Neurocomputing 169 (2015) 196–204

Contents lists available at ScienceDirect

Neurocomputing

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

Theoretical properties and implementation of the one-sided mean kernel for time series

Nicolas Chrysanthos *, Pierre Beauseroy, Hichem Snoussi, Edith Grall-Maës

Institut Charles Delaunay (UMR CNRS 6279 STMR), LM2S Université de Technologie de Troyes, 10010 Troyes, France

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form
23 November 2014
Accepted 24 November 2014
Available online 4 April 2015

Keywords:
Kernel methods
Structured data
Sequential data
Dynamic programming

ABSTRACT

In this paper we introduce a new kernel for sequences of structured data, investigate its properties and propose a fast implementation. We demonstrate using the theory of infinitely divisible kernels that this kernel is positive definite, that it is a radial basis kernel and that it reduces to a product kernel when comparing two sequences of the same length. We present an implementation of this kernel using dynamic programming techniques that leads to an algorithm of lower complexity than competing kernels. We illustrate that this kernel presents a consistent behavior in the context of sub-sampling of continuous time series. Finally we compare this kernel with the global alignment kernel in two classification tasks with real world data using support vector machines.

1. Introduction

Kernel methods have proven extremely useful for dealing with a wide range of problems [1], in particular they have been used to extract nonlinear features using extensions of linear methods [2] and to treat structured data as vectors in a Hilbert space [3]. In this publication we deal with the case of data organized as sequences, whose elements lie in an arbitrary space $\mathcal{X}$, which can be a space of structured data. We only assume $\mathcal{X}$ is endowed with a kernel k. We denote by $\mathcal{X}^*$ the space of finite sequences with elements in $\mathcal{X}$, such that $\mathcal{X}^* = \bigcup_{i=1}^{\infty} \mathcal{X}^i$. In the literature $\mathcal{X}$ and k are called respectively the *ground space* and the *ground kernel* whereas $\mathcal{X}^*$ is sometimes called the sequence space. The goal is now to design a sequence kernel k^* on $\mathcal{X}^*$ with suitable properties.

Framework: Let $\boldsymbol{x} = (x_1, \ldots, x_l)$ and $\boldsymbol{x}' = (x_1', \ldots, x_m')$ be two elements of $\mathcal{X}^*$. In the general case, these two elements may not have the same length, and thus one cannot use traditional vector-based approaches such as a Gaussian kernel in an Euclidean space to compare these sequences.

One solution is to define alignments between sequences. An alignment associates elements from one sequence to elements in another sequence such that the order of elements is preserved.

The well-known dynamic time warping method [4] for example can be defined using alignments. This algorithm seeks the best alignment between two sequences, and results in the so-called optimal assignment kernel. Although this kernel has been extensively used by practitioners it has been demonstrated recently that it is in fact not positive definite [5]. Since then some researchers have proposed alternatives, such as the global alignment kernel [6] or the spectrum kernel [7].

Methodology: The kernel studied in this work is novel in two regards. Firstly, we only consider a particular kind of alignments with repetitions, those in which *only the shorter sequence can have repeated elements*, hence the name "one-sided". Secondly, instead of only retaining the best alignment like in the optimal assignment we rather consider the mean (in a sense which shall be clarified) of all alignment scores.

Contributions: We demonstrate using the theory of infinitely divisible kernels that the proposed kernel is positive definite. We also illustrate many other interesting practical properties: it is a radial basis kernel, has no issues of diagonal dominance, and presents a consistent behavior in the case of time series sub-sampling. We propose an implementation of this kernel using dynamic programming techniques, which results in a complexity of $O(l \times (m-l))$ for a pair of sequences of respective lengths $l < m$, which is much faster than competing techniques which have a complexity in $O(l \times m)$.

