point, the local maximum points and local minimum points.

(4) Symbolic: for time series $Y_t=\{v_1, v_2, \ldots v_n\} \in R$, given a finite discrete symbols set $R=\{r_i|1<i<m\}$.

Definition 3: for the time series Y_t has a very high dimension n, in order to find its approximate with dimensions of m expressed Y_t, must meet the following three conditions:

(1) The m<<n efficient dimension reduction;

(2) Retains the essential features of Y_t;

(3) The method is no parameter.

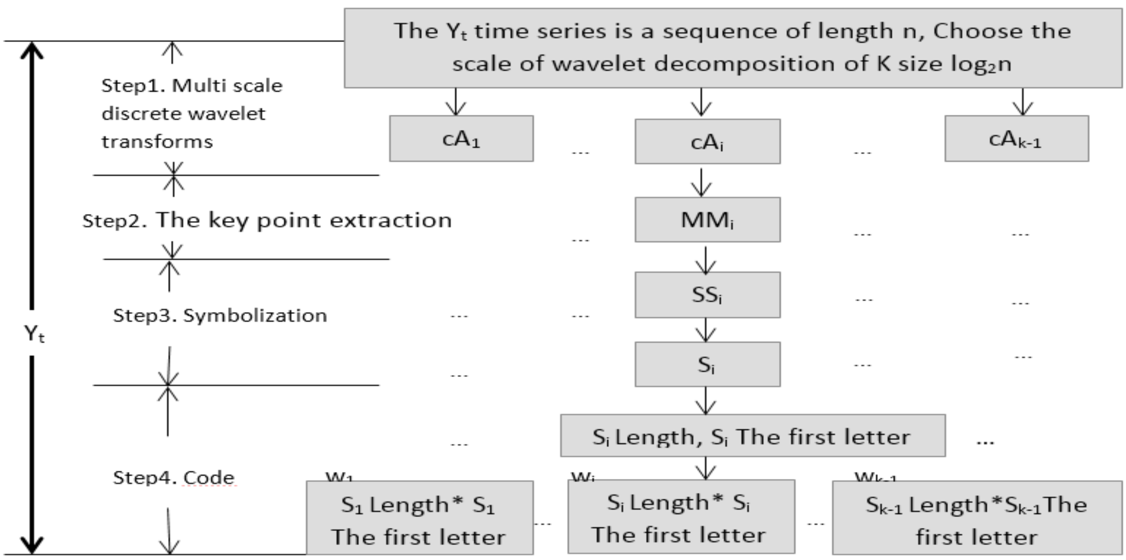

Figure1. NSAR said the overall framework of approximate algorithm

3.2 Model construction

This section will present a multiscale wavelet approximation coefficients based on time series symbolization of key points and symbolic coding sequence of approximate representation model of NSAR. The overall framework of the method is shown in Figure 1. From figure 1 can be found in the NSAR time series approximation representation is mainly divided into four parts: the first is the use of multiresolution discrete wavelet decomposition DWT to obtain the wavelet approximate coefficient; and then extract the key points of the wavelet approximation coefficients of each layer, the corresponding wavelet approximate coefficient of substitution with key point sequence; then we use the non-parametric approximation said method is symbolic of key points of each layer sequence, alternate key point sequence by symbolic sequence; finally, to encode the symbolic sequences of each layer, the approximate representation of each layer of code collected get original time series. The first two steps in the retention plays a very important role in the main information of time series; after the two steps is the key dimensionality reduction.

3.3 The algorithm description

First establish the following definition:The length of the signal required to be treated must be a multiple of 2, so for a time series of Y_t, before the wavelet approximation coefficients extraction, first of all need to detect n its length whether meet the requirements. If satisfied, the discrete wavelet transform can; otherwise, need to fill 0 until its length at the end of the Y_t to meet the requirements.

