



Dimensionality reduction for multivariate time-series data mining

Xiaoji Wan^{1,2} · Hailin Li^{1,2} · Liping Zhang¹ · Yenchun Jim Wu^{3,4} 

Accepted: 29 December 2021 / Published online: 19 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

A multivariate time series is one of the most important objects of research in data mining. Time and variables are two of its distinctive characteristics that add the complication of the algorithms applied to data mining. Reduction in the dimensionality is often regarded as an effective way to address these issues. In this paper, we propose a method based on principal component analysis (PCA) to effectively reduce the dimensionality. We call it “piecewise representation based on PCA” (PPCA), which segments multivariate time series into several sequences, calculates the covariance matrix for each of them in terms of the variables, and employs PCA to obtain the principal components in an average covariance matrix. The results of the experiments, including retained information analysis, classification, and a comparison of the central processing unit time consumption, demonstrate that the PPCA method used to reduce the dimensionality in multivariate time series is superior to the prior methods.

Keywords Covariance matrix · Piecewise representation · Principal component analysis · Time-series data mining

✉ Yenchun Jim Wu
wuyenchun@gmail.com

Xiaoji Wan
wanxiaoji@hqu.edu.cn

Hailin Li
blihailin@163.com

Liping Zhang
zhanglp@hqu.edu.cn

¹ College of Business Administration, Huaqiao University, Quanzhou 362021, Fujian, China

² Oriental Enterprise Management Research Center, Huaqiao University, Quanzhou 362021, Fujian, China

³ College of Humanities and Arts, National Taipei University of Education, Taipei 106, Taiwan

⁴ Graduate Institute of Global Business and Strategy, National Taiwan Normal University, Taipei 106, Taiwan

1 Introduction

A time series is a type of time-dependent and high-dimension data, which widely exist in the economy [1], finance [2], engineering [3], marketing [4], the Internet of things (IoT) [5], and other fields. In recent years, research on time-series data mining (TSDM) has attracted researchers in different fields. However, more research reports have been conducted on univariate time series (UTS) than multivariate time series (MTS). The reason is that the relationships among the variables in MTS are difficult to obtain accurately, and the high dimensionality of the variables is also an obstacle in MTS data mining. For example, the IoT temporal data collected by different sensors in mobile edge computing (MEC) are typically high dimension. To predict them more accurately, it is necessary to this high-dimensional characteristic.

In order to effectively decrease the complications of data mining in time-series datasets, some scholars adopt the methods used in dimensionality reduction, including feature selection [6, 7] and feature representation. Feature representation involves local auto patterns [8], discrete wavelet transformation [9, 10], shape space representation [11, 12], piecewise linear approximation [13, 14], piecewise aggregate approximation [15–17], and symbolic approximation [18–20]. However, these methods mainly focus on UTS in terms of the time axis. This means that a long time series can be reduced to a short sequence that retains the most important information. However, because MTS has two types of dimensions (time and variable), most of the above-mentioned methods do not succeed in reducing the dimensionality of MTS. Therefore, in doing so, the two distinctive characteristics of MTS must be considered simultaneously.

Some existing methods can be used to reduce the dimensionality in variable characteristics, such as singular value decomposition (SVD) [21, 22], principal component analysis (PCA) [23, 24], and independent component analysis (ICA) [25]. The principles of SVD and PCA are often the same and are based on projection transformation so as to ensure that the projected data have the maximum variance. According to the order of the variances, the first few principal components are retrieved and used to represent the original time series. ICA is an extension of PCA and factor analysis, which is derived from blind source separation. It can retrieve some independent components that are hidden in the datasets.

In general, the above-mentioned methods are often combined with distance measurements to mine MTS data. Krzanowski [26] presented a similarity method using the cosine angle between each corresponding principal component. Singhal and Seborg [27] proposed a novel distance measurement method S_{dist} to determine the similarity in the pair of principal components with the same order of variance. Karamitopoulos et al. [28] used PCA to obtain the transformation space of the queried time series. Moreover, they projected the other time series into space and reconstructed the new time series. The fitting error between two reconstructed time series was regarded as the distance between the query time series and the queried one. Goetschalckx et al. [29] jointly employed SVD, retraining, pruning, and clustering to achieve better compression in neural networks. Weng and Shen

[30] proposed a two-dimensional SVD (2dSVD) and applied Euclidean distance to measure the similarity between two MTS. Wu and Yu [31] used Fast ICA [32] and a suitable distance function to cluster the independent principal components for MTS. In addition, some scholars [33] also applied the methods of independent component analysis (ICA) and ensemble empirical mode decomposition (EEMD) to explore the underlying factors of single financial time series.

Principal components extracted from MTS are based on the covariance of every pair of the variables. Each variable is indeed a time series that describes all the information on the corresponding variable. However, as the length of MTS increases, the covariance might not be able to reflect the relationship between the two variables, which leads to incomplete principal components. In this paper, we segment MTS into subsequences, each of which is used to calculate a covariance matrix that reflects the relationships between variables in a more detailed fashion. Moreover, an average covariance matrix derived from all the covariance matrices can be obtained. We call the method a piecewise representation based on PCA (PPCA).

The proposed method has some advantages over the traditional methods as follows. (1) The local information described by the covariance matrix of the subsequence is taken into consideration, which provides more detailed information for representing MTS. (2) The average covariance matrix to some extent reflects the overall information on MTS, so some important characteristics are retained by PPCA. (3) The experimental results reveal that the quality of TSDM is not proportionate to the quantity of data information. This means that sometimes local information, when applied to representation and data mining, can yield good results.