2. Alignments and alignment scores

Let $\boldsymbol{x} = (x_1, \ldots, x_l)$ and $\boldsymbol{x}' = (x_1', \ldots, x_m')$ be two elements of $\mathcal{X}^*$. An alignment associates elements from one sequence to elements in another sequence such that the order of elements is preserved.

* Corresponding author.
E-mail addresses: nicolas.chrysanthos@gmail.com (N. Chrysanthos), pierre.beauseroy@utt.fr (P. Beauseroy), hichem.snoussi@utt.fr (H. Snoussi), edith.grall@utt.fr (E. Grall-Maës).

http://dx.doi.org/10.1016/j.neucom.2014.11.079
0925-2312/

Alignments can either introduce gaps or repetitions. Once an alignment has been found, it is then possible to compute an alignment score between the two sequences, by for example summing the pairwise distances between aligned elements.

2.1. Global alignments

In this work we will be interested in alignments that introduce repeating states. Formally we define an alignment π of length p between two sequences of lengths l and m as a pair (π_1, π_2) of p increasing indexes such that

$$1 = \pi_1(1) \leq \cdots \leq \pi_1(p) = l$$
$$1 = \pi_2(1) \leq \cdots \leq \pi_2(p) = m \quad (1)$$

and

$$\begin{pmatrix} \pi_1(i+1) - \pi_1(i) \\ \pi_2(i+1) - \pi_2(i) \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \quad (2)$$

We denote by $\mathcal{A}(\boldsymbol{x}, \boldsymbol{x}')$ the set of all alignments between $\boldsymbol{x}$ and $\boldsymbol{x}'$. The length p of an alignment π is denoted by $|\pi|$.

2.2. One-sided alignments

In this paper we will be interested in a particular subset of alignments that we have called one-sided alignments. These are the alignments *where only the shortest sequence can have repeated elements*. Suppose sequence $\boldsymbol{x}$ is shorter than $\boldsymbol{x}'$, such that $l \leq m$; thus the condition on the alignment π becomes

$$\begin{pmatrix} \pi_1(i+1) - \pi_1(i) \\ \pi_2(i+1) - \pi_2(i) \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \quad (3)$$

One should remark that between two sequences of the same length, there exists only one one-sided alignment, which is the trivial alignment $\forall i \in 1 \ldots l, \pi_1(i) = \pi_2(i) = i$. In the general case where $l \leq m$ it can be seen from Eq. (3) that there are $\binom{m-1}{l-1}$ one-sided alignments between $\boldsymbol{x}$ and $\boldsymbol{x}'$. We denote by $\mathcal{A}^-(\boldsymbol{x}, \boldsymbol{x}')$ the set of one-sided alignments between $\boldsymbol{x}$ and $\boldsymbol{x}'$.

2.3. Representation of alignments

It is possible to conveniently represent alignments between two sequences of lengths l and m as paths on a matrix of size (l,m). Note that in this section as well as in the rest of this work we shall always represent this matrix with the *shorter sequence as the vertical indexes*. At each step, the vertical position is given by π_1 which corresponds to $\boldsymbol{x}$ and the horizontal position is given by π_2 which corresponds to $\boldsymbol{x}'$. Note that Eqs. (1)–(3) impose restrictions on the paths on the matrix. Fig. 1 presents an example of both a global and a one-sided alignment between sequences of lengths 3 and 6. One can see in Fig. 1 that because of Eq. (3) and the fact that each path must start and end at respectively the left-hand top corner and the right-hand bottom corner (Eq. (1)) some parts of the alignment matrix in the one-sided case cannot be attained and are displayed with red stripes.

