Neurocomputing 167 (2015) 8–17

Contents lists available at ScienceDirect

Neurocomputing

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

A multi-scale smoothing kernel for measuring time-series similarity

A. Troncoso [a,*], M. Arias [b], J.C. Riquelme [c]

[a] Department of Computer Engineering, Pablo de Olavide University, Spain
[b] Department of Computer Science, Universitat Politècnica de Catalunya, Spain
[c] Department of Computer Science, University of Seville, Spain

ARTICLE INFO

Article history:
Received 30 October 2013
Received in revised form
18 March 2014
Accepted 11 August 2014
Available online 19 May 2015

Keywords:
Kernel
Similarity
Distance
Time-series classification

ABSTRACT

In this paper a kernel for time-series data is introduced so that it can be used for any data mining task that relies on a similarity or distance metric. The main idea of our kernel is that it should recognize as highly similar time-series that are essentially the same but may be slightly perturbed from each other: for example, if one series is shifted with respect to the other or if it slightly misaligned. Namely, our kernel tries to focus on the shape of the time-series and ignores small perturbations such as misalignments or shifts. First, a recursive formulation of the kernel directly based on its definition is proposed. Then it is shown how to efficiently compute the kernel using an equivalent matrix-based formulation. To validate the proposed kernel three experiments have been carried out. As an initial step, several synthetic datasets have been generated from UCR time-series repository and the KDD challenge of 2007 with the purpose of validating the kernel-derived distance over shifted time-series. Also, the kernel has been applied to the original UCR time-series to analyze its potential in time-series classification in conjunction with Support Vector Machines. Finally, two real-world applications related to ozone concentration in atmosphere and electricity demand have been considered.

1. Introduction

Time-series analysis is an important problem with application in domains as diverse as engineering, medicine, astronomy or finance [11,29]. In particular, the problem of time-series classification and prediction is attracting a lot of attention among researchers. One of the most successful and popular methods for classification and prediction are kernel-based methods such as support vector machines (SVM) [26,12,35,25]. Despite their popularity, there seem to be only a handful of kernels designed for time-series. This paper tries to fill this gap, and proposes a kernel exclusively designed for time-series. Moreover, using a standard trick, we are able to convert our kernel into a distance for time-series, therefore allowing us to use our kernel in distance-based algorithms as well.

A crucial aspect when dealing with time-series is to find a good measure, either a kernel similarity or a distance, that captures the essence of the time-series according to the domain of application.

For example, Euclidean distance between time-series is commonly used due to its computational efficiency; however, it is very brittle and small shifts in one time-series can result in huge changes in the Euclidean distance. Therefore, more sophisticated distances have been devised and designed to be more robust to small fluctuations of the input time-series. Notably, Dynamic Time Warping (DTW) [30] is held as the state-of-the-art method for comparing the similarity among time-series. The DTW is very powerful in the sense that it can deal optimally with contractions, expansions and shifts in time-series in addition to being able to handle time-series of different lengths. Unfortunately, computing the DTW distance is prohibitively costly for many practical applications [33]. Moreover, it cannot be used to define a positive definite kernel since it violates the triangle inequality [3].

Therefore, researchers are coming up with distances for time-series that approximate the DTW at lower computational costs either by adding global path constraints [30,36] or by reducing the number of instances e.g. in nearest neighbor classification [35].

In this paper we introduce a new kernel, called MUlti-Scale Smoothing Kernel (MUSS). The basic idea behind our kernel is to take into account many smoothed versions of the time-series and compute the similarity of the time-series as the aggregation of the similarities of the multiple smoothed versions of the original time-series. The underlying idea is that by smoothing the original time-series we will get rid of slight perturbations, and so the basic trends will become apparent and more easily detected. The main strength of this kernel is the integration of multiple time-scales, that is, at a high level, the MUSS kernel is a combination of linear kernels obtained by

* Corresponding author.
E-mail addresses: ali@upo.es (A. Troncoso), marias@lsi.upc.edu (M. Arias), riquelme@us.es (J.C. Riquelme).

http://dx.doi.org/10.1016/j.neucom.2014.08.099
0925-2312/

using several smoothed versions over different scales from the original time-series. In a sense, the kernel-derived distance that is proposed here tries to fix the brittleness of Euclidean distance without incurring in the high computational costs of DTW. Moreover, our kernel can easily be adapted to deal with multidimensional time series by considering multi-variate versions of the point-wise distance between time-series. In addition, we can derive a distance metric from the kernel definition that satisfies the triangle inequality.

The main goal of the proposed kernel is to recognize as similar time-series that may be slightly perturbed from one another. Namely, it tries to focus on the shape of the time-series and not so much on the details. It is conceivable that small errors in measurement or delivery of data may result in slight shifts or misalignments of time-series. Consequently, any data that is sent through complicated machinery can suffer from this type of misalignment as for example astronomic data, and could benefit from our kernel.

