Similarity Feature Extraction of EEG

Zhendong Mu

Institute of Information Technology, Jiangxi University of Technology, Nanchang, Jiangxi 330029, China

Keywords: Feature extraction; Similarity; Electroencephalograph (EEG) signal

Abstract. With the development of information technology nowadays, the analysis of mass data with time features has become possible. EEG feature extraction method is a key element of EEG research, this paper defines a method similarity of feature extraction. Five subjects are used to verify identification studies, which shows that the similarity method can extract to EEG features the subject. The analysis result of 400 collected EEG signals of the same mode shows that the similarity among features of the 400 samples is up to 80%.

Introduction

With the development of information technology nowadays, the analysis of mass data with time features has become possible. Therefore, the feature extraction and study of mass time series data have become the main research direction of data mining. A time series feature has two opposite characteristics: certainty and uncertainty. The certainty refers that in general many times of overall analysis of time series information corresponds to different results with a constant feature of time series. The uncertainty refers that for single time series information, such constant information is not complete or the time point when it occurs is uncertain. For solving the uncertainty, three mathematical theories are generated and they are fuzzy set theory, rough set theory and quotient space theory. These theories have been successfully applied in the fields, such as data mining, problem solving, pattern recognition, etc. Because corresponding features can be extracted from information itself using the rough set theory and this process needs no other knowledge, the extraction of many features is done using a rough set method.

During study of data mining, time series information includes many time-related information, such as meteorology, finance, management, astronomy, earthquake, etc. Such information is discrete and continuous and has different study methods respectively. Fox example, Tim Shimeall and Phil Williams[1] made trend analysis of computer network information security and divided the trend into: the trend of internal and external causes, temporal trend, space trend, associated trend and mixed trend. And they introduced trend analysis methods for different types of trend analyses.

In this article, the feature extraction of time series is realized by taking EEG signal as a study subject, collecting EEG signals via a proven EVP test mode, dynamically defining upper and lower approximation and a similarity calculation by using rough set method, and removing noise data gradually by adding time windows. The analysis result of 400 collected EEG signals of the same mode shows that the similarity among features of the 400 samples is up to 80%.

Feature extraction method

The EEG signal is a time series in the field of real numbers. As shown in Fig. 1, although the original EEG signal includes certain ERP features, they are hard to extract. The reason is that the EEG signal is a weak electric signal, the features of which does not occur completely in one or several signals and are easy to be disturbed by noise data. The feature extraction herein is done by referring to information table, using wave crest and valley change and adding windows. The detailed algorithmic method is as follows:

Step 1: n-sample EEG signal.

Step 2: Calculate the gradient of input EEG signals, keep crest values, and zero remaining signals, and the crest value vector on a time series is obtained in this way. Define the crest value to be 1 and the valley value to be -1, and the waveform vector of time series is obtained in this way. Classifying n-sample EEG signals: Classify the EEG signals of the same event into one category and mark them with 1, 2, 3..., respectively. The merging time span, TW is equal to 1.

Step 3: Define the window's size to be T and window's step to be S, and start calculating the similarity in the window of the similar type of EEG signals from start time point.

Step 4: According to the similarity calculated in Step 3, divide the wave crest and valley values of EEG signal into upper and lower approximation sets; and the span of merged points, TW is equal to TW plus 1;

Step 5: Make looping execution of Step 3 until all valley points are added to upper approximation or the merged span, TW is equal to 1/T.

Step 6: Output

The so-called event refers to the stimulus signals used to induce EEG signals from a subject during EEG signal collection;

The similarity used herein is divided into two parts: the shape similarity to indicate the change trend of time series and the value similarity to indicate the value of time series. If the definitions of two numbers ai and aj are operated as follows:

$$f(a_i, a_j) = a_i \oplus a_j = \begin{cases} 1 & a_i = a_j \\ 0 & a_i \neq a_j \end{cases}$$

The formula for calculating shape similarity Sis of n waveform vectors of time point x on a time series can be represented as follows:

$$Sis = \sum_{i,j=1}^{i,j=n} a_i \oplus a_j$$

For a classified threshold, with a set similarity, if exists, x time point of time series sample is of shape similarity.

