Pattern Recognition 71 (2017) 187–195

Contents lists available at ScienceDirect

Pattern Recognition

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

Indefinite Core Vector Machine

Frank-Michael Schleif[a,b,*], Peter Tino[a]

[a] School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK
[b] University of Applied Sciences Wuerzburg Schweinfurt, Sanderheinrichsleitenweg 20, 97074 Wuerzburg, Germany

ARTICLE INFO

Article history:
Received 27 December 2016
Revised 6 April 2017
Accepted 1 June 2017
Available online 3 June 2017

Keywords:
Indefinite learning
Krĕin space
Classification
Core Vector Machine
Nyström
Sparse
Linear complexity

ABSTRACT

The recently proposed Krĕin space Support Vector Machine (KSVM) is an efficient classifier for indefinite learning problems, but with quadratic to cubic complexity and a non-sparse decision function. In this paper a Krĕin space Core Vector Machine (iCVM) solver is derived. A sparse model with linear runtime complexity can be obtained under a low rank assumption. The obtained iCVM models can be applied to indefinite kernels without additional preprocessing. Using iCVM one can solve CVM with usually troublesome kernels having large negative eigenvalues or large numbers of negative eigenvalues. Experiments show that our algorithm is similar efficient as the Krĕin space Support Vector Machine but with substantially lower costs, such that also large scale problems can be processed.

1. Introduction

Learning of classification models for indefinite kernels received substantial interest with the advent of domain specific similarity measures. Indefinite kernels are a severe problem for most kernel based learning algorithms because classical mathematical assumptions such as positive definiteness, used in the underlying optimization frameworks are violated. As a consequence e.g. the classical Support Vector Machine (SVM) [1] has no longer a convex solution - in fact, most standard solvers will not even converge for this problem [2]. Researchers in the field of e.g. psychology [3], vision [4–6] and machine learning [7,8] have criticized the typical restriction to metric similarity measures. In fact in [8] for multiple examples from real problems it is shown that many real life problems are better addressed by e.g. kernel functions which are not restricted to be based on a metric. Non-metric measures (leading to kernels which are not positive semi-definite (non-psd)) are common in many disciplines. The use of divergence measures [9–11] is very popular for spectral data analysis in chemistry, geo- and medical sciences [12,13], and are in general not metric. Also the popular Dynamic Time Warping (DTW) [14] algorithm provides a non-metric alignment score which is often used as a proximity measure between two one-dimensional functions of different length. In image processing and shape retrieval indefinite proximities are often obtained by means of the inner distance [15] - another non-metric measure. Further examples can be found in physics, where problems of the special relativity theory naturally lead to indefinite spaces. Further prominent examples for genuine non-metric proximity measures can be found in the field of bioinformatics where classical sequence alignment algorithms (e.g. smith-waterman score [16]) produce non-metric proximity values. Multiple authors argue that the non-metric part of the data contains valuable information and should not be removed [6,7].

Furthermore, it has been shown [2,7,17] that work-arounds such as eigenspectrum modifications are often inappropriate or undesirable, due to a loss of information and problems with the out-of sample extension.

Due to its strong theoretical foundations, Support Vector Machine (SVM) has been extended for indefinite kernels in a number of ways [18–20]. Initial work focused on preprocessing the kernel matrix through heuristics to address the indefiniteness [21]. A recent survey on indefinite learning is given in [17]. In [2] a stabilization approach was proposed to calculate a valid SVM model in the Krĕin space which can be directly applied on indefinite kernel matrices. This approach has shown great promise in a number of learning problems but has intrinsically quadratic to cubic complexity and provides a dense decision model. This paper extends the work of [2] by deriving an equivalent optimization problem but within the Core Vector Machine (CVM) framework [22]. To ensure linear runtime complexity we combine the proposed

* Corresponding author at: School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK.

E-mail addresses: schleif@informatik.uni-leipzig.de, schleify@cs.bham.ac.uk, fmschleif@googlemail.com, fschleif@techfak.uni-bielefeld.de (F.-M. Schleif), pxt@cs.bham.ac.uk (P. Tino).

http://dx.doi.org/10.1016/j.patcog.2017.06.003
0031-3203/

indefinite CVM with a low rank kernel approximation using the Nyström approach [23]. The latter one will also serve as a key element to sparsify the final solution such that an easy out of sample extension becomes possible. We empirically demonstrate the effectiveness of the proposed approach in comparison to the KSVM.

