Neurocomputing 171 (2016) 744–753

Contents lists available at ScienceDirect

Neurocomputing

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

Accurate and efficient classification based on common principal components analysis for multivariate time series

Hailin Li [a,b,*]

[a] College of Business Administration, Huaqiao University, Quanzhou 362021, China
[b] Research Center for Applied Statistics and Big Data, Huaqiao University, Xiamen 361021, China

ARTICLE INFO

Article history:
Received 3 November 2014
Received in revised form
25 June 2015
Accepted 6 July 2015
Communicated by P. Zhang
Available online 15 July 2015

Keywords:
Classification
Common principal components analysis
Data mining
Multivariate time series

ABSTRACT

Multivariate time series are found everywhere and they are important data in the field of data mining, but their high dimensionality often hinders the quality of techniques employed for classifying multivariate time series. In this study, we propose an accurate and efficient classification method based on common principal components analysis for multivariate time series. First, multivariate time series are divided into several clusters according to the number of class labels, and the high dimensionality of multivariate time series can then be reduced by common principal components analysis, which gives the reduced principal component series sufficiently high variance. Second, each cluster is used to construct the corresponding reduced coordinate space formed by the eigenvectors of the common covariance matrix. Third, any multivariate time series without a class label can be projected onto these coordinate spaces and its label can be predicted based on the minimal variance of the reduced principal components series according to the different projections. Our experimental results demonstrated that the proposed method for the classification of multivariate time series is more accurate and efficient than existing methods. It is also flexible for multivariate time series with different lengths.

1. Introduction

Multivariate time series (MTS) are very common in many fields, such as stock time series in finance [1], electroencephalograms in medicine [2], and geophysical data recorded from monitoring networks [3]. To extract knowledge from these data, many techniques are used in the fields of data mining, particularly the classification of MTS.

In the field of time series data mining, classification is often applied to univariate time series (UTS) to address many scientific problems. However, due to the high dimensionality and numerical values of time series, classic methods such as ID3 [4], C4.5 [5], CART [6], and neural networks [7] have difficulty classifying these data. MTS have time-based and variable-based dimensions, both of which make the classic methods inefficient. However, the *K* nearest neighbors classifier method appears to be used widely to classify time series. For example, Keogh et al. [8] used this method to search for similar objects and to forecast the class labels of UTS. Yu et al. [9] proposed a nearest neighbor classifier, which is considered to be an effective method for UTS classification. Yang and Shahabi [10] also proposed an efficient *K* nearest neighbor search method for MTS, while Weng and Shen [11,12] proposed MTS classification based on 1-nearest neighbor classifier. In our previous studies [13–15] of time series data mining, we combined the 1-nearest neighbor classification method with other algorithms.

Nearest neighbor classifiers are very popular and valid in various fields, but the raw method is not used directly because the performance of these classifiers is not good with MTS. Therefore, before executing the nearest neighbor classifier, some preliminary processes must be performed: dimensionality reduction and the use of a similarity measure. Before MTS data mining, dimensionality reduction is used to reduce the computational time for related algorithms such as clustering, classification, motif/pattern recognition, and abnormal detection, while it can also improve the quality of the data mining results. Various methods can be used to reduce the dimensionality of MTS, including singular value decomposition (SVD) [16], principal components analysis (PCA) [17], independent components analysis [18], as well as their extensions and variations [19,11,12]. However, PCA may be the most commonly used method. PCA can be used to transform MTS into some principal component sequences, which have less dimensions than the original. Moreover, the principal component series (PCS) retains most of the information about the original MTS. In addition, a good function is very important for measuring the distance (or similarity) between two groups of features in a MTS. In particular, the nearest neighbor classifier is often based on a distance measure. Thus, a suitable distance function for

* Correspondence address: College of Business Administration, Huaqiao University, Quanzhou 362021, China. Tel.: +86 595 22693815.
E-mail address: hailin@mail.dlut.edu.cn

http://dx.doi.org/10.1016/j.neucom.2015.07.010
0925-2312/

measuring the similarity between two MTS items is also important for classification. In practice, Euclidean distance [20] and dynamic time warping (DTW) [21] are two of the most popular methods. The former is a fast method, but its quality is easily affected by abnormal data points, while it often requires that the time series are equal in length. DTW is a robust method, but the time and the space complexity of its computation are very large, which make it unsuitable for long time series with excessive numbers of variables.

In most cases, MTS have different lengths. Most of the traditional techniques can reduce the variable-based dimensions but the number of time-based dimensions is retained; thus, MTS with different lengths will yield sequences that represent the MTS with different lengths. For instance, PCA transforms MTS into the corresponding PCS with different lengths, which means that distance functions such as DTW must be used to measure the similarity. However, although DTW is an effective approach, it requires excessive amounts of time and space to measure distances.

