ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

A Review of Interpretable Classification of Multivariate Time Series

Chaojie Chang

School of Mathematics and Physics, China University of Geosciences (Wuhan), Wuhan 430074, China

Abstract

In recent years, as an important component of data mining, the classification task of MTS has received increasing attention from experts in various fields. Although there have been some research results on MTS classification methods, most of the studies in the literature focus on accuracy as the main research goal, and few of them focus on the interpretability of the results. In this paper, we summarize dozens of MTS classification algorithms with both accuracy and interpretability in three directions, introduce the principles and procedures of these methods, and analyze their advantages and disadvantages. It also provides an outlook on future research directions.

Keywords

Multidimensional Time Series; Multidimensional Time Series Classification; Interpretability.

1. Introduction

In most fields of science, data are measured over time and these observations lead to the collection of an ordered series of data, which we call a time series. Time series appear in many scientific and commercial applications, such as weather observations, wind energy forecasting, earthquake prediction [1] and action recognition [2]. Over time, multidimensional time series (MTS) emerge when multiple streams of interrelated data are recorded, i.e. the same thing with different parameters. For example weather observations (humidity, temperature), earth movements (3-axis) or satellite images (different spectra). In addition, through data transformation, some time-independent data such as genes, images can also be converted into the form of multidimensional time series, which are often found in medicine, finance, industry, meteorology and many other fields.

In fact, in a narrow sense, a multi-dimensional time series is a combination of multiple one-dimensional (univariate) time series, and each one-dimensional time series interacts or has certain interrelationships [3]. In a broader sense, a multivariate time series is a sequence of data generated by a particular system according to its chronological order, with each factor in the system generating a corresponding one-dimensional time series, such as geographic information systems, intelligent monitoring systems and aero-engine diagnostic systems, all of which generate large amounts of multivariate time series data. Therefore, for each classifier, dealing with such high-dimensional data is challenging in at least two ways: firstly, the MTS is characterised not only by individual feature values, but also by interactions between features of different dimensions. Second, each variable in the MTS may have a different relevance to the instance category. Among these variables, some are strongly correlated with the instance labels, while others may be contaminated by noise and have weak or no correlation. In other words, multivariate time series not only describe the patterns of variation of individual variables, but also reveal the interdependencies among the variables, so that those classification methods dealing with univariate time series cannot be transposed to the study of multivariate time series.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

In the past few decades, many algorithms for univariate time series classification (UTSC) have been developed, but multidimensional time series classification (MTSC) has received relatively little attention. Multi-dimensional time series classification is a key task to analyze and mine these multivariate data. With the increasing number of time data sets in different fields, the multivariate classification methods have attracted more and more attention from experts in different fields. In addition, one of the important research significance of data mining is to make people understand the key to practical problems, which we call interpretability. In recent years, although there have been some research results on multi-dimensional time series classification methods, most literature studies focus on accuracy as the main research goal, and few literature pay attention to the interpretability of the results. In recent years, although there have been some research achievements in multi-dimensional time series classification, most of the literatures focus on accuracy as the main research objective, and few literatures focus on the interpretability of the results. The traditional time series classification can help us to carry out effective data analysis to some extent, but the MTS classification method with speed, accuracy and interpretability will enable us to make the best response to a large number of problems in the shortest time. For example, from detecting when the patient is ill or has abnormal heart behavior, or whether the driver is in the best driving state, and the recognition of human activities, etc. Therefore, the construction of interpretable classification methods is a very hot topic. Its goal is to bulid accurate classification for complex tasks, and the constructed methods need to be easy for humans to understand. Thus, accuracy is not the only goal, and interpretability receives higher attention.

To sum up, while improving the accuracy of time series classification algorithm, it will be more useful for us to make it meet the requirements of interpretation in practical application. Based on this, this paper summarizes dozens of multi-dimensional time series classification algorithms with both accuracy and interpretability, introduces the principles and procedures of these methods, and analyzes their advantages and disadvantages.

In this paper, we introduce the classification method of multidimensional time series which is both accurate and interpretable. In Section 2, we introduce the basic definition of multidimensional time series classification. Section 3 mainly introduces the different classification methods of multidimensional time series, and divides MTS classification methods into three categories: General methods based on statistical feature extraction; Subsequence based method; Model based classification method. Section 4 introduces the further research direction of multidimensional time series representation and classification algorithm. Finally, the full text is summarized.

In this paper, we introduce the classification method of multidimensional time series which is both accurate and interpretable. In section 2, we introduce the basic definition of multidimensional time series classification. Section 3 mainly introduces different classification methods of multidimensional time series, and divides MTS classification methods into three categories: General method based on statistical feature extraction; Based on the sub-sequence method; Model-based classification methods. Section 4 introduces the further research direction of multidimensional time series representation and classification algorithm. Finally, it summarizes the whole paper.

2. Related concepts and definitions

In this section we give the relevant concepts and definitions used for multidimensional time series classification.