The algorithm description:

In order to make the algorithm is simple and fast, we choose the easy realization, good performance of Haar wavelet as the basic wavelet. Multi scale DWT specific decomposition steps are as follows: in the first layer decomposition, the original time series is decomposed into low frequency part of the Y_t, namely the wavelet approximate coefficient cA_1, and the high frequency part, i.e. the wavelet detail coefficients of cD_1; in the second layer decomposition, choosing the low-frequency part of the cA_1, to continue the decomposition, have a low frequency part and high frequency part cA_2. cD_2,, so iterative execution continues, followed on the low frequency part of every layer of the decomposition, $\log_n$ decomposition scale until the end of the completion, calculate the i layer wavelet approximation coefficients and detail coefficients cA_i cD_i respectively by using formula (1) and the formula (2):

$$cA_i = \frac{even\{cA_{i-1}\}+odd\{cA_{i-1}\}}{\sqrt{2}} \qquad (1)$$

$$cD_i = \frac{even\{cA_{i-1}\}-odd\{cA_{i-1}\}}{\sqrt{2}} \qquad (2)$$

The following work to be done is to extract the wavelet approximation coefficients, namely the extraction of first layer to the $\log_{2n-1}$ layer of the wavelet approximation coefficients$\{cA_{\log_2 n-1}, cA_{\log_2 n-2}, \cdots, cA_1\}$.

Extraction of wavelet approximation coefficients is mainly based on the following considerations: if the time sequence as the signal, then for most of the signal, the low frequency part is one of the most important parts of the signal, is an important part of the identification signal. And the high-frequency part represents only some subtle difference. With the voice of the people as an example, if you remove the high frequency part, although the voice sounds different, but still can know what to say; but, remove the low frequency part, only to hear a noisy. So the extraction of wavelet approximation coefficients can highlight the main feature, sequence, wavelet approximation coefficients contains the main characteristic information of the original series Y_t filter noise.

4. Experiments and result analysis

The experimental data of this section selects the simulation time series and real time series. As shown in Figure 2, the simulation time series include: sinusoidal periodic time series Y_1 and random white noise sequence Y_2. Figure 3 is the three real time series used: motor flux Y_3, sunspot number Y_4, and the volume of Y_5 shares of stock in a company, they are real time series.

First the basic characteristics of the five experimental series were qualitatively and quantitatively in the experiment described, i.e. four index calculation values: SE, PE, LZ and dish, and then according to the quantitative index values are given sequence specific description, as shown in table 1. Approximate representation and calculation of compression rate table 2 Results five experimental series.

The experimental results show that the dimension reduction of the same, compared with the existing symbolic approximate expression of SAX, NSAR can retain the original time series more basic information, performance and in time series similarity retrieval has better.

