International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:4, No:5, 2010

Comparative Studies of Support Vector Regression between Reproducing Kernel and Gaussian Kernel

Wei Zhang, Su-Yan Tang, Yi-Fan Zhu, and Wei-Ping Wang

Abstract—Support vector regression (SVR) has been regarded as a state-of-the-art method for approximation and regression. The importance of kernel function, which is so-called admissible support vector kernel (SV kernel) in SVR, has motivated many studies on its composition. The Gaussian kernel (RBF) is regarded as a "best" choice of SV kernel used by non-expert in SVR, whereas there is no evidence, except for its superior performance on some practical applications, to prove the statement. Its well-known that reproducing kernel (R.K) is also a SV kernel which possesses many important properties, e.g. positive definiteness, reproducing property and composing complex R.K by simpler ones. However, there are a limited number of R.Ks with explicit forms and consequently few quantitative comparison studies in practice. In this paper, two R.Ks, i.e. SV kernels, composed by the sum and product of a translation invariant kernel in a Sobolev space are proposed. An exploratory study on the performance of SVR based general R.K is presented through a systematic comparison to that of RBF using multiple criteria and synthetic problems. The results show that the R.K is an equivalent or even better SV kernel than RBF for the problems with more input variables (more than 5, especially more than 10) and higher nonlinearity.

Keywords—admissible support vector kernel, reproducing kernel, reproducing kernel Hilbert space, support vector regression.

I. Introduction

SUPPORT vector regression (SVR) [1] has been widely applied in the field of regression and approximation. It is a novel sparse kernel modeling method whose objective is to learn an unknown function based on a training set of N input-output pairs in a black box modeling approach [2]. It's shown that SVR possesses many advantages, e.g. no local optima, good ability of generalization, intrinsic regularization and the sparseness of support vectors, etc. These advantages encourage researchers focus on applying it into various fields, e.g. approximation [2], [3], prediction [4], [5] and other applications [6]. The tutorial can be seen in [7], [8].

It's well known that the approximation performance of SVR lies on the training data and kernel function. A kernel is called admissible support vector kernel (SV kernel) [8] if the Mercer's condition [9] is satisfied. Mercer's condition is one of popular methods to validate whether a prospective kernel is a positive definite function since any SV kernel should be capable of corresponding to a dot product in high dimensional feature space. Kernel function is regarded as a significant trick which benefits the computation of dot products in feature space using simple function defined on pairs of input patterns [10], [11]. In addition, the SV kernel implies the features of data in feature space since it contains all the information about the relative positions of data, i.e. choosing different kernels will produce different SVMs.

It's usually, however, time-consuming and demanding to validate a SV kernel. It's known that almost all the methods, e.g. Mercer's method, only tell us whether or not a prospective kernel is actually a dot product in a given space, but it does not show how to construct the feature map and the images of the input data in the feature space and even what the feature space is. The best choice of the best choice of a kernel for a given problem is still an open research issue [12], though there are some kernels, e.g. polynomial kernel $K(x,x') = (< x,x' > +1)^d$, Gaussian kernel (RBF) $K(x,x') = \exp(- \parallel x - x' \parallel /(2\sigma^2))$ and sigmoid kernel $K(x,x') = \tanh(v < x,x' > +c)$. It's found that the polynomial kernel is usually inferior in the problem with higher nonlinearity and sigmoid kernel performs closely to RBF but with complex form conditional satisfaction with Mercer's condition, and consequently seems obscure to the non-specialist [8]. Research has shown that RBF is not only theoretically well-founded but also superior in some practical classification applications [12], [13]. However, the performance of RBF is sensitive to the parameter σ [14], and there is no evidence that the RBF is the optimal choice for regression, especially dealing with multivariable complex functions.

Therefore, many researches are devoted to the study on the composition method of SV kernels and related properties, e.g. hybrid composition method based on some operations of kernels, e.g. positive linear combinations, integrals and products, etc. [8], [15], [16], multi-scale kernel [17] and wavelet kernel [18], [19] as well as the feature space of kernel mapping [20], such as reproducing kernel Hilbert space (RKHS) [21], [22], etc. Recently, the multi-scale kernel and RKHS becomes the research focuses. Although the former adopts techniques from wavelet theory and shift invariant spaces to construct a new class of kernels, it still bases on the framework of RKHS [17]. Therefore, we pay our attention to the kernel in RKHS.