Definition 1 (Multidimensional Time Series): A multidimensional time series can be thought of as a vector of multiple one-dimensional time series, denoted as $T=\{T_1,T_2,\cdots T_m\}$, where m is the dimension of the multidimensional time series and the length of each one-dimensional time series T_i is l.

Multidimensional time series matrix representation: A multidimensional time series X can be represented as a matrix of $n\times m$, recorded as $X=(X_1,X_2,\cdots X_m)=(x_{ij})_{n\times m}$.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n1} & \cdots & x_{n1} \end{bmatrix} \tag{1}$$

where m denotes the variable dimension, i.e. the number of attributes, n denotes the time dimension, X_i denotes a sequence of attributes of dimension i of length n, $t = 1, 2, \cdots, n$, x_{ij} denotes the observed values of the attributes of dimension j at time i. And in general $n \gg m$, when $m = 1$, the time series is a univariate time series.

Definition 2 (Multi-dimensional Time Series Data Set): Define a data set $D = \{T_1, T_2, \cdots T_s\}$ with s multivariate time series data, each of which has the same dimension, and D_i^j denotes the j th dimension of the i th multivariate time series.

Definition 3 (Multidimensional Time Series Classification): $P = \{M_1, M_2, \cdots, M_m\}$ is a training set containing m instances, where each instance M is a multidimensional time series of length l and dimension n. Any instance in the training set uniquely corresponds to a class label in the set C of class labels. The goal of multidimensional time series classification is to learn a mapping f of multidimensional time series observations to the classes to which they belong in the training set P.

Definition 4 (Time Series Subseries): Given a time series T of length l, a new series $\{t_i, t_{i+1}, \cdots t_{i+r-1}\}$ consisting of r consecutive observations intercepted from the i th observation is called a subseries of the time series T, where $1 \le i \le l - r + 1$.

Definition 5 (Time series similarity measure): Given 2 time series T and R both of length m, the similarity measure between these two time series is denoted as $Dist(T, S)$, as shown in the following equation:

$$Dist(T, S) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (x_i - y_i)^2} \tag{2}$$

Definition 6 (shapelets): shapelets are the most discriminative continuous subseries of a time series. The K optimal shapelets are selected in a one dimensional time series classification problem and denoted as $S \in R^{K \times M}$. A shapelets $S_k = \{s_{k1}, s_{k2}, \cdots, s_{kM}\} (k \in \{1, \cdots, K\})$ of length M is a subsequence of a time series (where M is the length of the shapelets). $X_i = \{x_{i1}, x_{i2}, \cdots, s_{k1}, s_{k2}, \cdots, s_{kM}, \cdots x_{iQ}\}$ $(i \in \{1, \cdots, N\}, k \in \{1, \cdots, K\}, M < Q)$.

Definition 7 (Distance between a Time Series and a Subsequence): The distance between a complete time series and a subsequence can be calculated using $SubDist(T, S)$, where $S_T^{|S|}$ is the set of all subsequences of length $|S|$ in the sequence T.

Definition 8 (Distance between Time Series and shapelets): the distance between the i th time series X_i, and the k th shapelet (denoted as S_k) is given by:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

$$D_{i,k} = \min_{j=1,\cdots,J} \frac{1}{M} \sum_{m=1}^{M} \left(X_{i,j+m-1} - S_{k,m} \right)^2 \tag{3}$$

The formula expresses the meaning that the Euclidean distance between S_k and all $J\left(J := Q - M + 1\right)$ subsequences of length M of X_i is calculated in turn, and the smallest of these distances is chosen as the distance between X_i and S_k.

Definition 9 (shapelets transformation [6]): After finding the K optimal shapelets, the distance between the K optimal shapelets and a time series is calculated as the new feature of the time series according to Definition 7, and the set of time series samples is mapped to the new feature space, the process called shapelets transformation. The data set X is transformed into $D = \left\{ \left(D_1, y_1\right), \left(D_2, y_2\right), \cdots, \left(D_N, y_N\right) \right\}$ .Where $D_i = \left\{ D_{i1}, D_{i2}, \cdots, D_{i,k}, \cdots, D_{i,K} \right\}$. Due to $K < Q$, the shapelets transformation process reduces the dimensionality of the time series data, while generally classifying new sample sets can be done with the classifier.

Definition 10 (Time series symbolisation): Given a time series T of length l, the process of converting the sequence T or its subsequences into a string consisting of discrete symbols using time series discretization techniques is known as time series discretization, also called time series symbolisation.

3. Classification Methods for Multidimensional Time Series

This section focuses on different classification methods for multidimensional time series with both accuracy and interpretability. We classify MTS classification methods into the following three categories: general methods based on the extraction of statistical features, sub-series-based methods, and model-based classification methods.