The remainder of this paper is organized as follows. In Sect. 2, we provide background materials and discuss related work about PCA. In Sect. 3, we present a new algorithm for representing MTS. Three kinds of evaluation experiments are described in Sect. 4. Finally, the conclusions and future directions are discussed in Sect. 5.

2 Background and related work

PCA is one of the most important methods used for reducing the dimensionality of MTS. It can handle four major distortions that should be considered, namely, offset translation, time warping, amplitude scaling, and noise. These properties indicate that it is a robust method for reducing dimensionality and retains the most important characteristics of MTS. Therefore, in this section, we explain how PCA works and the ways often used to measure the similarity between two groups of principal components.

2.1 Principal component analysis

Let us suppose that X denotes a multivariate time series with m variables, and its length is n . This means that an MTS can be written as $X_{n \times m}$, and the n observations with m variables, according to the time order, comprise the entire time

series. Let S denote an orthogonal matrix with m unitary column vectors of length m . The goal of PCA is to project MTS $X_{n \times m}$ onto a new space $S_{m \times m}$ through a linear transformation of Eq. (1).

$$Y_{n \times m} = X_{n \times m} S_{m \times m} \quad (1)$$

In this way, Y is the representation of X in the new space S . The property quality of Y is dependent of the orthogonal matrix S . This means that the better the new space S can describe the observations, the more noticeable the features are.

Actually, PCA is the linear transformation of the original variables and the coefficients. The coefficients make up the new space. To construct these coefficients (or the new space), PCA is often performed with singular value decomposition (SVD) to a covariance matrix of MTS X . If Σ denotes the covariance matrix of MTS X , it can be calculated with the following equation.

$$\Sigma = \text{cov}(X) = E[(X - E[X])(X - E[X])^T] \quad (2)$$

According to the properties of SVD, when a covariance matrix defined by Eq. (2) is decomposed by SVD, then we have

$$\Sigma = U \Lambda U^T \quad (3)$$

The matrix U can be used to denote the new space S and contains the variables' loadings for each principal component. Meanwhile, the diagonal elements of the matrix Λ are the corresponding variances. The bigger the variance is, the more information that the data project onto the corresponding vector.

However, according to Eq. (1), because Y is still equal to X in terms of dimensions, the dimensionality of X is not reduced. In fact, the dimensions of Y depend on the size of the space S (here, $S = U$). PCA picks up a new space coordinates system to describe the observations of MTS X . The new system often consists of k orthogonal column vectors of S , that is, $S(:, 1 : k)_{m \times k}$. Thus, the equation becomes

$$Y_{n \times k} = X_{n \times m} S_{m \times k} \quad (4)$$

In this way, PCA can reduce the dimensionality of MTS [34, 35]. The dimension is decreased from m to k , where $k < m$. The matrix Y also can be regarded as the feature matrix of the original MTS X .

Because of the performance of dimensionality reduction and feature extraction, PCA has not only been widely used in facial recognition but also applied to MTS data mining. Huang et al. [36] use PCA to split large MTS clusters into some smaller clusters. Barragan et al. [37] propose a method to recognize patterns in MTS based on the combination of wavelet features, PCA similarity metrics, and fuzzy clustering. The results demonstrate that it is efficient compared with other traditional approaches in a fault detection and diagnosis problem. Some other applications in MTS data mining are based on some measurements that are introduced in the next section. In addition, some extended versions of PCA are also applied to MTS data. 2dSVD is based on two-dimensional MTS matrices rather

than one-dimensional vectors, which considers MTS row-row and column covariance matrices to obtain a feature matrix [30].

Li [38] proposed an approach based on the full dataset and constructed a common principal component as a common projection space. They called this approach common principal component analysis (CPCA), which is based on the notion of a common subspace across all multivariate data items, and this subspace should be spanned by the orthogonal components. For an MTS dataset D , the common subspace $S = (S_1, S_2, \dots, S_N)$ can be defined by orthogonal components, as shown in Eq. (5).

$$\bar{\Sigma} S_i = \lambda_i S_i \quad (5)$$

where $\bar{\Sigma}$ is an average covariance matrix, that is, $\bar{\Sigma} = \Sigma_i / N$, and Σ_i is the covariance matrix of the i th MTS. $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ and $S = (S_1, S_2, \dots, S_k)$ are the eigenvalue vector and eigenvector matrix of the average covariance matrix $\bar{\Sigma}$, respectively. In this way, every MTS can be projected onto this subspace, and we can obtain the feature sequences for each MTS according to Eq. (4).

2.2 PCA-based measurements

PCA is employed to reduce the dimensionality of data such as images, speech, music, and MTS. The features extracted by PCA are the representation of the data. However, in most cases, valuable information and knowledge are still hidden in these features. In data mining, most of the algorithms, such as clustering, classification, and pattern recognition, need to measure the similarity (or distance) between two objects. For these reasons, PCA-based measures have been proposed to mine the knowledge in MTS datasets.

Krzanowski [26] used PCA to obtain the principal components and retained the first k components to represent the features of the original time series. Moreover, they regarded the sum of cosine values of every angle between all the combination of the selected principal components as similarity (S_{PCA}). Later, another method [39] modified the previous methods by weighting the angles with the corresponding variances. An improvement of similarity measurement (Extended Frobenius norm, Eros) based on the acute angles between the corresponding components, instead of all the components in the previous method, is proposed by [40]. Karamitopoulos et al. [41] proposed a distance measure that does not require the query object to be PCA represented for a time-series similarity search.

