Expert Systems With Applications 62 (2016) 116–130

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Hierarchical clustering of time series data with parametric derivative dynamic time warping

Maciej Łuczak*

Faculty of Civil Engineering, Environmental and Geodetic Sciences, Koszalin University of Technology, Śniadeckich 2, 75-453 Koszalin, Poland

ARTICLE INFO

Article history:
Received 10 March 2016
Revised 23 May 2016
Accepted 8 June 2016
Available online 9 June 2016

Keywords:
Time series clustering
Dynamic time warping
Parametric distance measure

ABSTRACT

Dynamic Time Warping (DTW) is a popular and efficient distance measure used in classification and clustering algorithms applied to time series data. By computing the DTW distance not on raw data but on the time series of the (first, discrete) derivative of the data, we obtain the so-called Derivative Dynamic Time Warping (DDTW) distance measure. DDTW, used alone, is usually inefficient, but there exist datasets on which DDTW gives good results, sometimes much better than DTW. To improve the performance of the two distance measures, we can combine them into a new single (parametric) distance function. The literature contains examples of the combining of DTW and DDTW in algorithms for supervised classification of time series data. In this paper, we demonstrate that combination of DTW and DDTW can also be applied in a method of time series clustering (unsupervised classification). In particular, we focus on a hierarchical clustering (with average linkage) of univariate (one-dimensional) time series data. We construct a new parametric distance function, combining DTW and DDTW, where a single real number parameter controls the contribution of each of the two measures to the total value of the combined distances. The parameter is tuned in the initial phase of the clustering algorithm. Using this technique in clustering methods requires a different approach (to address certain specific problems) than for supervised methods. In the clustering process we use three internal cluster validation measures (measures which do not use labels) and three external cluster validation measures (measures which do use clustering data labels). Internal measures are used to select an optimal value of the parameter of the algorithm, where external measures give information about the overall performance of the new method and enable comparison with other distance functions. Computational experiments are performed on a large real-world data base (UCR Time Series Classification Archive: 84 datasets) from a very broad range of fields, including medicine, finance, multimedia and engineering. The experimental results demonstrate the effectiveness of the proposed approach for hierarchical clustering of time series data. The method with the new parametric distance function outperforms DTW (and DDTW) on the data base used. The results are confirmed by graphical and statistical comparison.

1. Introduction

Clustering is a data mining technique which separates homogeneous data into uniform groups (clusters), where we do not have significant information about those groups (Rai & Singh, 2010). A special type of clustering is the clustering of time series, where a time series is an object that we identify as a (finite) sequence of real numbers (Antunes & Oliveira, 2001). Time series are classified as dynamic data since their characteristics change over time, subsequent values are linked in some way to the previous values, and their order is important. This is an integral part of the structure of time series (as opposed to non-serial data). Time series have applications in many different fields of science, engineering, economics, finance, medicine and health protection (Warrenliao, 2005). Clustering of time series can pose a significant challenge, because it requires the detection of patterns and relationships in large volumes of data. There are many problems that need to be solved when investigating time series objects, such as developing methods for detecting dynamic changes, anomaly and discord detection, process control, and character recognition (Chiş, Banerjee, & Hassanien, 2009; Faloutsos, Ranganathan, & Manolopoulos, 1994; Wang, Smith, & Hyndman, 2006).

Applications of time series clustering can be divided into several types of problems: anomaly or discord detection – finding unexpected and/or unwanted patterns in the series (Chan & Mahoney, 2005; Keogh, Lonardi, & Chiu, 2002; Leng, Lai, Tan, & Xu, 2009;

* Corresponding author.
E-mail address: mluczak@wilsig.tu.koszalin.pl

http://dx.doi.org/10.1016/j.eswa.2016.06.012
0957-4174/

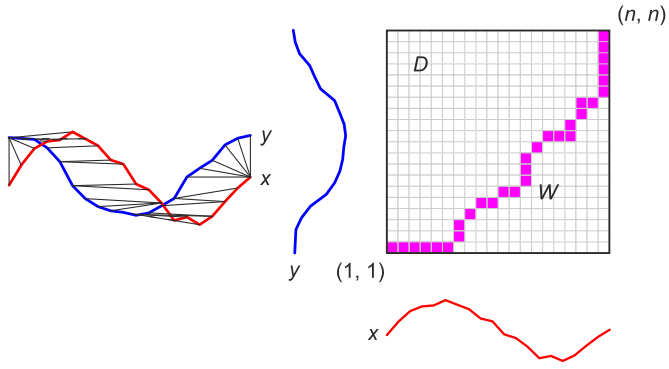

Fig. 1. Time series alignment and the corresponding warping path.

Wei, Kumar, Lolla, & Keogh, 2005); recognizing dynamic changes in the time series – correlation analysis (He et al., 2011); prediction and recommendation – various techniques of clustering and approximation enabling prediction of the future behavior of time series (Graves & Pedrycz, 2010; Pavlidis, Plagianakos, Tasoulis, & Vrahatis, 2006; Sfetsos & Siriopoulos, 2004); and pattern recognition – finding interesting patterns in databases, behavior patterns of sales or trading data (Aghabozorgi & Wah, 2014; Guan & Jiang, 2007; Kumar & Patel, 2002).