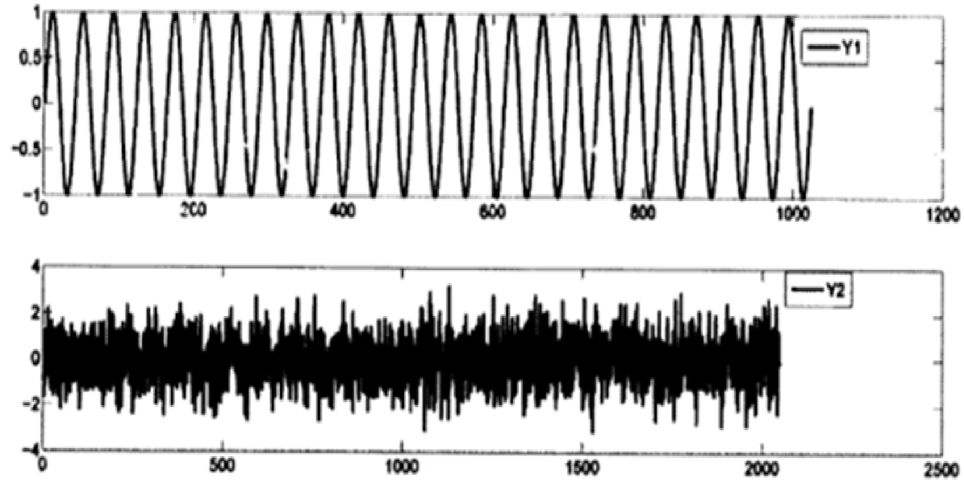

Figure2. Time series simulation

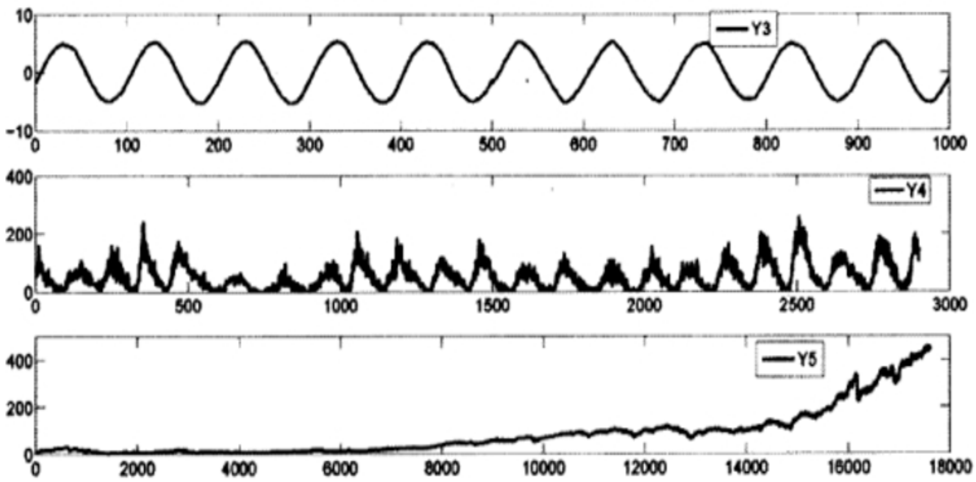

Figure3. Real time series

Table 1. Describe the basic characteristic of quantitative and qualitative experimental sequence

Sequence	Quantitative description				The qualitative description
	SE	PE	LZ	HE	
Y1	0.523	0.412	0.005	0.412	Periodic fluctuation, Comparison rules, The relatively stable
Y2	2.023	1.056	0.951	0.493	Close to the noise, More random component, Very unstable
Y3	0.463	0.351	0.010	0.742	Periodic fluctuation enhanced, The trend of the weak, A rule, Stable
Y4	0.747	1.047	0.105	0.827	Pseudo periodic fluctuation, The trend of enhancement, There are certain burr
Y5	0.243	0.964	0.021	0.965	The characteristics of obvious trend, Comparison rules, Stable

Table 2. At the same compression rate of two kinds of approximate representation method contrast retention characteristics

Sequence	Symbolic approximate representation		Compression ratio CR
	NSAR	SAX	
Y1	{-1,3,3,-13,15,-48,51,51,51}	{b,b,a,a,b,b,a,a,b}	0.801%
Y2	{1,3,-5,11,-24,-41,-88,175,333,678}	{a,a,b,b,b,a,b,b,a,b}	0.202%
Y3	{1,-3,5,-13,-20,21,21,21,25}	{b,b,b,b,a,a,a,a,a}	0.901%
Y4	{-1,2,-3,6,12,-43,-45,47,120,366,822}	{a,b,a,a,b,b,a,a,a,b,b}	0.381%
Y5	{1,1,1,-4,13,27,40,95,197,-413, -848,1645,-3417}	{a,a,a,a,a,a,a,a,b,b,b,b,b}	0.083%

1. Conclusions

In this paper, the performance of the existing approximate representation depends on the parameter itself brought some difficult to determine the optimum value of the problem, based on multi-scale DWT, key points, and symbolic coding technique nonparametric symbol approximation representation model -NSAR. The first model is the use of multi-scale wavelet decomposition DWT wavelet obtain approximate coefficients; then for each layer wavelet approximation coefficients extract key points, substitute the appropriate coefficients of wavelet approximation key sequence; Next key for each layer point sequence symbolized alternative key sequence symbol sequences; and finally, for each layer of symbolic coding sequence, each layer of code to collect up to get an approximate representation of the original time series. NSAR dimensions of the original sequence can achieve a great degree of reduction, if the original sequence of dimension n, can put a dimension reduction to $\log_{2n-1} \ll n$. Meanwhile, the model is no argument, and from the three aspects of the process with no arguments: first, in multiscale DWT, the decomposition scale choice is determined by the length of the sequence; the second, after the filter out noise Approximate on the wavelet coefficients extract key points, do not set a threshold for screening key points; and third, to symbolize the key point sequence, automatically determines the only two symbols to represent rising and falling trend. In addition, the experimental results show that under the same degree of reduction of dimension compared to conventional symbolic approximated SAX, NSAR can retain the original time series more basic information and a more in time sequence similarity search on excellent performance.