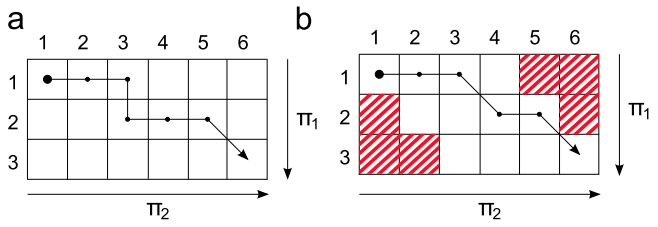

Fig. 1. Global and one-sided alignments. The parts of the matrix with red stripes in the one-sided case cannot be attained due to Eq. (3): (a) global alignment and (b) one-sided alignment. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

2.4. Examples of kernels defined with alignments

2.4.1. Optimal assignment kernel

Based on the popular dynamic time warping technique [4], the optimal assignment kernel considers only the "best" alignment between two sequences. When dealing with continuous values such that $\mathcal{X} = \mathbb{R}^d$ the best alignment is the one that maximizes the similarity:

$$k_{\mathrm{DTW}}(\boldsymbol{x}, \boldsymbol{x}') = \max_{\pi \in \mathcal{A}(\boldsymbol{x}, \boldsymbol{x}')} \prod_{i=1}^{|\pi|} k(x_{\pi_1(i)}, x'_{\pi_2(i)})$$

Although widely used in the literature, Vert [5] demonstrated that this kernel is in fact not positive definite and thus cannot be used as is in kernel methods.

2.4.2. Global alignment kernel

To circumvent this issue, Cuturi et al. [6] have proposed an alternative kernel, named the global alignment kernel. Contrary to the optimal assignment kernel, this one does not only consider the best alignment but rather sums the scores associated with all possible global alignments. As a consequence this kernel may prove more robust to quantify the similarity between two sequences, and they demonstrated that it was indeed positive definite under mild conditions:

$$k_{\mathrm{GAS}}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{\pi \in \mathcal{A}(\boldsymbol{x}, \boldsymbol{x}')} \prod_{i=1}^{|\pi|} k(x_{\pi_1(i)}, x'_{\pi_2(i)}) \quad (4)$$

3. The one-sided mean alignment kernel

The fact that we use only one-sided alignments for the definition of this kernel will let us introduce another formalism that will not only simplify the notations but also the demonstration of the main theorem of this contribution. The formalism we introduce is that of *dilatation operators*.

A dilatation operator is a function that maps a sequence to a longer sequence by repeating one or more of its elements while still keeping the order. We denote by $\xi_{l \to m}$ the set of dilatation operators that map sequences of length l to sequences of length m. The dilatation operators will be properly defined in later sections, but in the meantime one only has to know that there is a one-to-one correspondence between the set of one-sided alignments $\mathcal{A}^-(\boldsymbol{x}, \boldsymbol{x}')$ where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are two sequences of respective lengths $l \leq m$ and the set of dilatation operators $\xi_{l \to m}$. Note that consequently the cardinality of this set verifies $|\xi_{l \to m}| = \binom{m-1}{l-1}$.

Before delving into the technical details we provide in Table 1 a set of notations that may be helpful for reading the sequel.

3.1. Practical case: real values with Gaussian kernel

We start by giving an example of the one-sided mean kernel in the simple case where elements of sequences are real values: $\mathcal{X} = \mathbb{R}$, and the ground kernel k is the one-dimension Gaussian kernel $k(x, x') = \exp(-(x - x')^2)$ where $x, x' \in \mathbb{R}$. This is useful to get a sense of how this kernel is represented in most practical cases, before we delve into the more abstract setting of infinitely divisible kernels. As described in Table 1, let $\boldsymbol{x}$ and $\boldsymbol{x}'$ be two elements of $\mathcal{X}^*$. Furthermore we refer to the shorter and longer elements of $(\boldsymbol{x}, \boldsymbol{x}')$ as $\boldsymbol{x}_l$ and $\boldsymbol{x}_m$ respectively, with $l \leq m$ denoting the respective lengths of the sequences. Note that $\boldsymbol{x}_l$ and $\boldsymbol{x}_m$ are not to be mistaken for x_l and x_m which refer respectively to the lth and mth elements of $\boldsymbol{x}$. In the real case the one-sided kernel k^* is