Biomedical Signal Processing and Control 21 (2015) 82–89

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Technical note

Feature selection method based on mutual information and class separability for dimension reduction in multidimensional time series for clinical data

Liying Fang[a,b,c,*], Han Zhao[a,b,c], Pu Wang[a,b,c], Mingwei Yu[d], Jianzhuo Yan[a,b,c], Wenshuai Cheng[a,b,c], Peiyu Chen[a,b,c]

[a] College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China
[b] Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China
[c] Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China
[d] Hospital of Traditional Chinese Medicine, CPUMS, Beijing 100010, China

ARTICLE INFO

Article history:
Received 7 December 2014
Received in revised form 28 March 2015
Accepted 14 May 2015

Keywords:
Multidimensional time series
Dimension reduction
Feature selection
Mutual information
Class separability

ABSTRACT

In clinical medicine, multidimensional time series data can be used to find the rules of disease progress by data mining technology, such as classification and prediction. However, in multidimensional time series data mining problems, the excessive data dimension causes the inaccuracy of probability density distribution to increase the computational complexity. Besides, information redundancy and irrelevant features may lead to high computational complexity and over-fitting problems. The combination of these two factors can reduce the classification performance. To reduce computational complexity and to eliminate information redundancies and irrelevant features, we improved upon a multidimensional time series feature selection method to achieve dimension reduction. The improved method selects features through the combination of the Kozachenko–Leonenko (K–L) information entropy estimation method for feature extraction based on mutual information and the feature selection algorithm based on class separability. We performed experiments on the Electroencephalogram (EEG) dataset for verification and the non-small cell lung cancer (NSCLC) clinical dataset for application. The results show that with the comparison of CLeVer, Corona and AGV, respectively, the improved method can effectively reduce the dimensions of multidimensional time series for clinical data.

1. Introduction

Time-series analysis is widely used in many application fields, including medical data, financial data, moving-object tracking, human-computer interaction interface [1,2], etc. Data mining for time series has very important value, such as research on the classification, clustering or prediction of data, which can assist in finding the potential rules of time series data and provide support. Currently, most researches focus on univariate time series processing. However, with the development of data-collection technology, more and more multidimensional time series data become available, which contain a considerable amount of potentially valuable information. For example, diabetes clinical data, as a kind of time series data, contain abundant information including food intake, drugs intake and daily activities. The EEG data[1] which contain plentiful information on brain waves reflect correlations with certain genetic predisposition and disease. In Tanawongsuwan and Bobick [3], 22 markers are spread over the human body to measure the movements of body parts while walking. In medicine, EEG data from 64 electrodes placed on the scalp are monitored to examine the correlation of genetic predisposition to alcoholism [4]. Therefore, in recent years, multidimensional time series classification, dimension reduction and similarity search technology have become common concerns for researchers in the field of data mining [5–7].

A time series is a series of observations,

$$x_i(t); \qquad [i = 1, \ldots, d; \quad t = 1, \ldots, n] \tag{1}$$

* Corresponding author at: Beijing University of Technology, College of Electronic Information and Control Engineering, No. 100, Pingleyuan Street, Beijing 100124, China. Tel.: +86 13810101581.
E-mail address: fangliying@bjut.edu.cn (L. Fang).

[1] http://archive.ics.uci.edu/ml/datasets/EEG+Database.

http://dx.doi.org/10.1016/j.bspc.2015.05.011

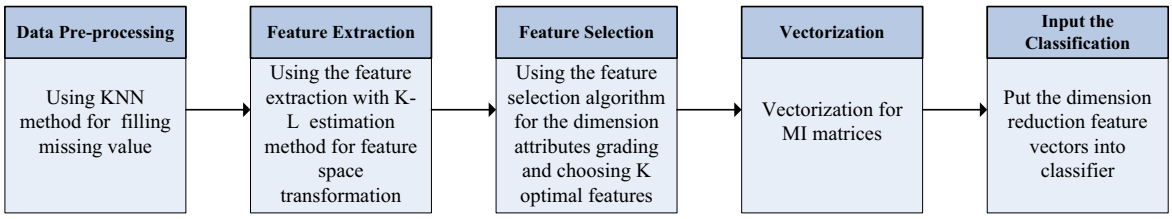

Fig. 1. MTS data dimension reduction process.

made sequentially through time where i indexes the measurements made at each time point t. It is called a univariate time series when d is equal to 1 and a multidimensional time series (MTS) when d is equal to or greater than 2. Due to the mass production of MTS data and the growing demand for classification in various fields, MTS classification techniques have been applied in many fields, such as the classification of RNA in bioinformatics, handwriting recognition and electrocardiogram (ECG) pattern matching. As MTS data are typical high-dimensional data [8], many features are either irrelevant or redundant. Moreover, dimension disasters, which are caused by excessive dimensions, exist in multidimensional feature space. Therefore, how to effectively select useful features for classification from the raw MTS data has become a current research hotspot with a high degree of difficulty.