For a time point x for n series samples, if mx1 stands for the maximum value of the vector of such time point, mx2 for the second maximum value, mi1 for the minimum value, and mi2 for the second minimum value, the similarity threshold of n time series values corresponding to the time point x can be calculated as follows:

$$\omega_x = \frac{|mx1 - mx2|}{|mi1 - mi2|}$$

If the time point x is of shape similarity, the similarity Siv of n vector values of the time point x can be calculated using the following formula:

$$Siv = \frac{\sum_{i=1}^{n}(a_i / \sigma)}{n\omega_x}$$

Where, ai refers to the vector value of the time point x, and the variance of vectors.

If the tolerance range interval for value similarity is [-Tx, Tx], the definition of upper and lower approximation sets of n time series samples is as follows:

$$B_-(x) = \{x \mid Siv_x \in [-Tx \; Tx]\}$$

$$B_-(x) = \{x \mid Sis \geq \lambda\}$$

Result

In order to verify the effectiveness of the method herein, we chose five subjects and each subject chose 400 effective EEG signals in the same event. Firstly, the superimposed average of such signals was calculated using superimposed and taken as feature bases.

For 400 samples from five subjects, the similarity of such subjects calculated using the above method is 1.8836, 0.9536, 2.1737, 2.3010 and 1.9181 respectively, as shown in Fig. 1. This indicates that the similarity calculation of time series features was successfully realized using the method.
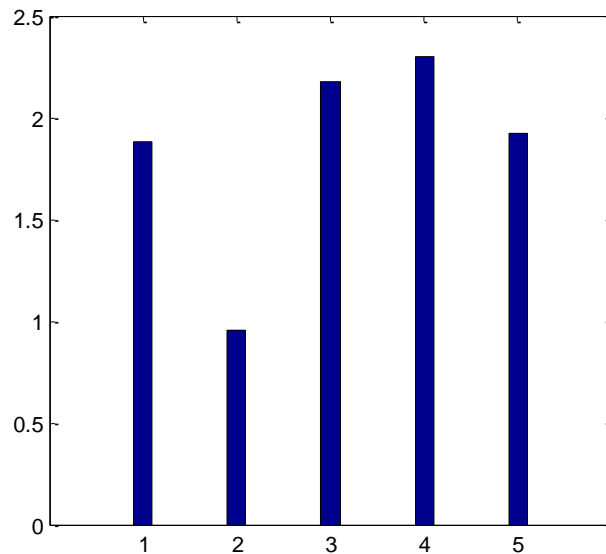

Fig. 1   EEG Similarity

References

[1] Shimeall, T. J., Williams, P. (2002). Models of information security trend analysis. Sensors Command Control Communications Intelligence Technologies for Homeland Defense Law Enforcement, p.p.43-52.

[2] Ostroff, J. S. (1989). Temporal logic for real-time systems. Research Studies Press, p.p.22-28.

[3] Bernad, D. J. (1996). Finding patterns in time series: a dynamic programming approach. Advances in knowledge discovery and data mining, p.p.229-248.

[4] Bazan, J., Skowron, A., Synak, P. (1994). Market data analysis: A rough set approach. ICS Research Reports, 6, 94.

[5] Golan, R., Edwards, D. (1994). Temporal rules discovery using datalogic/R+ with stock market data. In Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer, London, p.p.74-81.

[6] Golan, R. H., Ziarko, W. (1995). A methodology for stock market analysis utilizing rough set theory. In Computational Intelligence for Financial Engineering, 1995., Proceedings of the IEEE/IAFE 1995, p.p. 32-40.

[7] Anders, T. B. (1997). Mining Time Series Using Rough Set-A Case Study. In Proceeding of the First European Symposium, pp. 256-263.

[8] Yin, X., Shang, L.(2001) Study of Time Series Mining by Rough Set. Journal of Nanjing University, 37(3), p.p.182-187.

[9] Słowiński, R., Greco, S., Matarazzo, B. (2014). Rough-set-based decision support. In Search Methodologies, Springer, US, p.p. 557-609.

[10] Jia, X., Tang, Z., Liao, W., Shang, L. (2014). On an optimization representation of decision-theoretic rough set model. International Journal of Approximate Reasoning, 55(1), p.p. 156-166.

[11] Qian, Y., Zhang, H., Sang, Y., Liang, J. (2014). Multigranulation decision-theoretic rough sets. International Journal of Approximate Reasoning, 55(1), p.p. 225-237.

[12] Le Van Quyen, M., Martinerie, J., Baulac, M., Varela, F. (1999). Anticipating epileptic seizures in real time by a non‑linear analysis of similarity between EEG recordings. Neuroreport, 10(10), p.p. 2149-2155.