Acknowledgements

This work was supported by Shangluo university educational reform research project, and project number: 15jyjx117.

This work was supported by Shangluo university educational reform research project, and project number: 14jyjx104.

This work was supported by the Shaanxi Provincial Department of education scientific research projects, and project number: 12jk0950.

References

1. Agrawal R, Faloutsos C, Swami A.C. (2013) Efficient similarity search in sequence database. *Proceedings of 4th International Conference on Foundations of Data Organization and Algorithm*, Springer, New York , p.p. 69-84.
2. Keogh E, Chu S, Hart D et al. M. (2013) Segmenting Time Series: A Survey and Novel Approach. World Scientific Publishing Company: Singapore.
3. Sanghyun Park, Wesley W Chu, Jeehee Yoon et al. C. (2010) Efficient searches for similar subsequences of different lengths in sequence databases. *Proceedings of the 16th International Conference on Data Engineering*, Washington, U.S.A, p.p. 23-32.
4. K B Prat, E Fink. J.(2012) Search for patterns in compressed time series. *International Journal of Image and Graphics*, 2(1), p.p.89-106.
5. Aboy M, Hornero R, Ab Solo D, et a1.(2006) Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Transactions on Biomedical Engineering*, 53(11), p.p.2282 -2288.
6. Bagnall A, Janacek G. J. (2005) Clustering Time Series with Clipped Data. Machine Learning,

58(2), p.p. 151-178.
7. Agrawal R, Lin K-I, Sawhney H S, et a1.C.(1995) Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. *Proceedings of the 21th Very Large Data Bases*, VLDB Endowment, Zurich, Switzerland , p.p. 490 - 501.
8. Camerra A, Palpanas T, Shieh J, et a1.C. (2010) iSAX 2.0: Indexing and mining one billion time series. *Proc. of 2010 IEEE 10th International Conference on Data Mining*, Sydney, Australia, p.p. 58-67.
9. Fink E, Gandhi H S. J. (2011) Compression of time series by extracting major extrema. *Journal of Experimental &Theoretical Artificial Intelligence*, 23(2), p.p. 255-270.
10. Chawla S, Gionis A.C. (2013) Kmeans-:A unified approach to clustering and outlier detection. *Proceedings of the 2013 SIAM International Conference on Data Mining*, Texas, USA, p.p. 189-197.
11. Esling P, Agon C. J. (2012) Time-series data mining. *ACM Compute*r, 45(1), p.p.1-34.
12. Lin J, Li Y.C. (2009) Finding Structural Similarity in Time Series Data Using Bag-of-patterns Representation. *Proceedings of the 21st International Conference on Scientific and Statistical Database Management*, New Orleans, USA, p.p. 461-477.
13. Megalooikonomou V, Li G, Wang Q.C. (2004) A dimensionality reduction technique for efficient similarity analysis of time series databases. *Proceedings of the thirteenth ACM international conference on information and knowledge management*, ACM, Washington, USA, p.p.160-161.
14. Fu A W-C, Keogh E, Lau L Y, et al. J. (2008) Scaling and time warping in time series querying. *The VLDB Journal*, 17(4), p.p.899-921.

Framework of music Controller Based on Brain Computer Interface

Hu Jianfeng, Mu Zhendong, Yin Jinghai

Institute of Information Technology, Jiangxi University of Technology, Nanchang 330098, Jiangxi, China

Abstract
the study of Brain-Computer interface provides theoretical and practical basis for many brain electrical applications; for human, music can adjust mood, strengthen the immune ability, and improve the work efficiency. The existing music playback systems mainly focus on the convenience of control and simple music recommendation, but pay little attention to the mood of real-time users. Based on music control and recommendation, the paper designed one Brain-Computer interface system, which can intelligently control music; music control includes active control and passive recommendation, and passive recommendation is to automatically adjust music based on the physiological status of the users. Through the study on the recommendation accuracy, response time and comfort degree of the system, the system was proved to be feasible.
Key words: ELECTRO ENCEPHALO GRAPH, BRAIN-COMPUTER INTERFACE, MUSIC CONTROL , MUSIC RECOMMENDATION