RKHS owes the name to the so-called reproducing kernel(R.K) function, which could be regarded as a SV kernel. Although the basis concept and principle [23], frames [24], properties [25], and conceptual comparison of R.K to the other kernels, e.g. Mercer kernel, positive definite kernel (PDK) [26], etc., have been well studied, there are relatively little work on quantitative analysis and comparison in SVR

Wei Zhang is with the College of Information System and management, National University of Defense Technology (NUDT), Changsha, Hunan, 410073, China (e-mail:the_ant@163.com).

Su-Yan Tang is with the College of Information System and Management, NUDT, Changsha, Hunan, China. (e-mail: tsy2977162@163.com).

Yi-Fan Zhu is with the College of Information System and management, NUDT, Changsha,Hunan, China (e-mail: nudtzyf@hotmail.com).

Wei-Ping Wang is with the Graduate School, NUDT, Changsha, Hunan, China, (e-mail: wangwp@nudt.edu.cn).

based on some R.Ks with explicit forms. Firstly, there is a notorious problem i.e. parameter selection in SVR, which usually hinders the applicability of SVR. Secondly, very little work has been published on the methods for computing R.K, and consequently a limited number of R.Ks are available. And finally, it's shown that some operations of simpler R.K can compose more complex R.Ks [23]. It results in a capability of handling multiple inputs separately. In other words, the R.K can handle different input dimensions in a closed form with different nonlinear mapping functions based on the need of modelers or some credible prior knowledge, such as the independence among some input dimensions. However the conventional kernels, e.g. RBF, do not implicate the potential knowledge in their constructions.

In this paper, a new composition method of SV kernel based on R.K is proposed and two SV kernels with explicit forms are composed based on a simpler R.K in Sobolev RKHS $\mathcal{H}^1(\mathbb{R}; a, b)$. Subsequently, some systematic comparative studies on fitting precision and efficiency of the R.Ks to RBF are presented for eight synthetic problems under different criteria. The results show that these R.Ks perform closely to RBF in the problems with fewer input dimensions (less than 5) and relatively lower nonlinearity whereas superiorly in ones with more dimensions (more than 5, especially more than 10) and higher nonlinearity.

II. Preliminary

A. SVR Formulation

Given an training set $\mathcal{D} = \{(x_i, y_i), i = 1, ..., l\} \subset \Omega \times \mathbb{R}$, where Ω denotes the space of the input data (e.g. $\Omega = \mathbb{R}^d$, where d denotes dimensionality of input). All the SV algorithms aim at minimizing an upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data, which is based on the structural risk minimization principle [8]. It is to train a model as $y =< w, \phi(x) > +b$, which minimizes a general risk function as follows:

$$\frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{l} L(y_i, f(x_i)) \tag{1}$$

where w controls the flatness of the model, $\phi(x)$ is a mapping function, b is the bias, $< \cdot, \cdot >$ denotes the dot product, constant $C > 0$ determines the trade-off between error minimization and the maximization of the function flatness. In this paper, the ε-insensitive loss functionL_ε [1] is used

$$L_\varepsilon(y, f(x)) = |y - f(x)|_\varepsilon = \max\{0, |f(x) - y| - \varepsilon\} \tag{2}$$

where $\varepsilon \geq 0$ is a constant controlling the noise tolerances.

It's well-known that SVR can be formulated as the following quadratic programming (QP) problem [8] which can be solved efficiently by many well-documented optimization algorithms:

$$\begin{aligned} \min_{\alpha, \alpha^*} \ & \tfrac{1}{2} \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \\ & + \sum_{i=1}^{l} (\alpha_i + \alpha_i^*)\varepsilon - \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*) y_i \\ s.t. \ & \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C], i = 1, ..., l \end{aligned} \tag{3}$$

Consequently, the regression model takes a form as follows:

$$f(x) = \sum\nolimits_{i \in SV} (\bar{\alpha}_i - \bar{\alpha}_i^*) K(x_i, x) + b \tag{4}$$

where $i \in SV$ denotes the indices of support vectors (SVs), i.e. $x_i$with nonzero $\bar{\alpha}_i$ or $\bar{\alpha}_i^*$, $K(\cdot, \cdot)$is the kernel function

Obviously, the complexity of (4) depends only on the amount of SVs (ASV) and SV kernel rather than the dimensionality of the input space Ω. In fact, the SVs, which depend on the selection of kernel and coefficients of SV algorithm [12], can be automatically extracted. In other words, the major task of the SVM lies in the selection of its kernel [15].