Because Eros can measure the similarity of two MTS with unequal lengths and can obtain better results than other distance functions, such as dynamic time warping (DTW) [42, 43], Euclidean distance (ED) [44] and S_{PCA} [26], here we describe Eros in detail.

Eros is based on observations from both S_{PCA} and the Frobenius norm, which can easily calculate the similarity of two matrices. Suppose there are two MTS X_1 and X_2 of size $n_1 \times m$ and $n_2 \times m$, respectively. Let U_1 and U_2 be two right eigenvector matrices of their covariance matrices, Σ_1 and Σ_2 , respectively. The Eros similarity of MTS X_1 and X_2 can be defined as Eq. (6).

$$\begin{aligned} \text{Eros}(A, B, w) &= \sum_{i=1}^m w_i |u_{1i}, u_{2i}| \\ &= \sum_{i=1}^m w_i |\cos(\theta_i)| \end{aligned} \quad (6)$$

where u_{1i} and u_{2i} are the i th column orthonormal vectors of length m . u_{1i}, u_{2i} is the internal product of the two vectors. w is a weight vector that can be set by the eigenvalues of MTS datasets. For more detailed information about Eros, see [40].

Because Eros is based on an eigenvector matrix of the covariance matrix, it is easy to see that Eros is suitable for measuring the similarity of the MTS with the same number of variables and unequal lengths. However, it is independent of the information in other MTS. This means that the Eros between two MTS X_i and X_j depends only on information from the two MTS, which has nothing to do with the other MTS. By comparison, other methods are often based on the entire MTS dataset.

3 Piecewise representation based on PCA

Over time, the MTS will be longer, which might make the relationships between any two variables inaccurate. Moreover, traditional methods based on PCA are used to decrease the variable-based dimension by taking MTS as a whole into consideration. For an MTS dataset in which the lengths are short, in most cases, they reflect the relationships in the data observations. However, when the MTS in the dataset are long, the relationships become more complex. In other words, the local relationships sometimes are more important than the entire relationships.

To understand why the local relationships are important and motivate this work, we provide the following example. Suppose there exists an MTS X with two variables x_1 and x_2 , as shown in Fig. 1. The sequences of the two variables are quite similar, except for the two parts, with the time ranges (51,100) and (151,200). We draw the MTS with two variables in Fig. 2. It is easy to see that the two variables have a high correlation. x_2 is the drift variable of variable x_1 , which can also be regarded as an asynchronous relationship [42, 45]. In practice, drift phenomena exist in the

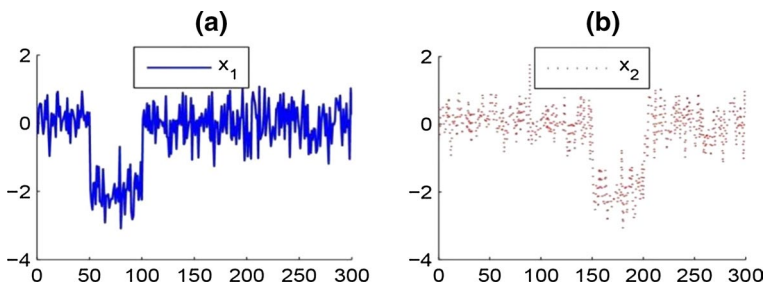


Fig. 1 Two variables, x_1 and x_2 , in an MTS X

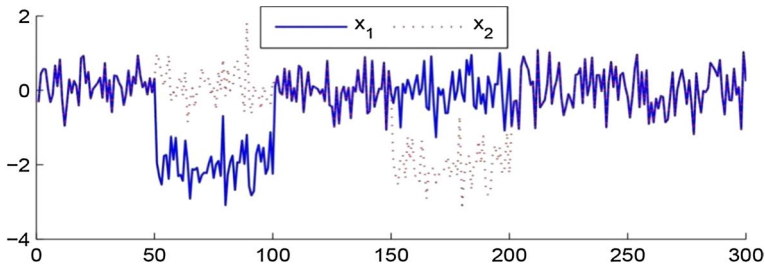


Fig. 2 The MTS X with two variables, x_1 and x_2 , with a length of 300

variables for MTS. For example, the stock portfolio should take the drift phenomena into consideration, which in economics is also called a comovement relationship between two stock markets. Therefore, it is necessary to consider the drift phenomenon when detecting knowledge in MTS databases.

However, PCA, which is used to calculate the principal components, often takes the MTS as a whole into consideration. In other words, the traditional PCA-based reduction in dimensionality considers all the information, rather than the local information. In some cases, such as the example mentioned earlier, the local information is more important than all of it, as it can detect the differences between the two variables. PCA is based on a covariance matrix or relationship matrix. Before extracting principal components, it must calculate the covariance matrix of MTS. For this example, the covariance matrix Σ_0 and the correlation coefficient matrix R_0 are

$$\Sigma_0 = \begin{pmatrix} 0.8405 & 0.0557 \\ 0.0557 & 0.8565 \end{pmatrix} \text{ and } R_0 = \begin{pmatrix} 1.0000 & 0.0656 \\ 0.0656 & 1.0000 \end{pmatrix}$$

According to statistical theory, the larger the absolute value of the covariance is, the more correlated the two variables will be. In addition, the closer to one the correlation coefficient is, the more correlated the two variables will be. The elements $\Sigma_0(1, 2) = \Sigma_0(2, 1) = 0.0557$ and $R_0(1, 2) = R_0(2, 1) = 0.0656$ deceptively indicate that the two variables are not very correlated. However, the two variables are indeed highly correlated. So, when PCA is used to reduce dimensionality and represent MTS, the true relationship between every pair of variables should be revealed.