3.1 Feature-based Classification Methods

Feature-based classifiers build models on temporal features, that is, they transform time series into feature vectors, which are then applied to traditional classification methods. Feature-based classification methods rely heavily on the features extracted from the original MTS data and are therefore more interpretable than other classifiers. As a result, feature-based methods are currently receiving increasing attention. In this section, we summarise the feature-based MTS classification methods.

Leading researcher Li et al. extended the previous work [7] in 2007 and the main difference with them is that the new feature vector extracts more information from the SVD. The improved method takes into account the singular value weighting of the first two main singular vectors. In both of Li's methods, correlations between variables can be extracted by the SVD. However, the high dimensionality of these eigenvectors leads to significant computational costs and is not interpretable.

Therefore, in order to be able to obtain classification methods with interpretability, Li et al [8] in 2016 proposed a Singular Value Decomposition (SVD) based method by which feature vectors are obtained for each matrix of MTS. The SVD reveals the geometric structure of the matrix. The first right singular vector gives the direction of greatest change in the multidimensional row vector, the second right singular vector is the direction of second greatest change, and so on. Since the first singular vector dominates and reveals the basic geometric structure of the matrix, the first singular vector is considered as the eigenvector after normalisation and some other matrix calculations. After SVD-based feature extraction, the general Support Vector Machine (SVM) classification algorithm is then applied to the feature vectors, which can be effectively used for the classification and recognition of multidimensional time series. The SVM classification algorithm has also been shown to outperform the classification of related similarity measures in terms of accuracy and CPU time, and has some interpretability through real data simulations. However, the disadvantage of this method is that the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

computational cost of the singular vector decomposition is very high. When a new time series is added to the training set, the feature vectors of the entire time series in the training set have to be recomputed.

In addition, the 2dSVD classification algorithm is an extension of the standard SVD and based on this, Weng proposes a multivariate time series classification method based on two-dimensional singular value decomposition [9] (2dSVD) since it explicitly captures the two-dimensional nature of the MTS samples. mts classification is difficult for traditional machine learning algorithms, mainly because of the different lengths of the mts samples, the many variables. In fact, the MTS sample is actually a two-dimensional data. The row-to-row and column-to-column covariance matrices of the MTS samples are extracted as feature matrices, and after obtaining the feature matrix of each MTS sample, a nearest neighbour classifier (1NN) is used for MTS classification. The performance of the algorithm depends on the selection of the number of principal feature vectors by choosing the appropriate parameters and , compared with two methods proposed by Li et al [7] for selecting feature vectors from MTS samples using the standard SVD. The experimental results indicate that the 2dSVD classification algorithm outperforms SVD on five real data sets and still has some interpretability.

Patrick [10] proposed the new MTS classifier WEASEL+MUSE. WEASEL+MUSE uses the classical method of constructing a multivariate feature vector on each vector and the basic principle of this classification method is as follows: first discrete features are extracted using a sliding window method of different sizes on each dimension. A cardinality test is then used to remove non-discriminatory features to obtain discriminatory features, and a machine learning logistic classifier is used on the final feature vector to obtain the final classification.The novelty of the WEASEL+MUSE algorithm is that it extracts and filters multivariate features from the MTS by encoding contextual information into each feature. Although the resulting feature set is small, it is identifiable and very useful for MTS classification. Experiments have shown WEASEL+MUSE to be one of the most accurate classifiers compared to the more advanced similarity metric classification methods or shapelet-based classification methods currently available. It performs well even for small datasets, whereas deep learning-based methods usually perform poorly. In addition, the method's excellent robustness and interpretability is demonstrated on data such as motion gesture recognition when studying application domains.

Most feature-based classification methods are based on univariate time series without considering multidimensional time series. However, due to the nature of some special algorithms, the classification algorithm of one-dimensional time series can be extended to multidimensional time series classification methods.Such as, Wistuba et al [11] proposed an ultra-fast (ShapeletsUltra-fast Shapelets, UFS) classification method based on redundant subsequences, which extracts Shapelets in each dimension of a multidimensional time series separately and improves the extraction speed of Shapelets to cope with the large data size of multidimensional time series.

In 2006, Weng et al [12] proposed a new feature extraction method based on supervised neighborhood preserving embedding (NPE). In fact, MTS classification is the problem of classifying a set of MTS samples into a group of predefined classes, and most of the existing MTS classification methods are not designed to preserve the intra-class local structure of the MTS data set. However, when classifiers are used to classify this class of problems [13], maintaining the intra-class local structure is crucial for the accuracy and validity of the classification results.And NPE is a new linear dimensionality reduction technique, which can maintain the local structure. This technology is not only suitable for training samples, but also suitable for testing samples.The principle is as follows: first, the MTS samples are processed by principal component analysis (PCA) to remove the smallest principal component; then, the MTS samples in the PCA subspace are projected into a low-dimensional space using NPE.The effectiveness of the proposed MTS classification method is demonstrated by comparing it with four different classifiers in simulated experiments on six real data sets. However, because the method is sensitive to the dimensionality of the embedding, the optimal value of how to find it needs to be further explored.In the simulation experiment, compared with four different classifiers, the effectiveness of the proposed MTS classification method is proved by experiments on

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

six real data sets. However, because this method is sensitive to the embedded dimension d , how to find the optimal value of d needs further exploration.