B. Conditions for SV Kernel

Kernel function is a crucial ingredient in SVR, and a kernel is called a SV kernel if it satisfies Mercer's condition[8], since the kernel used in QP formulation (3), has to be a positive definite function. This paper is mainly focus on SV kernels with positive definiteness that are appropriate for general discussion, though there are lots of works on replacing the QP by a linear programming (LP) [27], [28]. Obviously, any SV kernel also can be employed in a LP formulation.

Choosing different kernel functions will produce different SV algorithms and may result in different performances [15]. It is because, as stated in the previous section, different SV kernel implies different feature space, and consequently different reflection of the feature of the estimation function.

The question that raises now is, whether a function $K(s, t)$ corresponds to a dot product in a feature space. There are many researches, e.g. [1], [8], [29], [30]. The following theorems, including Mercer' and Bochner's theorem, represent the function.

Theorem 1: Let Ω be a closed subset of $\mathbb{R}^n$, $n \in \mathbb{N}$, μ is a Borel measure on Ω. Suppose $K \in L_\infty(\Omega^2)$ such that the integral operator $T_K : L_2(\Omega) \to L_2(\Omega)$ defined by

$$T_K f(\cdot) := \int_\Omega K(\cdot, x) f(x) d\mu(x) \tag{5}$$

is semi-positive. Let $\psi_i \in L_2(\Omega)$ be the eigenfuntion of T_K associated with the eigenvalue $\lambda_i \neq 0$ and normalized such that $\parallel \psi_i \parallel_{L_2} = 1$ and let $\overline{\psi_i}$ denote its complex conjugate. Then

(i) $(\lambda_i(T))_i \in l_1$

(ii) $\psi_i \in L_\infty(\Omega)$ and $\sup_i \parallel \psi_i \parallel_{L_\infty} < \infty$

(iii) $K(x, x') = \sum_{i \in \mathrm{N}} \lambda_i \overline{\psi_i(x)} \psi_i(x')$ (referred to as Mercer kernel) holds for almost all (x, x'), where the series converges absolutely and uniformly for almost all (x, x').

Less formally speaking this theorem means that if

$$\int_{\Omega \times \Omega} K(x, x') f(x) f(x') dx dx' \geq 0, for \ all f \in L_2(\Omega) \tag{6}$$

holds, we can write $K(x, x')$ as a dot product in some feature space, i.e. any function $K(x, x')$ who satisfies Mercer's condition is a SV kernel. Unfortunately, the validation is still of difficulty and intractability.

Theorem 2: Given a positive finite Borel measure μ on $\mathbb{R}$, the Fourier transform $\mathcal{Q}$ of μ, i.e.$\mathcal{Q}(t) = \int_{\mathrm{R}} e^{-itx} d\mu(x)$ is a continuous function, then $\mathcal{Q}$ is a positive definite function and

vice versa. In other words, every positive definite function is the Fourier transform of a positive finite Borel measure, i.e. the kernel takes the form $K(x,x') = \mathcal{Q}(x-x')$ is positive definite, and vice versa.

Here, the kernel function in theorem 2 is called translation invariant kernel, e.g. RBF $K(x,x') = \exp(- \parallel x-x' \parallel^2 /(2\sigma^2))$. Smola *et al.* [29] presented the following method for validating a SV kernel based on the Bochner's theorem [30].

Theorem 3: A kernel $K(x,x') = K(x-x')$ is an admissible SV kernel if and only if the Fourier transform

$$F[K](\omega) = \hat{K}(\omega) = (2\pi)^{-\frac{d}{2}} \int_{\Omega} e^{-i<\omega,x>} K(x)dx \qquad (7)$$

is nonnegative.