To address these issues, we propose a new method using PCA to represent MTS. It is called piecewise representation based on PCA (PPCA). Because the local information is important for a long MTS, we can segment the original MTS $X_{n \times m}$ along with the time direction into several short sequences $\hat{X} = \{\hat{X}_1, \hat{X}_2, \dots, \hat{X}_w\}$. Every short sequence \hat{X}_i has the local information on MTS. The covariance matrix Σ_i of every sequence \hat{X}_i is calculated, and then the average covariance matrix Σ_a can be obtained with Eq. (7).

$$\Sigma_a = \frac{1}{w} \sum_{i=1}^w \Sigma_i \tag{7}$$

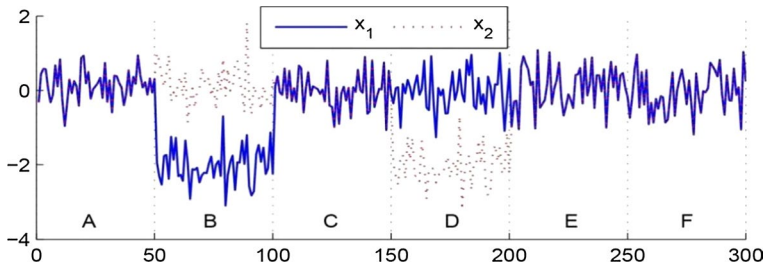


Fig. 3 An MTS with the variables x_1 and x_2 is segmented into 6 sequences

Table 1 The pseudo-code of the algorithm of PPCA

Input: An MTS $X_{n \times m}$ and w number of segments

Output: The feature sequence Y and the transformation space U

- (1) Segment the MTS X into w short sequences, that is, $\hat{X} = \text{Segment}(X, w)$
- (2) Calculate the covariance matrix Σ_i of the sequences \hat{X}_i , namely, $\Sigma_i = \text{cov}(\hat{X}_i)$ and $i = 1, 2, \dots, w$
- (3) Calculate the average covariance matrix Σ_a according to Eq. (7)
- (4) Use Eq. (8) to decompose Σ_a and obtain its eigenvector matrix as the transformation space U
- (5) Return feature sequence Y by $Y_{n \times m} = X_{n \times m} U_{m \times m}$

The average covariance matrix Σ_a in Eq. (7) is based on the local information on MTS. It comprehensively reflects the relationships between every two variables for MTS. For this example, when an MTS with a length of 300 is segmented into 6 equal short sequences, the length L is 50—that is, $L = 50$. The average covariance matrix Σ_a and the corresponding correlation coefficient matrix $R_a = \frac{\sum_{i=1}^w R_i}{w}$ can be obtained as follows:

$$\Sigma_a = \begin{pmatrix} 0.2488 & 0.1756 \\ 0.1756 & 0.2470 \end{pmatrix} \text{ and } R_a = \begin{pmatrix} 1.0000 & 0.7051 \\ 0.7051 & 1.0000 \end{pmatrix}$$

As shown in Fig. 3, the sequences marked B and D can be used to distinguish the two variables, and the other sequences, marked A , C , E , and F , are the contribution of the correlation. $\Sigma_a(1, 2) = \Sigma_a(2, 1) = 0.1756$ and $R_a(1, 2) = R_a(2, 1) = 0.7051$ are larger than Σ_0 and R_0 . This indicates that the two variables are highly correlated, which coincides with the true relationship.

Finally, SVD can be used to decompose Σ_a and obtain the transformation space U according to Eq. (8).

$$\Sigma_a = U \Lambda U^T \tag{8}$$

Thus, we can use Eq. (4) to reduce the dimensionality and obtain the feature sequences of MTS. The algorithm of PPCA can be described by the pseudo-code in Table 1.

According to the PPCA algorithm, if we choose the first k eigenvectors as the transformation space, that is, $U_{m \times k}$, where $k < m$, then the dimensionality reduction can be achieved with Eq. (4). However, the segment function in the algorithm can use either equivalent segmentation or adaptive division. In this paper, we only discuss feature representation based on equivalent segmentation. So, we can segment the MTS of length n into w multivariate sequences of length $L = n/w$. In addition, the time complexity of PPCA is the same as that of PCA: $O(nm^2) + O(m^3)$. In practice, PPCA has some additional time cost, which is caused by some auxiliary statements in the PPCA procedure, such as the segment calculation and calculation of the average covariance matrix. This analysis of time consumption is confirmed in the section on experimental evaluation.

4 Experimental evaluation

To test the performance of the proposed method, we designed an experimental evaluation made up of three parts: comparison of the retained information, classification evaluation, and time consumption analysis. Meanwhile, because PCA and other traditional PCA-based dimensionality reduction methods consider only the entire information, rather than the local information, we compared the proposed PPCA method to three existing ones: PCA [23, 23, 24], CPCA [38], and RTS. RTS denotes the Euclidean distance based on raw MTS.

In addition, three UCI datasets were selected in the experiments: EEG Eye State (EEGEye), the brain computer interface (BCI), and EEG database (EEG), respectively. The lengths of MTS in the first dataset were unequal, and the lengths of MTS in the other two datasets were equal, which allows the methods to perform on MTS datasets based on different lengths.

EEGEye has 19 MTS with lengths of more than 100. Moreover, each is different and has 14 variables. They are labeled 0 or 1, indicating the eye-open and eye-closed states. BCI has 2 classes, 316 training and 100 test trials of 28 EEG channels, and 500 samples. Here we only chose the training dataset with class labels. The EEG database has 20 MTS with 2 classes whose lengths are 256 and have 64 attributes.

