Expert Systems with Applications 42 (2015) 2305–2312

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Multivariate time series classification with parametric derivative dynamic time warping

Tomasz Górecki [a,*], Maciej Łuczak [b]

[a] Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland
[b] Department of Civil Engineering, Environmental and Geodetic Sciences, Koszalin University of Technology, Śniadeckich 2, 75-453 Koszalin, Poland

ARTICLE INFO

Article history:
Available online 13 November 2014

Keywords:
Dynamic time warping
Derivative dynamic time warping
Multivariate time series

ABSTRACT

Multivariate time series (MTS) data are widely used in a very broad range of fields, including medicine, finance, multimedia and engineering. In this paper a new approach for MTS classification, using a parametric derivative dynamic time warping distance, is proposed. Our approach combines two distances: the DTW distance between MTS and the DTW distance between derivatives of MTS. The new distance is used in classification with the nearest neighbor rule. Experimental results performed on 18 data sets demonstrate the effectiveness of the proposed approach for MTS classification.

1. Introduction

In recent decades, time series analysis has become one of the most popular branches of statistics. Time series are currently ubiquitous, and have come to be used in many fields of science. Data sets in the form of time series occur in many areas of human life. Recent developments in computing have provided the basic infrastructure for fast access to vast amounts of online data. This is especially true for the recording of time series data, for example in the medical and financial sectors. One of the major applications is time series classification. Multivariate time series (MTS) classification is an important problem in time series data mining. MTS classification is difficult for traditional machine learning algorithms mainly because of the dozens of variables (if an MTS sample is broken into univariate time series and each processed separately, the correlations among the variables could be lost) and different lengths of MTS samples.

Several approaches have been proposed for MTS classification. Maharaj (1999) used *p*-values and hierarchical clustering to classify stationary MTS. Geurts and Wehenkel (2005) classified subsequences instead of the whole MTS sample. Hayashi, Mizuhara, and Suematsu (2005) proposed an approach involving embedding MTS samples in a vector space and classifying them in the embedded space. Kadous and Sammut (2005) proposed an approach to MTS classification using metafeatures. Rodriguez, Alonso, and Maestro (2005) proposed to select literals from MTS samples with boosting and to use these literals with SVM. Yang et al. (2005) proposed a new feature subset selection method for MTS classification, based on common principal component analysis. Li, Khan, and Prabhakaran (2006) and Li, Khan, and Prabhakaran (2007) proposed two feature vector selection approaches for MTS classification by using singular value decomposition. The first approach considers only the first singular vector and the normalized singular values, while the second takes into account the first two dominating singular vectors weighted by associated singular values. Spiegel, Gaebler, Lommatzsch, De Luca, and Albayrak (2011) separated a time series into segments using SVD and then clustered the recognized segments into groups of similar context. Ghalwash, Ramljak, and Obradović (2012) integrated the Hidden Markov model and SVM classifier to make an early classification of MTS. Ghalwash and Obradović (2012) proposed using time series segments called shapelets in classification of MTS. Finally, Prieto, Alonso-González, and Rodríguez (2014) proposed stacking for multivariate time series classification.

However, these approaches do not consider explicitly the two-dimensional nature of MTS samples (an MTS sample is in fact one kind of two-dimensional matrix data). Weng and Shen (2008a) proposed a new approach for MTS classification using two-dimensional singular value decomposition, which is an extension of standard SVD. This method captures explicitly the two-dimensional nature of objects. Weng and Shen (2008b) tried also to use locality-preserving projections in the classification process of MTS. Weng (2013) presented an extension of the previous method which preserves the within-class local structure of the MTS.

* Corresponding author.
E-mail addresses: tomasz.gorecki@amu.edu.pl (T. Górecki), mluczak@wilsig.tu.koszalin.pl (M. Łuczak).

http://dx.doi.org/10.1016/j.eswa.2014.11.007
0957-4174/

This paper introduces a new shape-based similarity measure, called parametric derivative dynamic time warping DD_{DTW}, for multivariate time series data. There have been many measures proposed for univariate time series data, the most widely known being the Euclidean distance. The main problem with this measure is that the compared time series need to have the same length. A newer measure, DISSIM (Frentzos, Gratsias, & Theodoridis, 2007), provides a solution to this problem, but it is computationally costly and in general does not compare favorably with elastic measures. The family of elastic measures uses dynamic programming to align sequences with different lengths, and includes DTW (Berndt & Clifford, 1994), LCSS (Das, Gunopulos, & Mannila, 1997), edit distance with real penalty (ERP) (Chen & Ng, 2004), edit distance on real sequence (EDR) (Chen, Özsu, & Oria, 2005), derivative dynamic time warping (DDTW) (Keogh & Pazzani, 2001), and angular metric for shape similarity (AMSS) (Nakamura, Taki, Nomiya, Seki, & Uehara, 2013). The sequence weighted alignment model (Swale) (Morse & Patel, 2007) can be regarded as another elastic measure, but without employing dynamic programming. A major difference between DTW, LCSS, ERP, EDR on the one hand, and AMSS, DTW, DD_{DTW} on the other, is that those in the first group look only at individual data points, without considering the shapes of trajectories. AMSS and LCSS are less affected by outliers, but AMSS is more sensitive to short-term oscillations, which require preprocessing. There have also been other measures proposed, such as TQuEST (Aßfalg, Kriegel, Kunath, Pryakhin, & Renz, 2006) and SpADe (Chen, Nascimento, Ooi, & Tung, 2007). SpADe is similar to AMSS, DTW and our DD_{DTW} in the sense that it looks at the shapes of data. A critical difference, however, is that SpADe requires many parameters, which must be tuned for each data set, whereas AMSS and DDTW have no parameter to tune, and our DD_{DTW} has only one parameter.