Moreover, for kernels $K(x,x') = K(< x,x' >)$(dot-product kernel), there exists sufficient conditions for being admissible. We do not detail it as it's not concerned with in this paper, for further details see [31].

III. Perspectives of Reproducing Kernel

A. Definition of Reproducing Kernel

The abstract theory of RKHS has been developed over a number of years outside the domain of SVR [23]. A variety of applications, especially in data interpolation and smoothing, are dealt in a RKHS, because the RKHS provides a rigorous and effective framework for smooth multivariate interpolation of arbitrarily scattered data and for accurate approximation of general multidimensional functions [32], [33]. In this section, some basic concepts are introduced briefly. For more details on RKHS see e.g. [21], [23], [25], [34].

Defintion 1: Let $\Omega \subseteq \mathbb{R}^d$ be an arbitrary nonempty set, $\mathcal{H}$ is a Hilbert space of function $f : \Omega \to \mathrm{R}$ (short for$f \in \mathbb{R}^{\Omega}$). $\mathcal{H}$ is called a reproducing kernel Hilbert space (RKHS) if there exists $K : \Omega \times \Omega \to \mathbb{R}$, satisfies the following:

(i) $\forall x$, $K_x(y) = K(y,x)$ as a function of y belongs to $\mathcal{H}$.

(ii) The reproducing property: $\forall x \in \Omega$, and $\forall f \in \mathcal{H}$,

$$f(x) =< f, K_x > \qquad (8)$$

(iii) $\mathcal{H}$ is spanned by K, i.e., $\mathcal{H} = \overline{span\{K_x(\cdot)|x \in \Omega\}}$

Here, v is called the native space of K [25].

Defintion 2: (R.K) $K : \Omega \times \Omega \to \mathbb{R}$ is called a R.K of $\mathcal{H}$, if it satisfies the conditions (i) and (ii) in Definition 1.

The R.K possesses some basic properties, e.g. uniqueness, existence, positive definiteness, convergence and projection, etc. [23]. Additionally, there are some attractive properties which would be contributed to compose more complex R.Ks, that is, let K_i (i=1,2) is the R.K of the RKHS $\mathcal{H}_i$ with the norms $\| \cdot \|_i$, then

Property 1: $K = K_1 + K_2$ is the R.K of a RKHS $\mathcal{H}$ of all functions$f = f_1 + f_2$ with $f_i \in \mathcal{H}_i, i = 1,2$, and with the norm defined by $\| f \|^2 = \min\{\| f_1 \|_1^2, \| f_2 \|_2^2\}$, i.e. the minimum taken for all the decompositions $f = f_1 + f_2$ where $f_i \in \mathcal{H}_i, i = 1,2$.

Note that the property can be extended to the case where $K = \sum_{i=1}^{n} K_i$. In addition, the difference of R.Ks is also a R.K; more details see [23] for reference as well.

Property 2: The direct product of $\mathcal{H}_1$ and $\mathcal{H}_2$ possesses a R.K $K(x_1,x_2,y_1,y_2) = K_1(x_1,y_1)K_2(x_2,y_2)$.

From property 2, we see immediately that the kernel $K(x,y) = K_1(x,y)K_2(x,y)$ is positive definite as the restriction of the kernel $K(x_1,x_2,y_1,y_2)$ to the subset $\Omega_1 \subset \Omega$ consisting of the "diagonal" element $\{x,x\} \in \Omega$ as shown in [23]. Similarly to property 1, the product property also can be extended to the case where $K = \prod_{i=1}^{n} K_i$.

B. Relations between SV Kernel and Reproducing Kernel

It's necessary to discuss the relations between various kernels to validate whether a R.K can be used as a SV kernel. It is hoped that the discussion here would help to bridge the conceptual gap between some familiar kernels, e.g. positive (semi-)definite kernel (PDK), Mercer kernel and R.K, whereas some of the observations are not new or profound.