The experiments were implemented with a Windows 7 system on a one quad-core Intel i7-2640 M clocked at 2.80 GHz with 8 GB of memory. Moreover, the related programs were compiled with Matlab R2012b.

4.1 Retained information comparison

The feature representations $Y_{n \times k}$ based on PCA of the MTS data are often influenced by the component weight matrix (or transformation space) $U_{m \times k}$ and the variances matrix (or the singular values) Λ . In particular, the eigenvalue (the diagonal element of Λ) provides the variance of the corresponding component Y_i . Moreover, the bigger the eigenvalue is, the more retained information that can be obtained. So, in this experiment, PPCA is compared to the other two methods, PCA and CPCA, for the retained information.

The researchers performed the three methods with the BCI_Train dataset. In this case, the reduced length L for PPCA is 100. We obtain the percentage of the sum of the first k eigenvalues for each MTS. The average percentage can be regarded as the result of the retained information rate to be compared for each method. The retained information comparison is shown in Fig. 4, based on the different reduced dimensions.

The comparison shows that the three methods yield different retained information based on having different k . Moreover, PPCA seems to fall in between PCA and CPCA in this respect. However, we should point out that having more retained information does not always have good results in data mining. For example, the original MTS with full information often need to have its dimensionality reduced so that it can obtain better results, which indicates that too much concentration on retained information could obscure more important features in MTS. Therefore, the retained information from the proposed PPCA method is not the best, but it considers local features that can distinguish between two MTS. As analyzed in the previous section, we know that PPCA can handle the drift phenomena better than other methods. Actually, the information retained by PPCA is close to that of PCA, as shown in Fig. 5. When L equals the length of MTS, PPCA will degenerate to PCA.

4.2 Classification

MTS classification is one of the important technologies applied to data mining. In this experiment, we perform the four methods (PPCA, PCA, CPCA and RTS) on the three MTS datasets. We choose the nearest-neighbor classification as the classifier. We then let each MTS search the most similar object in the rest of the dataset. This means that if MTS X_i searches the most similar object in the datasets D , then X_i must compare the rest $X_j \in \{D - X_i\}$ using the four methods. Because all MTS in the datasets have labels, we can check the labels of MTS X_i and X_j . If they are identical, we regard it as the correct classification; otherwise, it is the wrong classification.

The classification should be combined with the distance functions. Because of the superiority of Eros, we use it as the distance function to measure the

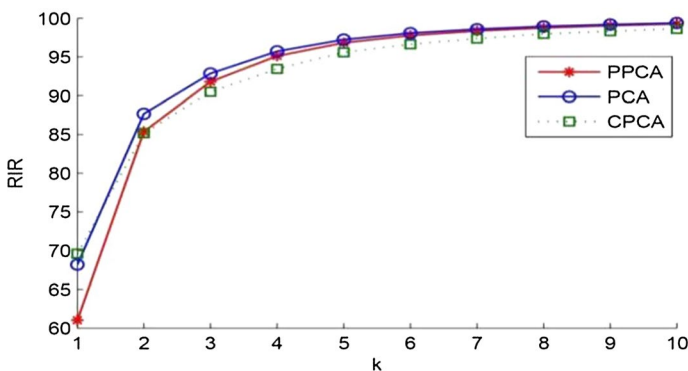


Fig. 4 Comparison of retained information based on different reduced dimensions k

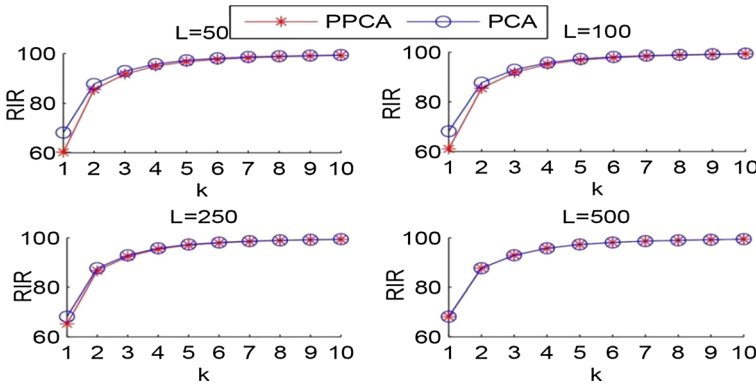


Fig. 5 Comparison of retained information between PPCA and PCA based on different reduced dimensions L

similarity of MTS. The more details can refer to the literature [40]. In this experiment, we segment MTS in the corresponding dataset according to different lengths of L . The average results can be returned when all the classifications for the chosen lengths are completed. As shown in Figs. 6 and 7, the average results in the two datasets can be compared based on different reduced dimensions k . For more detailed information, we give the classification results of PPCA and PCA based on different lengths of L and different reduced dimensions of k in the EEG dataset. As shown in Fig. 8, the results of the classification of PPCA according to different lengths of L are used to compare it to that PCA.

The comparisons in Figs. 6 and 7 show that PPCA is better than the other methods in the classification of MTS. Moreover, the classification error rate for the reduced dimensions is obviously lower than it is for the other methods. At the same time, Fig. 8 also indicates that the classification results of PPCA according to different lengths L are the best.