Hailin [14] proposed a multivariate time series classification method based on public principal component analysis. Firstly, the multivariate time series are divided into multiple clusters according to the number of category labels, and then the high dimension of the multivariate time series is reduced through common principal component analysis, so that the principal component series after dimension reduction has a sufficiently high variance. Secondly, each cluster is used to construct the coordinate space formed by the eigenvector of the common covariance matrix. Thirdly, any multivariate time series without category labels can be projected onto these coordinate spaces, and their labels can be predicted based on the minimum variance of the optimized principal component sequences depending on the different projections.The experimental results show that the multivariate time series classification method is more accurate and effective than the existing methods. It has certain accuracy and interpretability for multivariate time series with different lengths.

There are also feature-based MTS classification methods that combine accuracy and interpretability for MTS data in different domains, such as dealing with firefighting data and medical data, etc.

Zagorecki [15] proposed a generic method for classification of multivariate time series data based on feature engineering. 25 features were extracted from each univariate time series data and derived signals, most of them are are time domain features, while some are obtained by Fourier transform, so they belong to frequency domain features.

Zagorecki[15] proposed a general method for multivariable time series data classification based on Feature Engineering. The method extracts 25 features from each univariate time series data and the derived signals, most of them are are time domain features, while some are obtained by Fourier transform, so they belong to the frequency domain features.In order to obtain the relationship between different variables, the correlation coefficient between time series is added to the candidate feature set.Then, the cfssubsetEval algorithm is used for feature selection and then the random forest algorithm is used for classification. The effectiveness and interpretability of the method is demonstrated by the simulated experimental performance of firefighters in two mining competitions.

Xie [16] et al. proposed the Noise-Assisted Multivariate Empirical Modal Decomposition (NA-MEMD) algorithm to perform multi-scale decomposition of each variable time series of multivariate EEG signals to generate Intrinsic Mode Function (IMF) components at different frequencies.The sensitivity factor of each component is calculated using the Jason Shannon Distance (JSD) method, and the three components with the largest sensitivity values of each variable are reconstructed to generate a new signal, which is extracted using the Common Spatial Pattern (CSP), and the extracted feature values are input to a support vector machine for classification.This method first performs dimensionality reduction and feature extraction on the original time series, which makes it easier to perform classification work in low-dimensional space, but some important information may be lost when performing feature extraction, while feature extraction for multivariate time series often requires improvement of the original feature extraction method, and the parameters of the classifier are more complicated to adjust, and the selection of different classifiers will also have an impact on the classification results.

3.2 Subsequence-based Methods

Multidimensional time series data are very complex, and the often very large data sets, the high dimensionality of the time axis add significant challenges to the classification task. Compared to the traditional segmentation-based methods, the subsequence-based methods extract features from only some subsequences, while the segmentation-based methods extract features from all segments; In addition, the subsequence-based approach can further reduce the computational cost only if the extraction of subsequences is not too time-consuming.Thus, if discriminative subsequences can be extracted from the time series species, then it will make the classification results have some accuracy and interpretability.This section summarizes several general subsequence-based and shape-lets

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

subsequence-based MTS classification methods with interpretability. In addition, we further classify the MTS classification methods with interpretability based on shapelets subsequences into two cases, early classification and general classification.

Shi Mohan [17] delves into a time series classification algorithm that provides good interpretability based on shape similarity. This algorithm randomly draws discriminative subsequences from a time series, constructs a decision tree based on the distance from the instance to that subsequence, and generates a random forest with both accuracy and interpretability. Based on the existing research, a new decision tree structure is proposed for the forest, which embeds discriminative patterns from two different classes in the nodes and takes a comparison-based approach to decision making. The advantages of this approach are as follows: first, the classification accuracy is improved by introducing more information that is beneficial to the decision process. Secondly, the split threshold search process of the original structure is skipped, which reduces the training time consumption.

Lee [18] partitions a multidimensional time series into subsequences, each contained in a minimum bounding rectangle (MBR). Each MBR is stored in a database and indexed using an R-tree or its variants. The estimated MBRs are employed to speed up queries for similar actions. If two sequences are of unequal length, the shorter sequence glides from the beginning to the end of the longer sequence when the two sequences are compared. However, this makes the method unable to identify two similar sequences of different lengths or with local acceleration and deceleration.

Given that most classification methods are generally not interpretable, Ye and Keogh proposed the concept of shapelets [19] in 2009 as a new primitive representation of time series that can be highly predictive of the target. Unlike subsequences obtained by traditional segmentation methods, shapelets are subsequences obtained by supervised learning and are the most discriminative continuous subsequences of the time series.Therefore, Shapelet-based time series classification methods can provide better interpretability and at the same time can provide higher accuracy and have certain noise immunity. It should be noted that in the following, we divide the research direction of MTS classification based on Shapelet into two cases for consideration: the way of mining Shapelet itself and how to use the mined Shapelet for classification recognition.