Defintion 3: Let Ω be a subset of $\mathbb{R}^n$, $n \in \mathbb{N}$, $K : \Omega \times \Omega \to \mathbb{R}$ is symmetric and positive (semi-)definite (PD), if and only if for arbitrary finite sets $\{x_1,...,x_m\} \subseteq \Omega$, the matrix $\mathbf{K} = (K(x_i,x_j))_{1 \le i,j \le m}$ is symmetric and positive definite, i.e. $\forall m \in \mathbb{N}$, $\forall c_i \in \mathbb{R}$, for any $x_1,...x_m \in \Omega, i = 1,..m, K$ satisfies the following inequation

$$\sum_{i,j=1}^{m} c_i c_j K(x_i,x_j) \ge 0 \qquad (9)$$

Theorem 4: $K : \Omega \times \Omega \to \mathbb{R}$ is a SV kernel iff K is a PDK.

The proof is obvious. Refer to e.g. [35].

Theorem 5: $K : \Omega \times \Omega \to \mathbb{R}$ is a Mercer kernel iff K is a PDK.

Proof: if K is a Mercer kernel, i.e. there exists a map function Φ such that $K(t,s) =< \Phi(s), \Phi(t) >$. Then,

$$\begin{aligned} \sum_{i,j=1}^{m} c_i c_j K(x_i,x_j) &= \sum_{i,j=1}^{m} c_i c_j < \Phi(x_i), \Phi(x_j) > \\ &= \| \sum_{i=1}^{m} c_i \Phi(x_i) \|^2 \ge 0 \end{aligned}$$

thus, K is a PDK according to (9).

For the converse, if K is a PDK, K is a Mercer kernel according to Theorem 4 and 1, which completes the proof.

Theorem 6: $K : \Omega \times \Omega \to \mathbb{R}$ is a Mercer kernel iff there exists a RKHS $\mathcal{H}$ with R.K K, i.e. $\mathcal{H}_K(\Omega)$.

Proof: According to *Moore–Aronszajn* Theorem [23], any PDK K is associated with a space $\mathcal{H}_K(\Omega)$ and vice versa. Note that the Theorem 5 holds if K is a PDK, that is, K is a Mercer kernel, which completes the proof.

IV. Illustrative Examples

Almost all the researches on SVR in RKHS framework are limited to theoretic rather than practical study, e.g. representation theorem, prime and dual expression [36] etc., though it has proven that any R.K can be used as a SV kernel. On the one hand, the fact that the RBF is regarded as a R.K in an unknown RKHS and shows significant performance, in some sense, hinders the studies on the performance of more general R.K in practical applications. On the other hand, it's always a difficult and challenging task for computing a R.K with explicit form [23], [37]. In this paper, an exploratory research on the performance of SVR based general R.K will be discussed.

Noted that it is concerned as a time consuming and demanding task to conclude whether a function could strictly satisfy

Mercer's or Bochner's theorem or not. In fact, there are quite limited off-the-shelf SV kernels, especially the translation invariant kernels which strictly satisfy the Theorem 2 or 3, except for the R.K, namely$K_{\mathcal{H}}(x,y)$, in Sobolev RKHS $\mathcal{H}^1(\mathbb{R};a,b)$ [38]. Consequently, it is expected to provide a new alternative to compose more complex SV kernels for SVR.

Defintion 4: (Sobolev RKHS) Sobolev RKHS $\mathcal{H}^1(\mathbb{R};a,b)$ is a space consisting of all absolutely continuous functions $f(x)$, $x\in\mathbb{R}$ and with the following finite norm:

$$\| f \| = \int_{\mathrm{R}} a^2 \left|f'(x)\right|^2 + b^2 \left|f(x)\right|^2 dx < \infty,\ a,b>0 \quad (10)$$

The corresponding R.K $K_{\mathcal{H}}(x,y)$ is as follows:

$$K_{\mathcal{H}}(x,y) = \frac{1}{2ab} e^{-\frac{b}{a}|x-y|} = \frac{1}{2\pi}\int_{\mathrm{R}} \frac{\exp\left(i\omega(x-y)\right)}{a^2\omega^2+b^2} d\omega \quad (11)$$

Note that $K_{\mathcal{H}}$ in (11) is a SV kernel, since it is a translation invariant kernel, and $\hat{K}_{\mathcal{H}}(\omega) = (b^2+a^2\omega^2)^{-1} \geq 0$, where $\hat{K}_{\mathcal{H}}$ denotes the Fourier transform of $K_{\mathcal{H}}$. In other words, $K_{\mathcal{H}}$ satisfies Theorem 3.