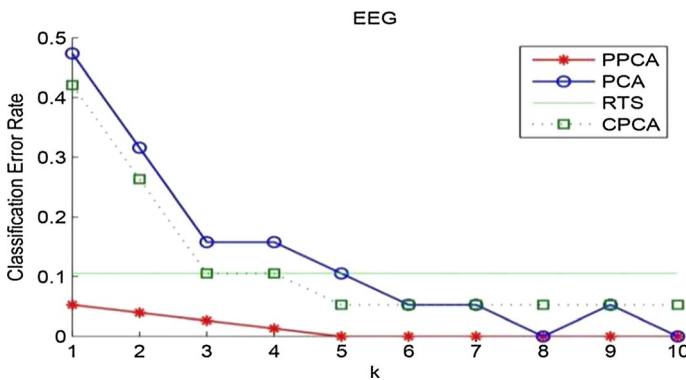


Fig. 6 The average results of the classification on the EEG dataset

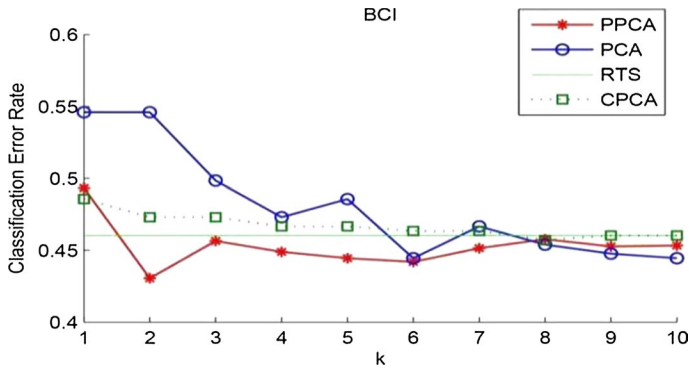


Fig. 7 The average results of the classification on BCI_Train dataset

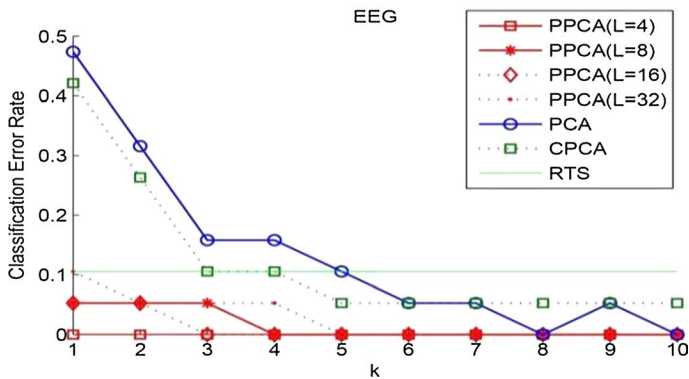


Fig. 8 The results of the classification based on different lengths L for PPCA

In addition to the experiments on these two datasets in which the lengths of MTS are equal, we also conduct an experiment on the dataset EEGEye with unequal lengths of MTS. Because CPCA and RTS are often not suitable for MTS with unequal lengths, here we only compare the classification results of PPCA to PCA, as shown in Fig. 9. The simple way to segment MTS according to the length (L) is $w = n/L$. This means that the length of the last sequence is $n - (w - 1) * L$, and the lengths of the other sequences are L . PPCA_Max, PPCA_Mean, and PPCA_Min are the maximum, the mean, and the minimum of the classification results according to different reduced dimensions (1, 2, ..., 10) for a particular length. Figure 9 demonstrates that PPCA, like PCA, is also suitable for MTS with unequal lengths.

On the whole, PPCA, like PCA, can handle MTS with random lengths. Moreover, the experimental results of classification show that PPCA outperforms the other methods, and the local information is very important for distinguishing MTS in this case. Moreover, combined with the retained information comparison, it is easy to see that the classification results are not proportional to the quantity of data information.

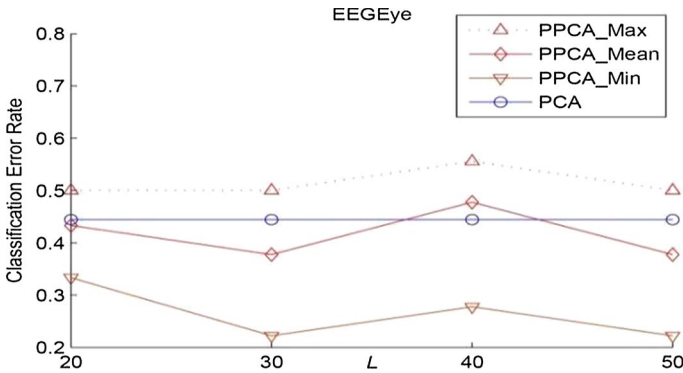


Fig. 9 The results of a classification of PPCA and PCA handling the MTS with different lengths on the EEGEye dataset

4.3 Time consumption

Time consumption is also an important factor in determining the performance of certain methods. Because CPCA and RTS often do not fit MTS with unequal lengths, in this experiment, we perform the four methods (PPCA, PCA, RTS, CPCA) on the BCI dataset with the same length of MTS using the above-mentioned classification. The CPU time consumption can be recorded based on different reduced dimensions. For PPCA, we take different segmented lengths (20,40,60,80,100) into consideration. A comparison of the four methods of CPU time consumption are shown in Figs. 10 and 11.

