AN EXPLICIT CONSTRUCTION OF A REPRODUCING GAUSSIAN KERNEL HILBERT SPACE

Jian-Wu Xu, Puskal P. Pokharel, Kyu-Hwa Jeong, Jose C. Principe

Computational NeuroEngineering Laboratory
Department of Electrical and Computer Engineering, University of Florida
Gainesville, FL 32611 U.S.A.

ABSTRACT

In this paper, we propose a method to explicitly construct a reproducing kernel Hilbert space (RKHS) associated with a Gaussian kernel by means of polynomial spaces. In contrast to the conventional Mercer's theorem approach that implicitly defines kernels by an eigendecomposition, the functionals in this reproducing kernel Hilbert space are explicitly constructed and are not necessary orthonormal. We also point out an intriguing connection between this reproducing kernel Hilbert space and a generalized Fock space. We give an experimental result on approximation of the constructed kernel to a Gaussian kernel.

1. INTRODUCTION

Recently, several powerful kernel-based learning algorithms have been proposed, among which are support vector machines [11], kernel principal component analysis [10], kernel Fisher discriminant [6] and kernel independent component analysis [2]. Kernel-based algorithms are nonlinear versions of linear algorithms where the data has been nonlinearly transformed to a high dimensional, possibly infinite dimensional, feature space where we only need to compute the inner product of transformed data via the kernel functions. The attractiveness of kernel-based algorithms resides in their elegant treatment of nonlinear learning problems and efficiency for high-dimensional problems. Kernel methods have been successfully applied to classification, regression, density estimation and etc [9].

The reproducing kernel Hilbert space plays an important role in kernel-based learning algorithms. An RKHS is a Hilbert space of special properties [1]. In fact, the feature space into which the data is mapped is an RKHS where the nonlinear mapping Φ constitutes the basis. The nonlinear mapping Φ is related to an integral operator kernel $K(x, y)$ which corresponds to the inner product of transformed data:

$$K(x, y) =< \Phi(x), \Phi(y) > . \tag{1}$$

This work was partially supported by NSF grant ECS-0300340

The essence of kernel-based learning algorithm is that the inner product of the transformed data can be *implicitly* computed in the RKHS without *explicitly* using or even knowing the nonlinear mapping Φ. Hence, by applying kernels one can elegantly build a nonlinear version of a linear algorithm based on inner products.

One of the most fundamental issues in learning problems is the selection of the data representation. In kernel-based learning algorithms, this translates into the choice of the functionals, or the appropriate feature space RKHS. The reason is that the nonlinear mapping has a direct impact on the kernel and thus, on the solution of the given learning problems. Different kernels (polynomial, sigmoid, Gaussian) very likely will result in different performances. The advantage of kernel-based learning algorithms becomes also a disadvantage. The general question of how to select the ideal kernel for a given problem remains an open issue.

Recently, there have been attempts to explicitly construct an RKHS. A. Rakotomamonjy *et al* proposed a method of building an RKHS and its associated kernels by means of frame theory [7]. Any vector in that RKHS can be represented by linear combination of the frame elements. But a frame is not necessary linear independent although it results in stable representation. In the present work, we take the polynomial space approach to construct explicitly an RKHS associated with one of the most popular kernels, the Gaussian kernel. By transforming a generalized Fock space [3] with a positive operator, we build an RKHS associated with Gaussian kernel. The functionals, are explicitly given by the polynomials. Unlike the Mercer's theorem approach, these functional are not necessary orthornomal.

This paper is organized as follow. First we briefly introduce the Mercer's theorem. The main result of the paper is presented in the section 3, where a polynomial space method is proposed to explicitly construct a reproducing kernel Hilbert space associated with Gaussian radial basis kernel. Then we proceed to point out the connection between the RKHS and the generalized Fock space. In section 5, we present a toy example simulation to show the effectiveness of approximation of the constructed kernel to a Gaussian kernel.

2. MERCER'S THEOREM

One of the fundamental theorem in reproducing kernel Hilbert space research field is the Mercer's theorem, which gives the eigen-decomposition of a positive definite function (kernel) by orthonomalizing the kernel space.

Mercer's theorem: Suppose R is a continuous symmetric non-negative kernel on a closed finite interval T × T. Let $\{\varphi_k(x)$, k = 1, 2,...$\}$ be a sequence of normalized eigenfunction of the kernel R, and $\{\lambda_k$, k = 1, 2, ...$\}$ be the sequence of corresponding non-negative eigenvalues. In other word, for all integers k and j,

$$\int_T R(x,y)\varphi_k(x)dx = \lambda_k\varphi_k(y), \quad x,y \in T \tag{2}$$

$$\int_T \varphi_k(x)\varphi_j(x)dx = \delta_{k,j} \tag{3}$$

where $\delta_{k,j}$ is the Kronecker delta function, i.e., equal to 1 or 0 according as $k = j$ or $k \neq j$. Then

$$R(x,y) = \sum_{k=0}^{\infty} \lambda_k\varphi_k(x)\varphi_k(y) \tag{4}$$

where the series above converge absolutely and uniformly on T × T [5].

It follows that $R(x,y)$ can be rewritten as an inner product between two vectors in the feature space, i.e.,

$$\begin{aligned} R(x,y) &= <\Phi(x), \Phi(y)> \\ \Phi: x &\mapsto \sqrt{\lambda_k}\varphi_k(x), \qquad k = 1,2,.... \end{aligned} \tag{5}$$

Kernel-based learning algorithms use the above idea to map the data in the input space to a high-dimensional, possibly infinite-dimensional, feature space, which is an RKHS with functionals $\Phi(x)$ and inner product defined by equation (4) via the nonlinear mapping Φ. Instead of considering the given learning problem in input space, one can deal with the transformed data $\{\Phi_k(x), k = 1,2,...\}$ in feature space. When the learning algorithms can be expressed in terms of inner products, this nonlinear mapping becomes particular interesting and useful since one can employ the *kernel trick* to compute the inner products in the feature space via kernel functions without knowing the exact nonlinear mapping.

Mercer's theorem is more like an existence theorem because it does not give the explicit formula for the functionals in the RKHS. While kernel-based learning algorithms work well even without this knowledge since every quantity is expressed in terms of inner product of functionals (i.e., the kernels), their functional form is still of great interest to optimize performance and to get an insight into the appropriateness of the representation. Ultimately, this will allow us to utilize the RKHS structure and expand the class of algorithms, beyond inner products, that can be developed in kernel space. In the next section, we will present a method to construct an RKHS with the popular Gaussian kernel, with explicit expressions for the functionals using polynomial spaces.

3. EXPLICIT CONSTRUCTION OF AN RKHS BY POLYNOMIALS

In this section, we present the main result of this paper of explicitly constructing an RKHS using polynomials. First we give the definitions of functionals and inner product of a general Hilbert space. Then, a kernel function is imposed on this general Hilbert space to make it a reproducing kernel Hilbert space. This approach of building an RKHS with polynomials can also be found in [3], which is called generalized Fock space. Our contribution in this paper is that it is an RKHS associated with Gaussian kernel that we explicitly construct by introducing new definitions of functionals and kernel function. The connection between the generalized Fock space and our result will be discussed in section 4.

First we construct an inner product space $\mathcal{H}$ by defining functionals and inner product. The evaluation of functional *f* at any given *x* is given by

$$f(x) = e^{-\frac{x^2}{2\sigma_0}} \sum_{k=0}^{n} \frac{f_k}{k!} x^k \tag{6}$$

where σ_0 is a constant and (*n* + 1)-tuple (f_0,f_1, ... , f_n) are the coefficients which uniquely characterize the polynomial *f*. Then the inner product between any two functionals f and h can be specified in the form

$$< f, h >= \sum_{k=0}^{n} \frac{\sigma_k}{k!} f_k h_k \tag{7}$$

where f_k and h_k are coefficients for f and h respectively and σ = (σ_0, σ_1, ..., σ_n) is a set of positive constants chosen *a priori*. It can be easily seen that this inner product space $\mathcal{H}$ is complete thus forming a Hilbert space.

In order to make $\mathcal{H}$ a reproducing kernel Hilbert space, we impose a kernel function *K* on $\mathcal{H}$ in the following form

$$K(x,y) = e^{-\frac{x^2+y^2}{2\sigma_0}} \sum_{k=0}^{n} \frac{1}{k!\sigma_k} (xy)^k. \tag{8}$$

It can be verified that the Hilbert space $\mathcal{H}$, equipped with such K, is a *reproducing kernel Hilbert space* and the kernel function $K(x,\cdot)$ is a *reproducing kernel* because of the following two properties of $K(x,y)$:

1. $K(x,y)$ as a function of y belongs to $\mathcal{H}$ for any fixed x because we can rewrite $K(x,y)$ as

   $$K(x,\cdot)(y) = e^{-\frac{y^2}{2\sigma_0}} \sum_{k=0}^{n} \frac{\left(\frac{x^k}{\sigma_k} \cdot e^{-\frac{x^2}{2\sigma_0}}\right)}{k!} y^k \tag{9}$$

   i.e., the constants $(x^k/\sigma_k \cdot e^{-\frac{x^2}{2\sigma_0}})$, $k = 0,1,...,n$ become the coefficients $f_k, k = 0,1,...,n$ in the definition of f, and thus

   $$K(x,\cdot) \in \mathcal{H}. \tag{10}$$

2. Given any $f \in \mathcal{H}$, the inner product between reproducing kernel and f yields the function itself,

$$< K(x,\cdot), f >= \sum_{k=0}^{n} \frac{\sigma_k}{k!} (\frac{x^k}{\sigma_k} e^{-\frac{x^2}{2\sigma_0}}) \cdot f_k$$
$$= e^{-\frac{x^2}{2\sigma_0}} \sum_{k=0}^{n} \frac{f_k}{k!} x^k = f(x). \qquad (11)$$

This is so called *reproducing property*.

The RKHS constructed above has the freedom to choose the degree of functionals, i.e., the dimension n of the kernel space $\mathcal{H}$. The most interesting case is that we might extend it to an infinite-dimensional RKHS provided that the norm of functional is finite as $n \to \infty$, i.e., given a sequence of positive weighting constants satisfying certain conditions $\sigma = (\sigma_0, \sigma_1, ...)$,

$$\| f \|^2 = < f, f >= \sum_{k=0}^{\infty} \frac{\sigma_k}{k!} f_k^2 < \infty. \qquad (12)$$

Then the functionals, inner product and reproducing kernel in $\mathcal{H}$ will be defined by (6), (7) and (8) with $n = \infty$.

In the special situation of weights

$$\sigma_k = \sigma_0^k, \qquad k = 1, 2, ... \qquad (13)$$

where σ_0 is a fixed positive constant, then the reproducing kernel (8) in the infinite-dimensional RKHS becomes

$$\begin{aligned} K(x,y) = & \; e^{-\frac{x^2+y^2}{2\sigma_0}} \sum_{k=0}^{\infty} \frac{1}{k!} (\frac{xy}{\sigma_0})^k \\ = & \; e^{-\frac{x^2+y^2}{2\sigma_0}} e^{\frac{xy}{\sigma_0}} \\ = & \; e^{-\frac{(x-y)^2}{2\sigma_0}} \end{aligned} \qquad (14)$$

which is the Gaussian kernel used widely in machine learning, function approximation, density estimation, support vector machine, and etc. The constant σ_0 turns out to be the kernel width. It controls the norm length of those functionals in RKHS. i.e., the spread of nonlinear mapped data sample in feature space.

Comparing this method with Mercer's theorem, we notice that there are two major differences between them.

1. First, we have given an explicit expression for the functionals in the RKHS associated with Gaussian kernel in terms of polynomials while Mercer's theorem never does that. We can get the exact evaluations for those functionals at each point in the RKHS. This enables us to know exactly the structure of the RKHS associated with the Gaussian kernel.

2. Second, the functionals we constructed above are not necessary an orthonormal basis, while the Mercer's theorem is realized by orthonormalizing the RKHS. This perspective provides also a general alternative to build an RKHS from known functionals besides from Mercer's theorem.

The method we constructed an RKHS enables us to have the explicit expression of the functional in RKHS associated with Gaussian kernel. Hence we can exactly know the nonlinear mapping Φ used in the kernel-based learning algorithm and so operate directly with the transformed data to extend the algorithms beyond inner products. Furthermore, as we have the control of the dimension of the RKHS $\mathcal{H}$, this might help the computational complexity issue in kernel-based learning algorithms through approximation of Gaussian kernel by polynomials as indicated in equation (8).

4. CONNECTION WITH GENERALIZED FOCK SPACE

The idea of Fock space was first proposed by Fock in [4] to be used in quantum mechanics, where quantum states are described in the way of passing from one single object to collections of objects. More recently Figueiredo introduced an "arbitrarily weighted Fock space", which was called generalized Fock space in [3]. The space is equipped with an appropriate weighted inner product, thus forming an RKHS. The proposed RKHS has been used in liearn/nonlinear system and signal analysis, where a number of problems are involving approximation and inversion of nonlinear functions/functionals, and nonlinear operators [3].

In the univariate case, a generalized Fock space F^n is an RKHS, where the functionals f, inner product and kernel function F are defined as follows respectively,

$$f(u) = \sum_{k=0}^{n} \frac{f_k}{k!} u^k \qquad (15)$$

$$< f, h >_F = \sum_{k=0}^{n} \frac{\sigma_k}{k!} f_k h_k \qquad (16)$$

$$F(u,v) = \sum_{k=0}^{n} \frac{1}{k!\sigma_k} (uv)^k \qquad (17)$$

where the real $(n + 1)$-tuple $(f_0, f_1, ..., f_n)$ completely characterizes f, $\sigma = (\sigma_0, \sigma_1, ..., \sigma_n)$ is a set of positive weighting constants which are chosen *a priori* according to the problems under consideration. It can be shown that this generalized Fock space is an RKHS.

Similar to the RKHS $\mathcal{H}$ we constructed in section 3, the generalized Fock space F^n has the freedom of choosing the space dimension. The interesting case is that when the space becomes infinite dimensional while the norm of the functional satisfying the same condition (12) as $n \to \infty$. Then the kernel function $F(u, v)$, defined by (17), will become an exponential kernel as $n \to \infty$,

$$K(x,y) = e^{\frac{xy}{\sigma_0}} \qquad (18)$$

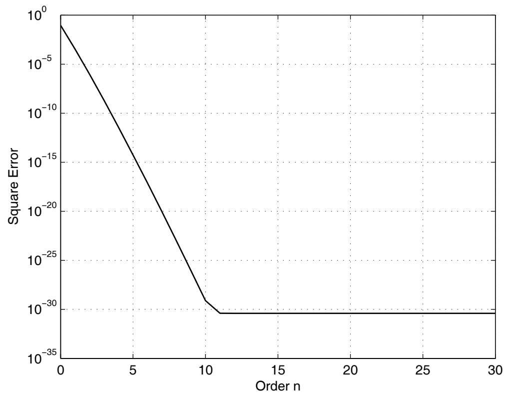

Fig. 1. Square Error between a Gaussian kernel and the constructed kernel in (8) versus the order of polynomials

given the same weights constraint as (13).