Suppose $\overline{x}\in\mathbb{R}^n$ is an input, where $x^i\in\mathbb{R}$ is the i^{th} component of $\overline{x}$. According to properties of R.K, we can obtain two complex R.Ks based on (11), that is,

$$(i) K_{PRK}(\overline{x},\overline{y}) = (\sum\nolimits_{i=1}^{n} K_{\mathcal{H}}(x^i,y^i))/n \quad (12)$$

$$(ii) K_{MRK}(\overline{x},\overline{y}) = \prod\nolimits_{i=1}^{n} K_{\mathcal{H}}(x^i,y^i) \quad (13)$$

Corollary 1: $K_{PRK}(\overline{x},\overline{y})$ (PRK for short) is a SV kernel.

Proof: Since $K_{\mathcal{H}}(x^j,y^j)$ is a R.K, then $K_{\mathcal{H}}(x^j,y^j)$is a PDK from Theorem 4. In other words, for $\forall m\in\mathbb{N}$, $x_1^i,...,x_m^i\in\mathbb{R}$, $\forall c_j\in\mathbb{R}$, $j=1,...,m$, we have

$$\sum\nolimits_{j,k=1}^{m} c_j c_k K_i(x_j^i,x_k^i) \geq 0 \quad (14)$$

Since for any $\bar{x}\in\mathbb{R}^n$, it can be uniquely composed by $x^i\in\mathbb{R}$, $i=1,...,n$, then we have

$$\begin{aligned}\sum_{j,k=1}^{m} c_j c_k K_{PRK}(\bar{x}_j,\bar{x}_k) &= \sum_{j,k=1}^{m} c_j c_k \sum_{i=1}^{n} a_i K_{\mathcal{H}}(x_j^i,x_k^i)\\ &= \sum_{i=1}^{n} a_i \sum_{j,k=1}^{m} c_j c_k K_H(x_j^i,x_k^i) \geq 0\end{aligned} \quad (15)$$

Therefore, PRK is a PDK according to (9), and consequently is a SV kernel from Theorem 4, which completes the proof.

Corollary 2: $K_{MRK}(\overline{x},\overline{y})$ (MRK for short) is a SV kernel.

Proof: In fact that $K_{\mathcal{H}}$is a Mercer kernel, since $K_{\mathcal{H}}$ is a R.K. From Mercer's theorem, $\forall m\in\mathbb{N}$, the following kernel Gram matrix $\mathbf{K}_{\mathcal{H}}$ of $K_{\mathcal{H}}$ to $x^1,...,x^m\in\mathrm{R}$

$$\mathbf{K}_{\mathcal{H}} := \left(K_{\mathcal{H}}(x^i,x^j)\right)_{i,j=1}^{m} \quad (16)$$

is positive (semi-)definite.

Using a classical Schur product theorem, it is easy to prove that the kernel Gram matrix $\mathbf{K}$ of MRK is also a positive (semi-)definite matrix. Then, MRK is a Mercer kernel, and also a SV kernel from Theorem 4 and 5, which achieves our assertion.

V. Synthetic Problems and Test Scheme

A. Features of Synthetic Problems

To test the performance of SVR based on different SV kernels, eight synthetic problems are selected and classified based on the features stated in [39], i.e.

(a) *Problem scale* (dimensionality of input). Three relative scales are considered, i.e. small scale (dimensionality is 2∼5, S for short), medium scale (dimensionality is 6∼9, M for short) and large scale (dimensionality≥10, L for short).

(b) *Nonlinearity of behavior*. Similarly to [39], the problems are classified into two categories: low-order nonlinearity (functions which are polynomial or that can be transformed to polynomial with degree less than 4, L for short) and high-order nonlinearity (otherwise, H for short).