Figure 10 shows that the CPU time cost is lower with PPCA than with CPCA and RTS but is slightly higher than PCA. The previous analysis of PPCA indicates that PPCA and PCA have the same complexity, that is $O(nm^2) + O(m^3)$. However, the time gap is caused by some auxiliary statements in the PPCA

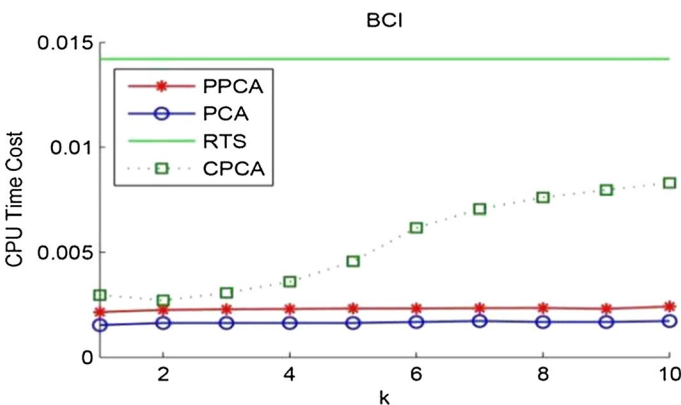


Fig. 10 A comparison of the four methods of CPU time consumption

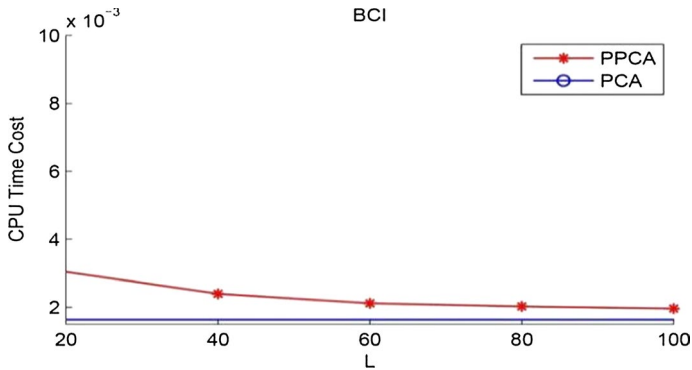


Fig. 11 The CPU time consumption of PPCA according to different segmented lengths L

procedure, such as determination of the segments and the calculation of average covariance matrix. In addition, as shown in Fig. 11, the CPU time cost of PPCA is decreased with increased length L and will be close to the cost with PCA.

5 Conclusions

This work focuses mainly on local information in MTS, which in some cases is very important for distinguishing multivariate time series. The proposed method (piecewise representation based PCA, or PPCA) segmented MTS into several sequences of equal length and calculated their covariance matrices so that more detailed relationships among the variables for MTS could be reflected. In addition, the average of all the covariance matrices was used to obtain the principal components, and the eigenvectors of the average covariance matrix were composed of the coordinates of reduced spaces. At the same time, we used the Eros distance function to measure the similarity between two coordinated systems in the experiments. The results demonstrate that PPCA can improve the quality of data mining technologies. Thus, we can conclude that PPCA takes the local information, rather than all the information into consideration, which can distinguish between two variables. Meanwhile, the average covariance matrix, to some extent, reflects the overall information in MTS, and some important characteristics are retained using PPCA. This work also reveals that the qualities of TSDM are not proportionate to the quantity of retained information by reducing the dimensionality of MTS based on PCA.

In this paper, the method used to segment MTS is equivalent division. However, this rigid segmentation sometimes makes the relationships between two variables inappropriate. Moreover, the parameter L , representing the segmented length, creates some problems in MTS data mining. Therefore, in future we should develop an adaptive and suitable method for segmenting MTS and making the feature representations more effective and robust.

Acknowledgements This work was supported by the Huaqiao University's High Level Talent Research Start-up Funding Project(14SKBS205), Ministry of Science & Technology, Taiwan (MOST 109-2511-H-003-049-MY3), Social Science Planning Project, Fujian(FJ2020B088) and National Natural Science Foundation of China (71771094).

Declarations

Conflict of interests We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence the work reported in this paper.

References

1. Davis RA, Song L (2020) Noncausal vector AR processes with application to economic time series. *J Econ* 216(1):246–267
2. Majumdar S, Laha AK (2020) Clustering and classification of time series using topological data analysis with applications to finance. *Expert Syst Appl* 162(1):113868
3. Yang D, Dong Z, Lim L et al (2017) Analyzing big time series data in solar engineering using features and pca. *Sol Energy* 153:317–328
4. Li H, Wu Y, Zhang S, Zou J (2021) Temporary rules of retail product sales time series based on the matrix profile. *J Retail Consum Serv* 60:102431
5. Yen NY, Chang JW, Liao JY et al (2020) Analysis of interpolation algorithms for the missing values in IoT time series: a case of air quality in Taiwan. *J Supercomput* 76:6475–6500
6. Müller IM (2021) Feature selection for energy system modeling: identification of relevant time series information. *Energy AI* 4:100057
7. Ahn GS, Hur S (2020) Efficient genetic algorithm for feature selection for early time series classification. *Comput & Ind Eng* 142:106345
8. Baydogan MG, Runger G (2016) Time series representation and similarity based on local auto patterns. *Data Min Knowl Discov* 30(2):476–509
9. Tamanna T, Rahman MA, Sultana A, Haque MH, Parvez MZ (2021) Predicting seizure onset based on time-frequency analysis of EEG signals. *Chaos Solitons & Fract* 145:110796
10. Sundarasekar R, Thanjaivadivel M, Manogaran G, Kumar PM, Varatharajan R, Chilankurti NK, Hsu C (2018) Internet of things with maximal overlap discrete wavelet transform for remote health monitoring of abnormal ECG signals. *J Med Syst* 42(11):1–13
11. Albertetti F, Grossrieder L, Ribaux O, Stoffel K (2016) Change points detection in crime-related time series: an on-line fuzzy approach based on a shape space representation. *Appl Soft Comput* 40:441–454
12. Gezawa AS, Bello ZA, Wang Q, Lei Y (2021) A voxelized point clouds representation for object classification and segmentation on 3D data. *J Supercomput*. <https://doi.org/10.1007/s11227-021-03899-x>
13. Papadakis SE, Kaburlasos VG (2010) Piecewise-linear approximation of non-linear models based on probabilistically/possibilistically interpreted intervals numbers (INs). *Inf Sci* 180:5060–5076
14. Si G, Zheng K, Zhou Z, Pan C, Xiang X, Kai Q, Zhang Y (2018) Three-dimensional piecewise cloud representation for time series data mining. *Neurocomputing* 316:78–94
15. Ren H, Liu M, Li Z, Pedrycz W (2017) A piecewise aggregate pattern representation approach for anomaly detection in time series. *Knowl Based Syst* 135:29–39
16. Li H (2017) Distance measure with improved lower bound for multivariate time series. *Physica A* 468:622–637
17. Fotso VSS, Nguifo EM, Vaslin P (2019) Grasp heuristic for time series compression with piecewise aggregate approximation. *RAIRO-Op Res* 53:243–259
18. Emmanuel M, Giraldez J (2019) Net electricity clustering at different temporal resolutions using a sax-base method for integrated distribution system planning. *IEEE Access* 7:123689–123697
19. He X, Shao C, Xiong Y (2016) A non-parametric symbolic approximate representation for long time series. *Pattern Anal Appl* 19(1):111–127