The simple method combining the nearest neighbor (1NN) classifier and some form of dynamic time warping (DTW) distance has been shown to be one of the best-performing univariate time series classification techniques (Ding, Trajcevski, Scheuermann, Wang, & Keogh, 2008). The expansion of DTW to multiple dimensions is only rarely found in the literature. There exist a few works which describe extensions of the DTW algorithm to include multiple dimensions. Gavrila and Davis (1995) described a type of multivariate DTW, but used it only for the recognition of human movement. An extension of DTW into two dimensions was proposed by Vlachos, Hadjieleftheriou, Gunopulos, and Keogh (2003) and Vlachos, Hadjieleftheriou, Gunopulos, and Keogh (2006), but not systematically tested. An extension of the method of Vlachos et al. (2003) was proposed by ten Holt, Reinders, and Hendriks (2007). They also used derivatives, but calculated DTW separately on feature derivatives and on feature values, and finally added these values. In these works the term multidimensional refers to the size of the feature vectors coming from the same modality. Consequently, these approaches use the conventional two-dimensional distance matrix, whose entries are calculated from multidimensional feature vectors. Also Mello and Gondra (2008) measured the similarity between two multidimensional (but not multimodal) series. Wöllmer, Al-Hames, Eyben, Schuller, and Rigoll (2009) introduced multidimensional dynamic time warping for multimodal data streams (they assumed bimodal data streams). Finally, Banko and Abonyi (2012) proposed algorithm called correlation based dynamic time warping (CBDTW) which combines DTW and PCA for highly correlated multivariate time series.

Our previous work (Górecki & Łuczak, 2013) contains the results of research on DTW for univariate time series where the derivative is added, and where parameterization involves both function and derivative. As was shown, our method outperforms classical DTW and DDTW. Because the addition of the first derivative gave such good results in the classification of univariate time series, we decided to research further and to use our technique to classify MTS. Our approach is therefore similar to the method of Vlachos et al. (2003). In contrast to that algorithm, however, we used the parametric approach, which allows us to choose the impact of each distance on the final distance measure between the MTS, and consequently on the quality of the classification. The new distance functions so constructed are used in the nearest neighbor classification method.

The main difference between the method proposed here and other methods which use DTW and other distance measures is the use of a combined approach. We use information from regular DTW and from its derivative version DDTW in one parametric distance measure DD_{DTW}. The parametric approach means that we can choose the size of the contributions from component distance measures for different data sets. Another advantage of our algorithm is that the parameter is not located within the distance DTW (as in some works by other authors), but outside that distance. This significantly reduces the computation time. An appropriate choice of the parameter in the new method means that the error from both components of the distance on the test data sets can be made even smaller. In spite of the need to tune a parameter in the training phase, the computational complexity does not depend on the number of parameters to search. However, in the testing phase of classification the method preserves the computational complexity of the component methods (DTW or DDTW). For all these reasons in combination, our method appears to be a universal method for the classification of MTS, able to identify for which data sets the impact of the derivative is helpful, and to what extent. At the same time, the parametric approach in the new method is a disadvantage as regards computation time. An algorithm (cross-validation in this paper) is required to seek the best value of the parameter on the training data, which unfortunately increases the computation time in the learning phase. We can, however, use the standard lower bound technique to reduce the computation time for the nearest neighbor method. For the DD_{DTW} distance measure the lower bound is a combination of lower bounds for the component distances (DTW and DDTW). It is also possible to construct a special algorithm (described in this paper) which accelerates the calculations on training data in the learning phase.

The remainder of the paper is organized as follows. We first (Section 2) review the concept of MTS and the dynamic time warping algorithm for MTS data. In the same section we introduce our parametric distance based on derivatives, and explain the optimization process and properties of the new distance measure. In Section 3 the MTS data sets used in the empirical comparison of methods are described, and we explain the experimental setup. Later in that section we present the results of our experiments on the described MTS, as well as statistical analysis of the examined methods. We conclude in Section 4 with discussion of possible future extensions of the work.

2. Methods

A (one-dimensional, univariate) time series is a sequence of observations ordered in time (or space) (Box, Jenkins, & Reinsel, 2008), where time is the independent variable. For simplicity and without any loss of generality, we assume that time is discrete. Formally, a time series x is defined as a sequence of real numbers in the form:

$$x = \{x(i) \in \mathbb{R} : \ i = 1, 2, \ldots, n\}.$$

The number n of data points in a given time series is called its length.

We define a multivariate (multi-dimensional) time series X as a finite sequence of univariate time series: