Neurocomputing 168 (2015) 190–199

Contents lists available at ScienceDirect

Neurocomputing

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

A local–global mixed kernel with reproducing property

Lixiang Xu [a,b,d], Xin Niu [b], Jin Xie [b], Andrew Abel [c], Bin Luo [a,*]

[a] School of Computer Science and Technology, Anhui University, Hefei 230601, Anhui, PR China
[b] Department of Mathematics & physics, Hefei University, Hefei 230601, Anhui, PR China
[c] Computing Science and Mathematics, University of Stirling, Stirling, FK9 4LA, Scotland
[d] Hefei VRview Information Technology Co., Ltd., Hefei 230088, Anhui, PR China

ARTICLE INFO

Article history:
Received 2 June 2014
Received in revised form
9 April 2015
Accepted 30 May 2015
Available online 9 June 2015

Keywords:
Reproducing kernel
Mercer kernel
Mixed kernel
Support vector machine

ABSTRACT

A wide variety of kernel-based methods have been developed with great successes in many fields, but very little research has focused on the reproducing kernel function in Reproducing Kernel Hilbert Space (RKHS). In this paper, we propose a novel method which we call a local–global mixed kernel with reproducing property (LGMKRP) to successfully perform a range of classification tasks in the RKHS rather than the more conventionally used Hilbert space. The LGMKRP proposed in this paper consists of two major components. First, we find the basic solution of a generalized differential operator by the delta function, and prove that this basic solution is a new specific reproducing kernel called a local *H*-reproducing kernel (LHRK) in RKHS. This reproducing kernel has good local properties, including odd order vanishing moment, and fast dilation attenuation. Second, in the RKHS, we prove that the LHRK satisfies the condition of Mercer's theorem, and prove that it is a typical polynomial kernel with global property, which also possesses the reproducing property. Furthermore, the novel specific mixed kernel (i.e., LGMKRP) proposed in this paper is based on these two different properties. Experimental results demonstrate that the LGMKRP possesses the approximation and regularization performance of a reproducing kernel, and can enhance the generalization ability of kernel methods.

1. Introduction

Kernel methods (KMs) have been successfully applied to a wide range of classification and recognition tasks [1,2]. The identification of an effective kernel for a given task is critical to most kernel based methods in machine learning [3–6]. For example, selecting the optimal kernel is a big challenge with regard to applying KMs to pattern recognition in Support Vector Machines (SVM). Much research has focused on proposing and evaluating a predefined parametric kernel, e.g., a polynomial or RBF kernel, but in this paper, we present a mixed kernel approach. There are many well developed techniques that have been previously proposed, including diffusion kernels [7], marginalized kernels [8], graph-based spectral kernels [9] and graph kernels [10,11]. In recent years, additional kernels have been proposed, including an improved Fisher kernel for large-scale image classification [12], a family of kernel descriptors to provide a unified and principled framework to turn pixel attributes into compact patch-level features [13], Fourier kernel learning [14], domain transfer multiple kernel learning [15], and the Quantum Jensen–Shannon Graph Kernel [10–11].

Kernels can be divided into two categories, local kernels (e.g., Gaussian kernel) and global kernels (e.g., polynomial kernel). A local kernel can present good interpolation abilities, meaning that only data points that are close to each other have an influence on the kernel values [5,16,17]. In comparison, a global kernel possesses interpolation ability as well as extrapolation ability. This means that it allows data points that are far away from each other to have an influence on the kernel values [18–20].

In order to attempt to obtain the best properties from both the approaches, a potentially more appealing approach is to learn a composite kernel from a fixed set of base kernels. This general framework is known as a mixed kernel framework. It has been shown in [21–23] that the use of this approach can result in both good interpolation and extrapolation abilities. In recent years, many frameworks and reviews of the mixed kernel approach have been presented. For example, an unbiased least squares support vector regression model with a composite kernel was proposed for reducing the computational complexity of a kernel machine's online modeling [24]. The constraints of time and memory will reduce the learning performance of SVM when it is used to solve a large number of samples. In order to solve this problem, a novel algorithm called Granular SVM based on mixed kernel function was proposed [25]. A challenging problem of pose recognition using simultaneous color and depth information was solved in [26] with the use of a local–global multi-kernel approach. Zhu et al. [23]

* Corresponding author. Tel.: +86 13856040965.
E-mail addresses: xulixianghf@163.com (L. Xu), luobin@ahu.edu.cn (B. Luo).

http://dx.doi.org/10.1016/j.neucom.2015.05.107
0925-2312/

proposed a method called mixed kernel canonical correlation analysis to achieve easy yet accurate implementation of dimensionality reduction. Another example is in [27], where a distribution spectrum was obtained by a continuum fitting method using a mixed Gaussian and exponential kernel function.

However, the kernel methods listed provide only a very limited discussion of the reproducing characteristics of the kernel function and do not fully discuss the case of a typical polynomial kernel with global property which possesses the reproducing property. There has been limited research focused on the reproducing characteristics of mixed kernels in Reproducing Kernel Hilbert Space (RKHS). The reproducing kernel function possesses good local properties (which will be discussed in more depth later in this paper), such as odd order vanishing moment, fast dilation attenuation, symmetry and regularity, and therefore, their utilization has many potential benefits. Therefore, in this paper, we propose the creation of a local–global mixed kernel with reproducing property (LGMKRP), which is based on the two different properties of the LHRK and global polynomial reproducing kernel (GPRK).

There are three main contributions of this paper. Firstly, we find the basic solution of a generalized differential operator, and prove that this basic solution is a new specific reproducing kernel, which is called a local H-reproducing kernel (LHRK). Some important and relevant properties of the LHRK are then discussed. Secondly, in the RKHS, we prove that the LHRK satisfies the condition of Mercer's theorem and demonstrate that the data in the neighborhood of a test point has a significant influence on its kernel value. Moreover, we also prove that a typical polynomial kernel with the global property possesses the reproducing property, known as polynomial RKHS, which contains a polynomial reproducing kernel. Based on this, we propose a novel specific mixture of kernels, which we call a local–global mixed kernel with reproducing property (LGMKRP), which is based on both of the different properties (i.e., local and global). This proposed method is evaluated with a range of extensive experiments. These compare our approach with a range of typical kernels using a number of publicly available datasets. The results of these experiments confirm the effectiveness of our method.

The remainder of this paper is divided as follows. Firstly, we describe the RKHS and provide some illustrative examples in Section 2. Section 3 discusses a reproducing kernel in RKHS and provides some important theorems and research, including the basic methodology, the properties of the H-reproducing kernel, the reproducing property of the polynomial kernel, model selection, and parameter tuning. In Section 4 we present a number of experimental results and discuss several in-depth application areas. Finally, the paper is concluded in Section 5.

2. Background

2.1. Reproducing Kernel Hilbert Space

Let $F(E)$ be the linear space comprising of all complex-valued functions on an abstract set E. Let H be a Hilbert space (possibly finite-dimensional) equipped with inner product $<\cdot,\cdot>_H$. Let

$$h: E \to H \tag{1}$$

be a Hilbert space H-valued function on E. We consider the linear mapping L from H into $F(E)$ to be defined by

$$f(q) = (Lg)(p) = <g, h(p)>_H. \tag{2}$$

The fundamental problems in the linear mapping in (1) are firstly the characterization of the function $f(p)$ and secondly the relationship between g and $f(p)$.

The approach to solve these fundamental problems is to form the function $K(p,q)$ on $E \times E$ defined by

$$K(p,q) = <g(q), g(p)>_H. \tag{3}$$

$R(L)$ denotes the range of L for H and we introduce the inner product in $R(L)$ induced from the norm

$$\|f\|_{R(L)} = \inf\{\|g\|_H; f = Lg\}, \tag{4}$$

then, from [28] we have Definition 1.

Definition 1. For the function $K(p,q)$ defined by (2), the space $[R(L), <\cdot,\cdot>_{R(L)}]$ is a Hilbert space satisfying the properties that

(1) for any fixed $q \in E$, $K(p,q)$ belongs to $R(L)$ as a function in p;
(2) for any $f \in R(L)$ and for any $q \in E$,

$$f(q) = <f(\cdot), K(\cdot,q)>_{R(L)}. \tag{5}$$

Further, the function $K(p,q)$ satisfying (1) and (2) is uniquely determined by $R(L)$. Furthermore, the mapping L is an isometry that maps from H onto $R(L)$ if and only if $\{h(p); p \in E\}$ is complete in H.

In Definition 1, the properties (1) and (2) of the function $K(p,q)$ are defined as the reproducing property of $K(p,q)$ in the Hilbert space $R(L)$, and the kernel $K(p,q)$ is called a reproducing kernel. A Hilbert space containing a reproducing kernel is known as a RKHS.

For clarity, we provide two simple and concrete examples of reproducing kernels in RKHS [28].

Example 1. Let $(e_1, e_2, \cdots, e_n)$ be an orthonormal basis in H and define

$$K(x,y) = \sum_{i=1}^{n} e_i(x)\overline{e}_i(y). \tag{6}$$

Then for any y in E,

$$K(\cdot,y) = \sum_{i=1}^{n} \overline{e}_i(y) e_i(\cdot), \tag{7}$$

belongs to H and for any function

$$\varphi(\cdot) = \sum_{i=1}^{n} \lambda_i e_i(\cdot). \tag{8}$$

In H, we have

$$\begin{aligned} \forall y \in E \ &<\varphi(\cdot), K(\cdot,y)>_H \\ &= <\sum_{i=1}^{n} \lambda_i e_i(\cdot), \sum_{i=1}^{n} \overline{e}_i(y) e_i(\cdot)>_H \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \overline{e}_j(y) <e_i, e_j>_H \\ &= \sum_{i=1}^{n} \lambda_i e_i(y) = \varphi(y). \end{aligned} \tag{9}$$

By Definition 1, we can therefore obtain $K(\cdot,y)$, the reproducing kernel of H.

Example 2. Let $K(i,j) = \sigma_{ij}$(*delta* function or Kronecker symbol, equal to 1 if $i=j$, to 0 otherwise).

Then

$$\forall j \in N,\ K(\cdot,j) = (0,0,\cdots,0,1,0,\cdots) \in H, (1 \text{ at the} j-\text{th place}), \tag{10}$$

$$\forall j \in N,\ \forall x = (x_i)_{i \in N} \in H,\ <x, K(\cdot,j)>_H = \sum_{i \in N} x_i \overline{\sigma}_{ij} = x_j. \tag{11}$$

So $K(\cdot,\cdot)$ is the reproducing kernel of H.