20. Zhang C, Chen Y, Yin A, Qin Z, Jiang Z (2018) An improvement of PAA on trend-based approximation for time series. In: Proceedings of the 18th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP'18) pp. 248–262
21. Oang KY, Yang C, Muniyappan S, Kim J, Ihee H (2017) SVD-aided pseudo principal-component analysis: a new method to speed up and improve determination of the optimum kinetic model from time-resolved data. *Struct Dyn* 4(4):044013
22. Kousika N, Premalatha K (2021) An improved privacy-preserving data mining technique using singular value decomposition with three-dimensional rotation data perturbation. *J Supercomput* 77:10003–10011
23. Granato D, Santos JS, Escher GB (2018) Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: a critical perspective. *Trends Food Sci Technol* 72:83–90
24. Yue X, Zhang H (2020) Grasshopper optimization algorithm with principal component analysis for global optimization. *J Supercomput* 76:5609–5635
25. Feng L, Zhao C, Huang B (2019) A slow independent component analysis algorithm for time series feature extraction with the concurrent consideration of high-order statistic and slowness. *J Process Control* 84:1–12
26. Krzanowski W (1979) Between-groups comparison of principal components. *J Acoust Soc Am* 74(367):703–707
27. Singhal A, Seborg DE (2005) Clustering multivariate time-series data. *J Chemom* 19:427–438
28. Karamitopoulos L, Evangelidis G, Dervos D (2010) PCA-based time series similarity search. *Data Min Ann Inf Systems* 8:255–276
29. Goetschalckx K, Moons B, Wambacq P, Verhelst M (2018) Efficiently combining SVD, pruning, clustering and retraining for enhanced neural network compression. In: Proceedings of the 2nd International Workshop on Embedded and Mobile Deep Learning pp. 1–6
30. Weng X, Shen J (2008) Classification of multivariate time series using two-dimensional singular value decomposition. *Knowl Based Syst* 21(7):535–539
31. Wu E, Yu P (2005) Independent component analysis for clustering multivariate time series data. *Adv Data Min Appl* 8:474–482
32. Issoglio E, Smith P, Voss J (2021) On the estimation of entropy in the FastICA algorithm. *J Multivar Anal* 181:104689
33. Xian L, He K, Wang C, Lai K (2020) Factor analysis of financial time series using EEMD-ICA based approach. *Sustain Futur* 2:100003
34. Xu J, Hugelier S, Zhu H, Gowen AA (2020) Deep learning for classification of time series spectral images using combined multi-temporal and spectral features. *Anal Chim Acta* 1143:9–20
35. Li H (2017) Distance measure with improved lower bound for multivariate time series. *Phys A Stat Mech Appl* 468(1):622–637
36. Huang Y, Gertler J, McAvoy T (1999) Fault isolation by partial PCA and partial NLPKA. *IFAC Proc Vol* 32(2):7647–7652
37. Barragan JF, Fontes CH, Embirucu M (2016) A wavelet-based clustering of multivariate time series using a multiscale SPCA approach. *Comput Ind Eng* 95:144–155
38. Li H (2016) Accurate and efficient classification based on common principal components analysis for multivariate time series. *Neurocomputing* 171:744–753
39. Johannesmeyer MC (1999) Abnormal situation analysis using pattern recognition techniques and historical data. PhD dissertation, University of California, Santa Barbara
40. Yang K, Shahabi C (2004) A PCA-based similarity measure for multivariate time series. In: Proceedings of the 2nd ACM International Workshop on Multimedia Databases pp. 65–74
41. Karamitopoulos L, Evangelidis G, Dervos D (2008) Multivariate time series data mining: PCA-based measures for similarity search. In: Proceedings of the 2008 International Conference in Data Mining pp. 253–259
42. Li H, Du T (2021) Multivariate time-series clustering based on component relationship networks. *Expert Syst Appl* 173:114649
43. Li H (2021) Time works well: dynamic time warping based on time weighting for time series data mining. *Inf Sci* 547:592–608
44. Kim H, Kim HK, Kim M, Park J, Cho S, Im KB, Ryu CR (2019) Representation learning for unsupervised heterogeneous multivariate time series segmentation and its application. *Comput Ind Eng* 130:272–281

45. Li H (2014) Asynchronism-based principal component analysis for time series data mining. *Expert Syst Appl* 41:2842–2850

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.