It can be noticed that there are similarity and difference between the RKHS $\mathcal{H}$ constructed in section 3 and the generalized Fock space F^n. The definitions for the inner product inside the two spaces are the same, while the functionals and kernel function are different. The relationship of the two spaces $\mathcal{H}$ and F^n is connected by a theorem in [1], which we will cite below without proof.

Theorem: Let H_1 and H_2 be two reproducing kernel Hilbert spaces of the same definitions of inner product, then there exits a positive operator with bound not greater than 1 that transforms H_1 into $H_2 \subset H_1$.

Comparing the definitions of functionals for two spaces, we can see that the $e^{-x^2/2\sigma_0}$ plays the role of a positive operator with bound not greater than 1, thus transforming the generalized Fock space F^n into $\mathcal{H}$ such that $\mathcal{H} \subset F$.

5. SIMULATIONS

We present a simple simulation here to show the effectiveness of approximation of the polynomial functionals to the Gaussian kernel. In the simulation, we calculate the square error between the Gaussian kernel and the proposed kernel in (8) of order n. The kernel width is chosen to be 1, and the range of the calculated data is from -5 to 5. The figure plots the square error versus order n. It suggests from the plot that with only order 11 we can effectively approximate the Gaussian kernel by polynomials. This also indicates that for practical purpose it is sufficient to work with much smaller dimensional space in stead of infinite dimensional space for a Gaussian kernel in kernel-based learning algorithms.

6. CONCLUSION

We have proposed a method to construct an RKHS associated with the Gaussian kernel by means of a polynomial space. Unlike the Mercer's theorem which orthonomalizes the RKHS by eigendecomposition of kernel, this method yields an RKHS with functionals that are not necessary orthonormal. Furthermore, it gives the explicit expressions of functionals which constitute the RKHS. This way of constructing the RKHS enables us to have the exact knowledge of the nonlinear mapping used in kernel-based learning algorithms based on the Gaussian kernel. The control of the dimension of the RKHS opens the possibility to deal with the computational complexity of kernel-based learning algorithm by approximating the Gaussian kernel via polynomials. On the other hand, we point out the close relationship between the RKHS associated with Gaussian kernel and the generalized Fock space.

7. REFERENCES

[1] N. Aronszajn, "Theory of Reproducing Kernels", *Trans. of American Mathematical Society*, Vol. 68, No. 3, pp. 337-404, May, 1950.

[2] F. R. Bach and M. I. Jordan, "Kernel Independent Component Analysis", *Journal of Machine Learning Research*, Vol. 3, pp. 1-48, 2002.

[3] Rui J. P. De Figueiredo, "A Generalized Fock Space Framework for Nonlinear System and Signal Analysis", *IEEE Trans. of Circuits and System*, Vol. CAS-30, No. 9, pp. 637-647, May, 1983.

[4] V. Fock, "Konfigurationsraum und zweite quantelung", *Z. Phys.*, Vol. 75, pp. 622-647, 1932.

[5] J. Mercer, "Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations", *Philosophical Trans. of Royal Society of London, Series A,* Vol. 209, pp. 415-446, 1909.

[6] S. Mika, G. Ratsch, J. Weston, B. Scholkopf and K.-R. Muller, "Fisher discriminant analysis with kernels", in *Neural Networks for Signal Processing IX*, pp. 41-48, New Jersey,1999.

[7] A. Rakotomamonjy and Stephane Canu "Frames, Reproducing Kernels, Regularization and Learning", *Journal of Machine Learning Research*, Vol.6, pp. 1485-1515, 2005.

[8] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, New York, Ungar, 1955.

[9] B. Scholkopf and A. Smola, *Learnig with Kernels*, MIT Press, 2002.

[10] B. Scholkopf, A. Smola and K. -R. Muller "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", *Neural Computation* Vol. 10, pp. 1299-1319, 1998.

[11] V. N. Vapnik, *Statistical Learning Theory*, New York, Wiley, 1998.