There are many different methods for the clustering of time series. In hierarchical clustering (Kaufman, Rousseeuw, & Corporation, 1990) clusters are found by an agglomerative or divisive algorithm. An agglomerative algorithm starts with each element in a single cluster, and subsequently clusters are combined into larger superclusters. By contrast, a divisive algorithm starts with one large cluster containing all of the elements, and divides them into smaller subclusters. Examples of improved methods of hierarchical clustering appear in Karypis, Han, and Kumar (1999), Guha, Rastogi, and Shim (1998), and Zhang, Ramakrishnan, and Livny (1996). In partitioning clustering, the data are divided into k groups, each containing at least one element of the dataset. The best-known algorithms of this type include k-means clustering (MacQueen, 1967), where each cluster has as its prototype the average of all the objects in the cluster. The idea behind the k-means algorithm is to minimize the average distance of elements from the cluster center (prototype). The averaged element of objects in a given cluster does not have to be an object belonging to the dataset (Petitjean et al., 2014; Petitjean & Gançarski, 2012; Petitjean, Ketterlin, & Gançarski, 2011). Another example is the k-medoids algorithm (Kaufman et al., 1990), where the prototype of a cluster is an object of the cluster which minimizes the average distance from other elements of the cluster. The k-means and k-medoids algorithms are used in many works, for example Lin, Vlachos, Keogh, and Gunopulos (2004), Bagnall and Janacek (2005), and Beringer and Hullermeier (2006). In turn, the fuzzy c-means (Bezdek, 1981) and fuzzy c-medoids (Krishnapuram, Joshi, Nasraoui, & Yi, 2001) algorithms give soft clusters, where each object can have a degree of membership in more than one cluster. Other approaches to clustering include model-based clustering (Shavlik & Dietterich, 1990), density-based clustering (Ester, Kriegel, Sander, & Xu, 1996) and grid-based clustering (Sheikholeslami, Chatterjee, & Zhang, 1998).

Time series clustering algorithms use various kinds of distance or similarity/dissimilarity measures. Each distance may reflect a different kind of similarity between time series. Shape-based measures, reflecting similarities in the time and shape domain of the time series include the Euclidean distance (ED), Dynamic Time Warping (DTW) distance (Sakoe & Chiba, 1971; 1978), Longest Common Sub-Sequence (LCSS) distance (Banerjee & Ghosh, 2001; Vlachos, Kollios, & Gunopulos, 2002) and Minimal Variance Matching (MVM) distance (Latecki et al., 2005). Distances reflecting model-based similarities include the Hidden Markov Models (HMM) distance (Smyth, 1997), and Auto-Regressive Moving Average (ARMA) distance (Kalpakis, Gada, & Puttagunta, 2001).

Information on time series clustering can be found in several review papers, for example: Aghabozorgi, Shirkhorshidi, and Wah (2015), Rani and Sikka (2012), Kavitha and Punithavalli (2010), and Liao (2005).

The Derivative Dynamic Time Warping (DDTW) distance is a measure computed as a DTW distance between (first) derivatives of the time series (Keogh & Pazzani, 2001). Pure DDTW is less useful as a universal distance measure. Used, for example, in a classification algorithm with the nearest neighbor method, it gives mostly poor results compared with DTW on a large data base of time series. However, there are a number of datasets where DDTW gives very good results. It seems best, in this case, to use a combination of the two distances. Such an approach to the process of (supervised) classification has been applied in Górecki and Łuczak (2013); 2014); 2014) (for univariate time series) and in Górecki and Łuczak (2015) (for multivariate time series). In this paper, we show that a similar technique can also be applied in the case of (one-dimensional) time series clustering. We use a distance DD_{DTW} which is a convex combination of the DTW and DDTW distances. We focus on one of the simplest methods of clustering, namely agglomerative hierarchical clustering with average linkage and a

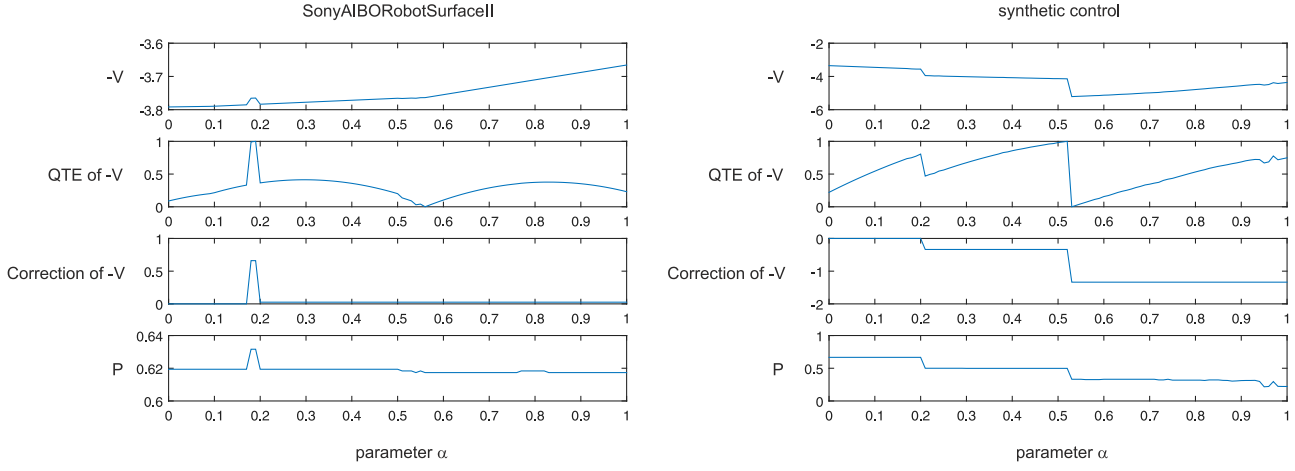

Fig. 2. The internal measure (negative) Inter-group Variance (−V), its quadratic trend extraction (QTE) and total correction with respect to the external measure Purity (P).