Some researchers have now extended shaplets-based one-dimensional time series classification methods [11, 20-22] to multivariate time series classification. For example, in 2012, Ghalwash[23] proposed a method called Multivariate Shapelets Detection (MSD) based on the timing of shapelets, which extracts time series feature patterns from all dimensions of the time series and classifies the time series by searching for the earliest and closest feature patterns.They extended the definition of a local shapelet to a multivariate context consisting of multiple segments, each extracted from only one component.By evaluating on eight gene expression datasets, the experiments showed that the MSD approach outperformed the baseline approach and achieved highly accurate classification using only 40%-64% of the time series, which can be used for early and patient-specific classification of multivariate time series.

He.tal [23] proposed a method called MCFEC (Early Classification Mining Core Feature), which focuses on discovering hidden information from the data to perform early classification of multivariate time series in an interpretable way. First, different shapelets are obtained as independent core features for each variable; then, two MTS early classification methods, mcfecc classifier and mcfecc - qbc classifier based on core features, are introduced. The experimental results show that the method is more effective and efficient than the previous MSD method.

Rakthanmanon [25] proposed a fast shapelet discovery algorithm that randomly selects shapelets in each dimension of a multivariate time series to complete the classification, which solves the biggest problem in time series classification methods dealing with shapelets: finding time series is very time consuming. Experiments show that this algorithm is two to three orders of magnitude more efficient than the current state-of-the-art algorithms and effectively improves the efficiency of multivariate time series classification. However, there is no significant difference in classification accuracy compared to other methods.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

In order to be able to improve the efficiency and classification accuracy of the algorithm to some extent, Grabocka [26] et al. proposed a fast classification algorithm (called SD classification method) for multivariate time series. This method avoids the prediction accuracy of measuring similar shapelets in the Euclidean distance space by online clustering or pruning techniques. In addition, the algorithm incorporates supervised shapelet selection to filter out only those candidates that improve the classification accuracy. It is shown experimentally that the proposed method is 3-4 orders of magnitude faster than the fastest existing shapelet discovery methods, while providing better prediction accuracy. Furthermore, the effectiveness of the algorithm is demonstrated by running results on four real multivariate datasets showing that the method can classify MB-level data in a few seconds and GB-level data in a few minutes.

In the other hand, Huiyun Zhao [27] proposed a multivariate time series classification method based on shapelets learning. The principle is as follows: first, a new one-dimensional shapelets learning framework is established, and for each one-dimensional time series in the multivariate time series, a regularized least squares loss objective function is established, and the objective function is solved by an accelerated Nesterov optimization method [28] to obtain shapelets; then the one-dimensional data are feature space transformed and SVM is used for classification . This learning framework greatly improves the classification efficiency of shape-lets based one-dimensional time series. Based on this, the shape-lets-based one-dimensional time series classification method is used to classify each dimension of the multivariate time series, and the final classification result of the multivariate time series is subsequently determined by voting on the classification results of each dimension. Experiments prove that the proposed method can achieve high classification accuracy in the multivariate time series classification problem.

Another representative research direction is to transform the relationship between Shapelet and time series into new features as a representation of the original time series for classification. For example, the Shapelets Transform (ST) proposed by Hills et al [6] is one of the typical ones, which calculates the distance between a Shapelet and its closest time series fragment to obtain discrete features, and uses various classical classification methods on top of it. All made some improvements, mining the Shapelet after discretizing the original sequence, and using logistic regression methods for classification.

Although shapelet-based classifiers are more interpretable and accurate than many state-of-the-art classifiers, shapelets have the limitation that the training process for shape-let-based classification is offline and the timing of the training process may be uncontrollable.

3.3 Model-based Classification Methods

Model-based approaches are also known as change similarity-based approaches. This approach uses statistical models such as Hidden Markov models [30-32], long and short time memory, kernel function models [33-35] and other statistical models such as fitted regression [36, 37] to deal with time series classification problems.

The original Hidden Markov Model (HMM) first appeared in the 1960s and was mainly used in the fields of stock price prediction, speech recognition and signal analysis. With the development of time, more and more researchers have applied HMM to the classification of time series.

Xuan Zhang [38] proposed the Adaptive Hidden Markov Model withAnomaly States (AHMMAS), which uses the wavelet transform to learn the features for each category of stock data, and feeds the learned feature values of each category into the Hidden Markov Model (HMM), respectively. Thus, the likelihood function in the model is calculated separately, where the category with the highest value of the likelihood function is the category to which the new stock data belongs. Although the model has good performance in time series classification and has some theoretical support and strong interpretability because it is based on the traditional time series analysis theory, the model is not applicable to variable time series and time series containing a lot of noise, and requires sufficient prior knowledge to confirm the various parameters of the model and the class of the model, therefore,

the model is suitable for classification scenarios such as detection and classification, and It cannot extract time series similarity features and correlation features.

Gupta [39] proposed an early classification method for multivariate time series with some classification accuracy. He is applying HMM to multivariate time series of traffic systems to identify outdoor environments for timely classification, estimating MR for each category label using a GP classifier, and then training the HMM for that variable based on MPLs with different labels: states correspond to time series sample labels, and the set of states $S = \{S_1, S_2, \cdots S_l\}$,State Transfer Matrix $A = (a_{ij})_{i \times l}$,

$$a_{ij} = P\left(State\ S_j\ at\ t+1 \mid State\ S_i\ at\ t\right) \tag{4}$$

The sequence of states that are likely to be hidden by the input samples is obtained, followed by the classification process using the MPL idea. In addition, the estimated maximum likelihood ratio is used to construct an integrated classifier that can predict the class labels of MTDs using a class forwarding approach. The performance of the proposed method is evaluated in simulated experimental species, and the experimental results show that the method achieves good classification results on existing datasets in the fields of transportation systems, traffic congestion, and similarity analysis of driving simulations with some interpretability.

Support vector machine (SVM) [32, 40, 41] is a data mining method based on statistical learning theory proposed by Vapnik et al. Among many machine learning algorithms, support vector machines have been widely used as a machine learning algorithm with better classification effect and stability [42], and many scholars have applied SVM algorithm to the classification work of time series data [43, 44].

Chalwash et al [45] first integrated the Hidden Markov Model (HMM) and Support Vector Machine (SVM) models in 2012 and proposed the Early Classification Model (ECM) for the classification of multivariate time series. Classification Model (ECM) for classification of multivariate time series. This model is an integration of the widely used HMM and SVM models, and although it is not a new technique per se, it has not been used for early classification of multivariate time series classification so far. It achieves very good classification results on real datasets: in experiments based on published datasets of drug treatment responses in multiple patients, the ECM uses only 40% of the time series on average and is able to outperform some baseline models in terms of accuracy.

In recent years, deep learning algorithms have been applied to various fields, for example, there are many deep learning methods that have also been applied to problems about time series classification. Classification based on deep learning models is also a model-based classification method, such as temporal convolutional neural networks, residual neural networks, recurrent neural networks, etc. Compared with general neural networks, the features of temporal convolutional neural networks include the use of mean square error instead of cross entropy as the loss function, the use of local average pooling layers, and the use of convolutional computation for all dimensions together rather than independently for each dimension as in MCDCNN.

Yi Zheng et al [46] proposed the use of Multi-Channel Deep Convolutional Neural Network (MC-DCNN) for multivariate time series classification, which performs convolution and pooling operations on each channel through a long and short term memory network and a three-layer convolutional network to obtain each The time series features of the channels are extracted separately from the individual univariate time series, then the channels are stitched together to merge the features, and finally the traditional MLPs are concatenated for classification. The MC-DCNN model is evaluated on two real datasets separately, and the experimental results show that the proposed MCDCNN model outperforms the general classification methods on both datasets, especially the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

accuracy is significantly improved on the weakly labeled data set. Although the method can better handle the classification task of multivariate time series, the method performs convolutional feature extraction for each variable time series, and the more complex the constructed network model is when the variable dimension is too large.

There are many other studies on model-based time series classification methods, for example, Sebastiani et al [47] applied Markov Chain to the time series classification problem to capture the variation characteristics in the training data; Lin et al [48] proposed an innovative HMM model to abstract the signal classification problem into a time series classification problem to deal with; Deng et al [49] used ARMA model for time series classification, etc., which will not be described here.

4. Further Research Directions

Multidimensional time series have been greatly developed in the past decade or so, but there are still shortcomings in the existing studies, which provide certain directions for our future research.

(1) Combining multidimensional time series representation techniques with classification methods. A large number of representation algorithms are currently proposed for time series queries or for similarity measures, however, few representation algorithms are proposed for interpretable time series classification algorithms, so there is a need to propose a time series representation form that can be applied to interpretable time series classification algorithms.

(2) In terms of multi-dimensional time series classification algorithms. Most of the current research is devoted to the indexing and prediction of multivariate time series, and the research on classification methods for multivariate time series still has much room for development. In addition, although a large number of time series classification algorithms have been proposed in recent years, however, few of these classification algorithms, except for the shapelet-based classification algorithm with good interpretability, have been studied to meet the interpretability requirements for time series classification.

(3) Support Vector Machine (SVM) or kernel method for multidimensional time series classification. It is found that most of the SVM methods applied on time series ignore the time order in the original data and run slowly or with poor interpretability, and the kernel method based on DTW has low classification accuracy on time series data and does not have interpretability. Therefore, it is desired to have classification methods that have high classification accuracy while finding discriminative subsequences and can handle large-scale data sets. Also, how to apply kernel methods to the discovery of time-series association rules is a problem to be solved.

(4) Research on visualization of multidimensional time series data. Humans as three-dimensional animals can not perceive the high-dimensional space, but time series data precisely has the characteristics of high dimensionality, in order to better understand and analyze time series data, it is necessary to use linear or non-linear dimensionality reduction methods, such as popular learning methods and MDS, etc., to display time series data in low-dimensional space and discover the distribution between different categories of time series, so as to enable better classification processing.

(5) Classification of special multidimensional time series. For example, the classification of multidimensional time series containing missing data, unbalanced data, etc. It is found that in performing the task of unbalanced time series classification, the setting of cost-sensitive weights is often based on the existing literature, but the different weight ratios also affect the final classification effect, so how to set the weight ratios correctly and scientifically is also another research direction of unbalanced time series classification.

5. Conclusion

The paper firstly introduces the classification methods of multidimensional time series with both accuracy and interpretability from three different aspects. Due to the special nature of multi-dimensional time series data, there are still many imperfections in the existing research results, such

as how to combine multi-dimensional time series representation and classification methods, how to classify special multi-dimensional time series, and the research on the similarity measure of multi-dimensional time series, which need further efforts from researchers.

References

[1] Li W M, Li F, Liu J W, et al. Similarity Search Over Data Stream using LPC-DTW: International Conference on Machine Learning & Cybernetics, 2007[C].

[2] Lewandowski, M., Makris, et al. Structural Laplacian Eigenmaps for Modeling Sets of Multivariate Sequences[J]. Cybernetics, 2014.

[3] Hsiao C Y, Pu C, Chihying H. Learning Causal Relations in Multivariate Time Series Data[J]. Economics E Journal, 2012,1(11):1-43.

[4] Dosilovic F K, Brcic M, Hlupic N. Explainable Artificial Intelligence: A Survey: MIPRO 2018 - 41st International Convention Proceedings, 2018[C].

[5] Grabocka J, Wistuba M, Schmidt-Thieme L. Fast classification of univariate and multivariate time series through shapelet discovery[J]. Knowledge and Information Systems, 2016,49.

[6] Hills J, Lines J, Baranauskas E, et al. Classification of time series by shapelet transformation[J]. Data Mining and Knowledge Discovery, 2014,28(4):851-881.

[7] Li C, Khan L, Prabhakaran B. Feature Selection for Classification of Variable Length Multiattribute Motions, 2007[C].

[8] Li C, Khan L, Prabhakaran B. Real-time classification of variable length multi-attribute motions[J]. Knowledge and Information Systems, 2006,10(2):163-183.

[9] B X W A, A J S. Classification of multivariate time series using two-dimensional singular value decomposition[J]. Knowledge-Based Systems, 2008,21(7):535-539.

[10] Schäfer P, Leser U. Multivariate Time Series Classification with WEASEL+MUSE[J]. 2017.

[11] Wistuba M, Grabocka J, Schmidt-Thieme L. Ultra-Fast Shapelets for Time Series Classification[J]. Computer Science, 2015.

[12] Weng X. Classification of multivariate time series using supervised neighborhood preserving embedding: 2013 25th Chinese Control and Decision Conference, 2013[C].

[13] Cai D, He X, Han J, et al. Orthogonal Laplacianfaces for Face Recognition[J]. IEEE Transactions on Image Processing, 2006,15(11):3608-3614.

[14] Li, Hailin. Accurate and efficient classification based on common principal components analysis for multivariate time series[J]. Neurocomputing, 2016,171:744-753.

[15] Zagorecki A. A Versatile Approach to Classification of Multivariate Time Series Data: 2015 Federated Conference on Computer Science and Information Systems, 2015[C].

[16] ZHENG Qian, QIAO Dan, LANG Xun, et al. Classification of motor imagery signals using noise-assisted fast multivariate empirical mode decomposition [J]. Chinese Journal of Intelligent Science and Technology, 2020,2(3):240-250.

[17] Shi Mo-han. Research on time series classification algorithms for interpretable gait recognition[D]. Beijing Jiaotong University, 2019.

[18] Seok-Lyong L, Seok-Ju C, Deok-Hwan K, et al. Similarity search for multidimensional data sequences: Proceedings of 16th International Conference on Data Engineering（Cat. No.00CB37073), 2000[C].2000. 29 Feb.-3 March 2000.

[19] Ye L, Keogh E J. Time series shapelets: a new primitive for data mining: Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, 2009[C].

[20] Grabocka J, Schilling N, Wistuba M, et al. Learning time-series shapelets, 2014[C].

[21] Hou L, Kwok J T, Zurada J M. Efficient Learning of Timeseries Shapelets: Thirtieth Aaai Conference on Artificial Intelligence, 2016[C].

[22] Mueen A, Keogh E, Young N E. Logical-Shapelets: An Expressive Primitive for Time Series Classification: Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, 2011[C].