1.1. Indefinite kernels and existing approaches

Domain specific proximity measures, such as alignment scores in bioinformatics [24], the edit-distance for structural pattern recognition [25], shape retrieval measures (e.g. the inner distance [15]) and many other ones, generate non-metric or indefinite similarities or dissimilarities. Classical learning algorithms such as kernel machines assume metric properties in the underlying data space and may not be applicable for this type of data.

Only few machine learning methods have been proposed for non-metric proximity data, e.g. the indefinite kernel fisher discriminant (iKFD) [26,27] or the probabilistic classification vector machine (PCVM) [28]. The iKFD is a classical fisher discriminant approach, maximizing the between class variance of the classes, but formulated in the Krěin space, by using an equivalence relation to the classical kernel Fisher Discriminant Analysis.[1] In its original formulation, iKFD provides models which are naturally non-sparse and has cubic runtime complexity. The PCVM, on the other hand, constitutes a probabilistic model, operating with basis functions in the input space without the need for the existence of feature space (through Mercer kernel). While the iKFD is a batch optimization algorithm the PCVM is formulated by a gradient descent strategy with potentially slow convergence for a number of problems. The PCVM algorithms has cubic complexity in the first iterations with a substantial speed-up during further iterations due to an inherent sparsification strategy.

Recently the Krěin space Support Vector Machine (KSVM) was proposed in [2] leading to an SVM equivalent formulation, but fully formalized in the Krěin space by replacing the SVM minimization problem with a stabilization problem. As shown in [2] it turns out that solving the stabilization problem (detailed in [2],sec 2) can be achieved by flipping the negative eigenvalues of the kernel spectrum. It is shown in [2] that this strategy has a theoretical foundation and by solving the stabilization problem one can obtain the solution in the original Krěin space. This allows us to classify any new point without having to transform it.iKFD and PCVM have been found to be very effective but unlike KSVM, they are not based on the sound theoretical framework of the SVM structural risk minimization principle (SRM) [1]. Furthermore, there are a number of other advantages of KSVM as outlined in [2]. Hence, it is very attractive to obtain a low cost SVM formulation in the Krěin space, which is the focus of this paper.

1.2. Contributions

We consider the problem of training a Core Vector Machine with an indefinite kernel. The present paper is based on [2] in which the stabilization idea is proposed and on effective Nyström approximation concepts given in [34], both applicable to indefinite kernels. To ensure linear runtime complexity in contrast to at least quadratic costs of the KSVM we derive an indefinite Core Vector Machine using a low rank kernel approximation which solves the original indefinite SVM problem at low costs. We also suggest a sparsification procedure to simplify the out of sample extension. The Nystöm approximation is not necessary to obtain an indefinite Core Vector Machine, but to keep linear runtime and memory complexity which is lost otherwise.

[1] We do not detail the approach here because the paper will focus on an extension of KSVM.

2. Krěin space SVM

The Krěin Space SVM (KSVM) [2], replaced the classical SVM minimization problem by a stabilization problem in the Krěin space. The respective equivalence between the stabilization problem and a standard convex optimization problem was shown in [2]. Let $x_i \in X, i \in \{1, \ldots, N\}$ be training points in the input space X , with labels $y_i \in \{-1, 1\}$, representing the class of each point. The input space X is often considered to be $\mathbb{R}^d$, but can be any suitable space due to the kernel trick. For a given positive C, SVM is the minimum of the following regularized empirical risk functional

$$J_C(f, b) = \min_{f \in \mathcal{H}, b \in \mathbb{R}} \frac{1}{2} \|f\|_{\mathcal{H}}^2 + CH(f, b) \tag{1}$$

$$H(f, b) = \sum_{i=1}^{N} \max(0, 1 - y_i(f(x_i) + b))$$

Using the solution of Eq. (1) as $(f_C^*, b_C^*) := \arg\min J_C(f, b)$ one can introduce $\tau = H(f_C^*, b_C^*)$ and the respective convex quadratic program (QP)