To overcome the problems mentioned above, we propose an accurate and efficient MTS classification method. The main motivations of our study are summarized as follows. First, we analyze the weaknesses of the traditional common PCA (CPCA) applied to the field of time series data mining. The effectiveness and the efficiency of CPCA are regarded as shortcomings of the algorithms used to extract knowledge from MTS datasets; thus, it is necessary to design a novel method to address these difficulties. Second, the traditional methods based on PCA often fail to handle MTS data with different lengths; therefore, the proposed method should consider various lengths and improve the quality of PCA for mining time series.

In this study, various MTS clusters are transformed to construct the corresponding reduced subspaces by CPCA. Each space is then organized by the eigenvectors of the common covariance matrix in a cluster. The MTS without class labels in the test dataset are projected onto the different subspaces and the minimal variance in the reduced PCS according to different projections can specify the label values for the MTS in the test dataset. The two main contributions of our proposed method are as follows. First, MTS items with the same label in a cluster are used to construct the subspace, which means that the PCS of any MTS item projected onto the corresponding subspace has a large variance. Thus, when the variance in the PCS derived from different subspaces is larger, the projected item and those in the corresponding cluster are more similar. Second, we treat the largest variance in the different subspaces as a classifier with high efficiency, which improves the performance of the proposed method. Three advantages may be obtained, as follows. (1) Our proposed method is faster at classifying MTS compared with the existing methods based on PCA. (2) The quality of the classification results obtained by the proposed method is often better than that with traditional method. (3) The proposed method is suitable for the classification of MTS datasets where the lengths of the MTS items are different. The results of our experimental evaluation also indicated that the proposed method is more accurate and efficient.

The remainder of this paper is organized as follows. Background and related work are introduced in Section 2. In Section 2, we describe the proposed new classification method. The results of experimental evaluations of the proposed method are presented in Section 4. In the final section, we give our conclusions and discuss future work.

2. Background and related work

Due to the high dimensionality of MTS, techniques for dimensionality reduction are very important for time series data mining, and PCA [15,17,10] is one of the most commonly used methods. In addition, compared with the traditional methods, CPCA can often improve the performance of the algorithms used in MTS datasets. In this section, we introduce both these methods and we review related work.

2.1. PCA

PCA is used widely for MTS dimensionality reduction. PCA can transform a MTS $X=\{x_1, x_2, \ldots, x_m\}$ into a PCS $Y=\{y_1, y_2, \ldots, y_m\}$, where m is the number of variable-based dimensions and y_i denotes the ith principal component sequence. Moreover, the first principal component sequence y_1 contains most of the information about the original MTS and y_2 contains the second highest amount of information, and so on. In fact, each principal component sequence y_i is a linear transformation of the variables in the original MTS and the coefficients defined in this transformation are considered as weight vectors, i.e.,

$$y_i = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{mi}x_m, \quad i = 1, 2, \ldots, m, \tag{1}$$

where a is the corresponding weight and x_j denotes the jth variable of the MTS. Moreover, the first principal component sequence y_1 has the largest variance, $\lambda_1 = Var(y_1)$, the second principal component sequence accounts for the largest portion of the remaining variance, $\lambda_2 = Var(y_2)$, and so on. In this manner, the first p component sequences may retain most of the variance present in all of the original m variables, where $p < m$. Thus, the dimensionality reduction for a MTS with m variables can be achieved by projecting it onto the p- dimensional subspace (also called the coordinate space). The subspace can be constructed by an eigenmatrix of the covariance matrix Σ of X. According to the SVD, the covariance matrix can be decomposed by

$$\Sigma = U\Lambda U^T, \tag{2}$$

where U contains the weights for the principal component sequences and the matrix Λ has the corresponding variances, which means that the first column vector of U is the weight vector of the first principal component sequence and its variance is the first element of the matrix Λ along the diagonal.

To reduce the dimensionality of MTS, the first p principal component sequences are retained, which means that the first p eigenvectors are used to construct the subspace, i.e., $A_{m\times p} = U(1:m, 1:p)$. Thus, the reduced PCS can be formed by

$$Y_{n\times p} = X_{n\times m}A_{m\times p}, \tag{3}$$

where n is the length of the MTS, where it is often the case that $p < m$ and $p < n$. In this manner, we can transform a MTS with a size of $n \times m$ into a reduced representation with a size of $n \times p$.

In addition to dimensionality reduction and feature representation using PCA, some distance functions are often used to measure the similarity between two representations following the transformation of the MTS by PCA. The angles between all the combinations of the selected principal components can be used to measure the similarity [22]. Another approach was proposed by [23] for modifying previous methods by weighting the angles with the corresponding variances. Ref. [17] addressed the issue of similar principal components present in a time series by using the different values of the variables. Refs. [24,10] proposed *Eros* based on the acute angles between the corresponding components, which can measure the similarity better and faster than previous methods. A fast similarity search for MTS using a projection comparison based on PCA was proposed by Karamitopoulos et al. [25]. In addition, since PCA is based on SVD, some methods based on SVD [11,12] have been applied to dimensionality reduction and as similarity measures for MTS.