[23] Ghalwash M F, Obradovic Z. Early classification of multivariate temporal observations by extraction of interpretable shapelets[J]. Bmc Bioinformatics, 2012,13(1):195.

[24] He G, Duan Y, Peng R, et al. Early classification on multivariate time series[J]. Neurocomputing, 2015,149(pt.b):777-787.

[25] Rakthanmanon T, Keogh E. Fast Shapelets: A Scalable Algorithm for Discovering Time Series Shapelets, 2013[C].

[26] Grabocka J, Wistuba M, Schmidt-Thieme L. Fast classification of univariate and multivariate time series through shapelet discovery[J]. Knowl. Inf. Syst., 2016,49(2):429-454.

[27] Zhao Hui-yun, Pan Zhi-song. Multivariate time series classification based on shaftets learning [J]. Computer Science, 2018,45(5):6.

[28] Liu J, Yuan L, Ye J. An efficient algorithm for a class of fused lasso problems: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010, 2010[C].

[29] Fang Z, Peng W, Wei W. Efficient Learning Interpretable Shapelets for Accurate Time Series Classification: 2018 IEEE 34th International Conference on Data Engineering (ICDE), 2018[C].

[30] Bagnall A J, Janacek G J. A Run Length Transformation for Discriminating Between Auto Regressive Time Series[J]. Journal of Classification, 2014.

[31] Corduas M, Piccolo D. Time series clustering and classification by autoregressive metric. Comput Stat Data Anal 52: 1860-1872[J]. Computational Statistics & Data Analysis, 2008,52(4):1860-1872.

[32] Zhang Y, Liu B, Ji X, et al. Classification of EEG Signals Based on Autoregressive Model and Wavelet Packet Decomposition[J]. Neural Processing Letters, 2017,45(2):1-14.

[33] Smyth P. Clustering Sequences with Hidden Markov Models[J]. Advances in neural information processing systems, 1999,9:648-654.

[34] Srivastava P K, Desai D K, Nandi S, et al. HMM-ModE – Improved classification using profile hidden Markov models by optimising the discrimination threshold and modifying emission probabilities with negative training sequences[J]. BMC Bioinformatics, 2007,8.

[35] Yakhnenko O, Silvescu A, Honavar V G. Discriminatively Trained Markov Model for Sequence Classification: IEEE International Conference on Data Mining, 2005[C].

[36] Chen H, Tang F, Tino P, et al. Model-based kernel for efficient time series analysis: Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'13), 2013[C].

[37] Jeong Y S, Jayaraman R. Support vector-based algorithms with weighted dynamic time warping kernel function for time series classification[J]. Knowledge-Based Systems, 2015,75(feb.):184-191.

[38] ZHANG Xuan. Empirical research on stock price prediction based on hidden Markov model and support vector machine [D]. Shandong University, 2019.

[39] An Early Classification Approach for Multivariate Time Series of On-Vehicle Sensors in Transportation [J]. IEEE Transactions on Intelligent Transportation Systems, 2020,21(12):5316-5327.

[40] Bilal M, Hanif M S. High Performance Real-Time Pedestrian Detection Using Light Weight Features and Fast Cascaded Kernel SVM Classification[J]. Journal of Signal Processing Systems, 2019,91(2):1-13.

[41] Kafai M, Eshghi K. CROification: Accurate Kernel Classification with the Efficiency of Sparse Linear SVM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017:1.

[42] Liu Dong-qi, Chen Zhi-jian, Xu Yin, et al. Hybrid SVM algorithm oriented to classifying imbalanced datasets[J]. Application Research of Computers, 2018,35(4):5.

[43] LIU Dong, LI Su, CAO Zhi-dong. Time sequence prediction of SVM based on improvedABC algorithm [J]. Transducer and Microsystem Technologies, 2018,37(5):4.

[44] JIA Song-da, PANG Yu-song, YAN Gao-wei. Multi-task LS-SVM for application of time series prediction. Computer Engineering and Applications, 2018, 54(3):233-237.

[45] Ghalwash M F, Ramljak D, Obradovic Z. Early classification of multivariate time series using a hybrid HMM/SVM model: Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on, 2012[C].

[46] Yi, Zheng, Qi, et al. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification[J]. Frontiers of Computer Science, 2016.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202209_8(9).0019

[47] Sebastiani P, Ramoni M, Cohen P R, et al. Discovering Dynamics Using Bayesian Clustering: Advances in Intelligent Data Analysis, Third International Symposium, IDA-99, Amsterdam, The Netherlands, August 1999, Proceedings, 1999[C].

[48] Tien-Ho L, Naftali K, Ziv B J. Alignment and classification of time series gene expression in clinical studies[J]. Bioinformatics, (13):i147.

[49] Kan D, A. W M, M. C N. Learning to recognize time series: combining ARMA models with memory-based learning: Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97. 'Towards New Computational Principles for Robotics and Automation', 1997[C].1997,10-11 July 1997.