$$\min_{f \in \mathcal{H}, b \in \mathbb{R}} \frac{1}{2} \|f\|_{\mathcal{H}}^2 \quad s.t. \sum_{i=1}^{N} \max(0, 1 - y_i(f(x_i) + b)) \leq \tau \tag{2}$$

where we detail the notation in the following. This QP can be also seen as the problem of retrieving the orthogonal projection of the null function in a Hilbert space $\mathcal{H}$ onto the convex feasible set. The view as a projection will help to link the original SVM formulation in the Hilbert space to a KSVM formulation in the Krein space. First we need to repeat a few definitions, widely following [2]. A Krěin space is an *indefinite* inner product space endowed with a Hilbertian topology.

Definition 1 (Inner products and inner product space)**.** Let $\mathcal{K}$ be a real vector space. An inner product space with an indefinite inner product $\langle \cdot, \cdot \rangle_{\mathcal{K}}$ on $\mathcal{K}$ is a bi-linear form where all $f, g, h \in \mathcal{K}$ and $\alpha \in \mathbb{R}$ obey the following conditions:

Symmetry: $\langle f, g \rangle_{\mathcal{K}} = \langle g, f \rangle_{\mathcal{K}}$,linearity: $\langle \alpha f + g, h \rangle_{\mathcal{K}} = \alpha \langle f, h \rangle_{\mathcal{K}} + \langle g, h \rangle_{\mathcal{K}}$and $\langle f, g \rangle_{\mathcal{K}} = 0 \; \forall g \in \mathcal{K}$ implies $f = 0$.

An inner product is positive definite if $\forall f \in \mathcal{K}$, $\langle f, f \rangle_{\mathcal{K}} \geq 0$, negative definite if $\forall f \in \mathcal{K}$, $\langle f, f \rangle_{\mathcal{K}} \leq 0$, otherwise it is indefinite. A vector space $\mathcal{K}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{K}}$ is called inner product space.

Definition 2 (Krěin space and pseudo Euclidean space)**.** An inner product space $(\mathcal{K}, \langle \cdot, \cdot \rangle_{\mathcal{K}})$ is a Krěin space if there exist two Hilbert spaces $\mathcal{H}_+$ and $\mathcal{H}_-$ spanning $\mathcal{K}$ such that $\forall f \in \mathcal{K}$, $f = f_+ + f_-$ with $f_+ \in \mathcal{H}_+$, $f_- \in \mathcal{H}_-$ and $\forall f, g \in \mathcal{K}$, $\langle f, g \rangle_{\mathcal{K}} = \langle f_+, g_+ \rangle_{\mathcal{H}_+} - \langle f_-, g_- \rangle_{\mathcal{H}_-}$. A finite-dimensional Krěin-space is a so called pseudo Euclidean space (pE).

If $\mathcal{H}_+$ and $\mathcal{H}_-$ are reproducing kernel hilbert spaces (RKHS), $\mathcal{K}$ is a reproducing kernel Krěin space (RKKS). For details on RKHS and RKKS see e.g.[35]. In this case the uniqueness of the functional decomposition (the nature of the RKHSs $\mathcal{H}_+$ and $\mathcal{H}_-$) is not guaranteed. In [36] the reproducing property is shown for a RKKS $\mathcal{K}$. There is a unique symmetric kernel $k(x, x)$ with $k(x, \cdot) \in \mathcal{K}$ such that the reproducing property holds (for all $f \in \mathcal{K}, f(x) = \langle f, k(x, \cdot) \rangle_{\mathcal{K}}$) and $k = k_+ - k_-$ where k_+ and k_- are the reproducing kernels of the RKHSs $\mathcal{H}_+$ and $\mathcal{H}_-$.

As shown in [36] for any symmetric non-positive kernel k that can be decomposed as the difference of two positive kernels k_+ and k_-, a RKKS can be associated to it. In [2] it was shown how the classical SVM problem can be reformulated by means of a stabilization problem. This is necessary because a classical norm as used in Eq. (2) does not exist in the RKKS but instead the norm is reinterpreted as a projection which still holds in RKKS and is