In this work, two ways of computing the kernel are presented: a recursive formulation and an equivalent matrix-based formulation. To evaluate the proposed kernel three experiments have been carried out. As an initial step, several synthetic datasets have been generated from UCR time-series repository [20] and the KDD challenge of 2007 [19], with the purpose of validating our kernel-derived distance over shifted time-series. In particular, a comparison with DTW and Euclidean distances shows that our kernel-derived distance outperforms the Euclidean distance and is competitive with respect to the DTW distance while having a much lower computational cost. The DTW distance is designed to deal with misalignments and shifts optimally. Therefore, our objective is not to beat the DTW, but to approach its performance without incurring its high computational cost. On the other hand, the Euclidean distance has been considered as baseline distance. In the second experiment, the proposed kernel has been applied to the original UCR time-series [20] to analyze its potential in time-series classification using an SVM. In this case, the proposed kernel shows a remarkable performance when comparing with a kernel based on DTW [10] and a linear kernel. Finally, two real-world applications related to ozone concentration in atmosphere and electricity demand have been considered to show the performance of the MUSS kernel over very specific datasets. In this case, an accuracy ranging from 97% to 99% has been obtained.

The paper is structured as follows. Section 2 presents the most relevant related works found in literature. Section 3 describes our time-series kernel and its corresponding derived distance. The experimental results are presented in Sections 4–6. Finally, Section 7 concludes with a summary of our main contributions and possible directions for future work.

2. Related work

Similarity and distance measures for time-series are a crucial ingredient in solving time-series classification and forecasting problems [29,15]. For this reason, many distances have been proposed. For example, [1] defines a distance between two time-series representing the convexities/concavities of two shape contours. In [4] the authors modify the Euclidean distance with a correction factor based on the complexity of the input time-series.

The success and popularity of Support Vector Machines has motivated researchers to design kernels that capture similarity between time-series and sequences. For example, [32] define a kernel for the particular task of handwritten character recognition. In this work, the authors approximate each time-series by a linear combination of piecewise polynomial functions and the kernel is based on the product of the coefficients and functions that form part of this approximation. Another family of kernels for time-series based on Echo State Network [23,27] with a deterministic reservoir architecture is proposed in [6]. Their kernel is defined by Gaussian kernel with the L_2 distance between the corresponding readouts for each time-series from the same reservoir.

It is well-known that the DTW distance is not a distance in a strict sense as it does not satisfy the triangle inequality and, therefore, it cannot be used to define a positive definite kernel [3]. Despite this disadvantage, many variants of DTW and definitions of kernels based on DTW have been recently proposed in the literature. As an example, [17] use Gaussian kernel and the DTW distance with a special support vector machine, which has the ability to handle non positive-definite kernel matrices.

Another example of the use of (a weighted variant of) DTW for time-series classification, this time based on nearest neighbors is [18]. The weights penalize instances with higher phase difference between a reference point and a testing point with the purposes of minimizing the distortion caused by outliers.

More recently, non-linear kernels have been proposed for time-series classification. In [10] a new kernel based on the DTW distance is defined by global alignments (GA-DTW). In particular, the kernel is defined as the sum of the exponential function of the distances for all possible alignments. However, this kernel has a high computational cost and similar constraints on alignments to that of [30] are presented to speed-up the computation in [8]. The same author presents another kernel based on the idea that similar time-series should be fit well by the same models [9]. The author used autoregressive models and thus the name of autoregressive kernel. In particular, these two global alignment-based and autoregressive kernels defined in [8] and [9] have been recently used in machine olfaction applications in conjunction with SVM [33]. An extension of SVM based on nonlinear dynamical systems theory is presented in [16]; here it is shown that these non-linear methods perform better and faster than the DTW distance-based methods. However, linear kernels may still be preferred over their more accurate non-linear counterparts due to their interpretability, computational efficiency and the lack of metaparameters that need tuning.

A kernel for periodic time-series arising in the field of astronomy is presented in [34]. This kernel is similar to a global alignment kernel as it consists of the sum of the exponential function of the inner products for all possible shifts of a time series instead of alignments.

Finally, another kernel for time-series is proposed in [22]. In particular, the time series are represented with a summarizing smooth curve in a Hilbert space and the learning method of the kernel is based on Gaussian processes.

In the lasts years, several approaches have been proposed to combine multiple kernels instead of using a single kernel. In [28] a combination of kernels for long-term time-series forecasting is presented. In particular, a kernel that takes into account the seasonality of the time-series to improve the performance of the predictor is combined with the well-known Gaussian or rational quadratic kernel. A detailed description can be found in [14].

Due to the fact that annotation of class labels in time-series is very expensive, researchers are exploring the semi-supervised methodology to the problem of time-series classification. The main strategies within this line of work are the extension of well-known semi-supervised techniques for static data classification to time-series problems [21], and the definition of new distances for time-series that work well in semi-supervised classification [7].

3. Kernel description

This section presents the notation used in this paper and also provides the definitions underlying the proposed kernel.