Feature extraction and feature selection are the main methods of dimension reduction [9]. Not only can they reduce classification errors, but can also improve classification efficiency. Currently, feature selection methods are used widely in MTS including CLeVer [10] and AGV [11] based on PCA and Corona [12] based on a correlation coefficient method. However, they can only identify linear relationships among dimensions, and their calculations are more suited to dealing with equal length samples of MTS. However, unequal length data are indeed the norm in clinical follow-up because patients may die or otherwise be lost from the dataset. Mutual information (MI) is an important concept in information theory. MI can be applied to nonlinear transformation and extraction of high-order statistics. Therefore, we consider using MI for feature extraction to transform the different lengths of samples to equal length. Meanwhile, by the nonlinear relationship in multidimensional feature space, we can effectively reduce dimensions through feature selection. However, the probability density estimation method has a great influence on MI computation which implies whether the method can effectively and efficiently express the typical features to promote the accuracy of feature selection. Thus, it is significant to choose an applicable probability density estimation method for MI feature extraction in MTS. In addition, the feature subset evaluation criterion is the key issue in feature selection and its quality directly impacts the final result. The class separability criterion is one of the important evaluation criteria. Between-class distance criterion is one of the commonly used methods. We get better class separability by minimizing within-class distance and maximizing between-class distance simultaneously. The purpose of feature selection is to choose the feature subsets with larger class separability. However, since the redundant variables have an obvious effect on the result of classification, while the between-class distance criterion cannot eliminate the redundant variables, we consider that introduce a criterion with redundancy variable to eliminate redundancies and irrelevant features. We then introduce the improved method which can effectively choose the optimal features and reduce dimensions.

This paper aims to break the limitation that correlation matrices in traditional MTS feature selection method can only measure the linear relationships between variables. We improve the feature selection method based on mutual information and class separability. We first compute the MI value by a probability density estimation method to extract the linear and nonlinear relationship between variables through MI matrices. By considering the existence of redundancies we next introduce the feature selection algorithm based on class separability to eliminate redundancies and make high correlation between the chosen feature subsets and the target class. We then use the improved method for dimension reduction processing on MTS as is shown in Fig. 1. Finally, we verify that if the improved method can effectively reduce dimensions through the contrast experiments based on classification accuracy with an SVM classifier.

The remainder of this paper is organized as follows. Section 2 introduces the feature extraction method based on MI. Section 3 introduces the feature selection algorithm based on class separability. The experiment and result with the improved method is followed in Section 4, which is followed by conclusion in Section 5.

2. Feature extraction method based on MI

This section introduces the MI feature extraction method, which involves some basic concepts of entropy and MI as are shown in Refs. [13–15].

In general, a MTS can be expressed as a $d \times n$ matrix $[x_{i,t}]^{d\times n}$. Each matrix expresses one sample. Assume that these research data include several samples and that two of the samples are $[x_{i,t}]^{d\times n_1}$ and $[x_{i,t}]^{d\times n_2}$. Generally speaking, each variable of within-sample sampling time has the same length. However, the length of two samples of between-sample sampling time t as n_1 and n_2 is not always the same. Therefore, each MTS sample is expressed by a $d \times t_j$ matrix $[x_{i,t}]^{d\times t_j}$.

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1,t_j} \\ \dots & \dots & \dots & \dots \\ x_{i1} & \dots & \dots & x_{it_j} \\ x_{d1} & x_{d2} & \dots & x_{dt_j} \end{bmatrix}, \qquad \left(i = 1, 2\ldots, d; \quad t = 1, 2, \ldots, t_j\right) \quad (2)$$

where $x_{i,t}$ denotes the sampling value of the variable x_i with the ith dimension at time point t. M_j substitutes for the jth sample matrix $[x_{i,t}]^{d\times t_j}$ as is shown in Fig. 2. t_j Denotes the sampling time length of the jth sample. X_i shows the index sequence of the ith dimension. Because each sequence X_i has different degrees of importance to classification, X_i is expressed in different colors and that a deeper color means a higher degree of importance. However, under the initial condition, for degree of importance for each sequence is unknown, the colors are shown in random depth. Fig. 3 shows a MTS dataset with n samples and each sample is a matrix with dimension d and sampling time length t_j. For any given sample, the degree of importance of each sequence is initially unknown.

By the definition of information entropy and MI, the probability density distribution of random variables must be approximately estimated before MI calculation. One kind of probability density estimation method based on nearest neighbor is introduced in [16], which has good effect used in [17,18] as well. The advantage of this method is that there is no need to estimate the probability density distribution function for any variables.