Information Systems 51 (2015) 27–42

Contents lists available at ScienceDirect

Information Systems

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

A distance based time series classification framework

Hüseyin Kaya [a,*], Şule Gündüz-Öğüdücü [b]

[a] Computational Science and Engineering, Istanbul Technical University, Turkey
[b] Computer Engineering, Istanbul Technical University, Turkey

ARTICLE INFO

Article history:
Received 16 September 2014
Received in revised form
3 February 2015
Accepted 15 February 2015
Recommended by F. Korn
Available online 23 February 2015

Keywords:
Time warping
Alignment
Time series classification

ABSTRACT

One of the challenging tasks in machine learning is the classification of time series. It is not very different from standard classification except that the time shifts across time series should be corrected by using a suitable alignment algorithm. In this study, we proposed a framework designed for distance based time series classification which enables users to easily apply different alignment and classification methods to different time series datasets. The framework can be extended to implement new alignment and classification algorithms. Using the framework, we implemented the k-Nearest Neighbor and Support Vector Machines classifiers as well as the alignment methods Dynamic Time Warping, Signal Alignment via Genetic Algorithm, Parametric Time Warping and Canonical Time Warping. We also evaluated the framework on UCR time series repository for which we can conclude that a suitable alignment method enhances the time series classification performance on nearly every dataset.

1. Introduction

A time series is produced by recording successive measurements of some quantity over time. There are many application fields working with time series [1–3]. For instance, in the search for exoplanets a time series is created by periodically recording the brightness of a target star, which is called a star light. The time series is then classified with respect to a set of references in order to find unexpected dips due to a planet passing directly between the star and the observer. More than a thousand potential exoplanet has been discovered in the Kepler project by time series analysis [4].

Time series classification plays an important role in many applications such as signature verification [5], speech and handwriting recognition or change detection in mutation analysis. It has been a topic of great interest which has two fundamental components: (i) classification and (ii) alignment. In the classification component, the well known methods, such as the k-Nearest Neighbor (k-NN) and Support Vector Machines (SVM) [6], can be used in their original forms. In these methods, the distance between a pair of time series is calculated by using standard Euclidean metric because of its easy and fast implementation. A time series is created by recording measurements over time. The imperfections of measurement device or differences in the examined subjects result in distortion in the time axis called time drifts or retention time difference [7] which reduces the classification performance. In order to eliminate the distortions, one should make non-linear adjustments often called alignment [8].

Alignment involves the elimination of temporal variations by stretching or compressing the time axis of one or both time series [9]. Another equivalent definition of alignment is to create a mapping between a pair of time series by fitting a warping function. Although the alignment methods may follow very different strategies, they all aim to produce a mapping which is then used to correct the time drifts in

* Corresponding author. Tel.: +90 5556218176.
E-mail address: hkayabilisim@gmail.com (H. Kaya).
URL: http://timewarping.org (H. Kaya).

http://dx.doi.org/10.1016/j.is.2015.02.005
0306-4379/

the time series. The newly produced and time corrected time series are called aligned time series.

Alignment methods can be used to improve the performance of a classifier by integrating the alignment method into the distance calculation. In this setup, whenever a distance calculation between a pair of time series is requested by the classifier, the time series are first aligned, then the Euclidean distance of the aligned time series is returned to the classifier. By using such an approach, the alignment becomes an integral part of the distance calculation so that it is usually considered as a distance measure [10].

The methods proposed in the studies on alignment techniques are usually considered as new alternatives to the standard Euclidean distance metric or other distance measures based on different alignment methods. For instance, Dynamic Time Warping (DTW) [11], one of the first alignment techniques emerging from the spoken word recognition field, is perceived as a new distance measure. DTW is indeed a generalization of the Minkowski distance which can handle time series of different lengths [12]. However, many other distance measures, such as the Euclidean distance, cannot be applied to time series of different lengths. In such cases, the time series first need to be "aligned" by re-interpolating them to equal length. This implies that a distance measure does not always behave as an alignment technique but rather works in tandem with alignment methods. Therefore, in this study, it is preferred to distinguish alignment methods from distance measures even if the alignment is a variant of some distance measure.

Another challenge in time series classification is the application specific adjustments either in alignment or data processing steps. For instance, multi-dimensional time series are often converted to one dimension by summing, averaging or using other dimensional reduction algorithms because of the fact that the majority of the alignment algorithms are designed to work only with one-dimensional time series [5]. Amplitude normalization, baseline correction or measuring the quality of the alignment are the other minor tasks to be handled in times series classification [13,14]. As a result, many alignment techniques proposed in the literature are entangled with the application specific tunings which make them unusable for standalone alignment purposes. For the same reason, the performance of the proposed alignment methods can be evaluated on a very limited number of application domains. The studies about time series classification have also a reasonable tendency to focus only on the field of study. Therefore, there are only a few studies trying to test their methods on datasets from different application domains.

In order to overcome the difficulties highlighted above, a public time series classification framework is proposed to make a clear distinction between classification and alignment.[1] The proposed framework enables one to freely change the alignment method and observe its performance without dealing with classification. Likewise, different classification techniques can be tested by keeping alignment method fixed. New classification or alignment algorithms can also be integrated to the framework by implementing the related interfaces. The framework also benefits parallel computing resources if available.

By using the proposed framework, two classification and four alignment methods were tested on 40 different datasets kindly provided by Eamonn Keogh of University of California, Riverside [15]. The most significant outcome of the experiments is that using an alignment method dramatically improves the classification performance on nearly every dataset. The second finding is that the performance of alignment techniques heavily relies on the characteristics of investigated dataset as such an example is analyzed in the experiments. Parallel programming feature of the framework was also tested by utilizing the facilities in High Performance Computing Center of Turkey. The framework has been designed as an open source project, so that researchers can implement their own algorithms.

The rest of this paper is organized as follows: In Section 2 we surveyed the literature on time series classification algorithms. In Section 3, we presented the framework and its current classification and alignment implementations. In Section 4, we experimented the framework with the dataset. The experimental setup and the results were also given. In the last section, we gave a conclusion and future work.

2. Related work

The studies in time series classification can be analyzed in three main categories with respect to the classification scheme: feature based, model based and distance based [16]. Our classification framework falls into the third category.

2.1. Feature based classification

It is a widely used strategy to reduce the dimensionality of time series by transforming into a set of features with the help of a suitable feature extraction method because directly using a high dimensional time series in its raw format is computationally expensive. In this way, the time series that cannot be directly used in the classification are converted into a smaller and more suitable forms. There are dozens of different types of feature extraction methods some of which are summarized here.

The spectral methods such as Discrete Fourier Transform [17] and Discrete Wavelet Transform (DWT) [18] transform the time domain into a frequency domain in which only a few low-frequency harmonics are retained for representation because of the fact that in many practical signals the most of the "energy" is concentrated on first low-frequency harmonics.

Eigenvalue methods are most commonly used feature extraction methods based on finding optimal set of features by applying eigenvalue analysis techniques such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) to a set of time series [19]. Since the features are determined in optimal sense, the performance of eigenvalue based methods is usually better than spectral methods especially if the input time series have several spikes of abrupt jumps [19]. Although PCA and SVD give optimal set of features, an extensive survey claims that they become unfeasible for out-of-core (disk

[1] The framework is available at http://timewarping.org.