(c) *Smoothness of performance behavior*. In this paper, the two forenamed features, i.e. *problem scale* and *nonlinearity order* are major research focus, therefore the noisy behavior is artificially created using local variations of a smooth function as shown in Table I. "No" denotes smooth without any noise and "Yes" denotes noisy behavior.

TABLE I
FEATURES OF SYNTHETIC PROBLEMS

Problem No.	Scale (No. of variables)	Non-linearity order	Noisy behavior	Symbol
P1	Small (l =2)	Linear	No	S-L
P2	Small (l =2)	Low-order nonlinear	No	S-L
P3	Small (l =2)	Low-order nonlinear	Yes	S-L
P4	Small (l =2)	High-order nonlinear	No	S-H
P5	Medium (l =6)	Low-order nonlinear	No	M-L
P6	Medium (l =6)	High-order nonlinear	No	M-H
P7	Large (l =10)	Low-order nonlinear	No	L-L
P8	Large (l =10)	High-order nonlinear	No	L-H

A summary of the features of the eight synthetic problems is given in Table I, and some symbols will be used in the next section. These problems utilized in this paper are or similar to the problems in [39], which are listed in the Appendix.

B. Parameter Selection in SVR based on Genetic Algorithm and Data Sampling based on Latin Hypercube Design

Parameter selection is a notorious problem since SV algorithm is very sensitive to the adequate choice of parameter values [7], which makes it hard for non-experts. Fortunately, there are only a handful of parameters, i.e. 1) regularization constant C, 2) tolerance error ε, 3) coefficients of SV kernel itself, e.g. kernel width σ in RBF and a, b in MRK and PRK, will impact on the prediction performances. It is, however, a combinatorial optimization problem, and also a NP-hard problem, to select a segment from thousands of their infinite combinations. Lots of papers have shown that genetic algorithm (GA) [40], [41] is useful to solve the combinatorial problem without prior knowledge. The GA based on GAOT toolbox with its standard settings is used to obtain the best parameters evolutionally [42], since the strategy of setting parameters is not our research focus.

Furthermore, the performance of SVR under different training and validating sample sizes, i.e. *small data set* (S for short) and *large data set* (L for short), is also compared. In order to sample more valuable training and validating sample, a good design of experiment (DOE) is very important [43], [44], because the information included in the training data set determines the performance of regression and prediction. Here, a conventional reduced sampling technique, i.e. Latin Hypercube sampling (LHS), is employed to sample data.

In this paper, two observation sets $\{x_i, y_i\}_{i=1}^{n}$, i.e. large set (n=200, L for short) and small set ($n = 100$, S for short), are generated by LHS, where half of them are selected randomly as training data and others as validating data to assess the accuracy of newly predicted points.

Remark: some tags will be introduced to denote different test schemes with different synthetic problems listed in Table I , e.g. "S-L-S" denotes S-L synthetic problems trained with *small data set*, where the third letter denotes the sample size.

C. Metrics for Performance Measures

To evaluate the performance of SVR based on general R.K, two qualitative criteria, i.e. fitting precision and efficiency, are used to compare the performance of MRK, PRK to that of RBF.

1) *Fitting Precision*. Including accuracy and robustness, where (i) accuracy means the capability of predicting the system response over the design space of interest and (ii) robustness means the capability of achieving good accuracy for different problems types and sample sizes.
2) *Efficiency*. The computational effort required for training a SVR and predicting new data sets.

To provide a more complete picture of precision and efficiency, the criteria above can be measured by several quantitative metrics, i.e. R square (R^2), relative average absolute error ($RAAE$), relative maximum absolute error ($RMAE$), which are used to measure the fitting precision, and modeling time (MT), amount of SVs (ASV) are employed to evaluate the efficiency. The $R^2 RAAE$ and $RMAE$ are given in (17)-(19), respectively.

$$R^2 = 1 - \sum_{i=1}^{l} (y_i - \hat{y}_i)^2 \Big/ \sum_{i=1}^{l} (y_i - \overline{y})^2 \qquad (17)$$

$$RAAE = \sum_{i=1}^{l} |y_i - \hat{y}_i| \Big/ \sum_{i=1}^{l} |y_i - \overline{y}| \qquad (18)$$

$$RMAE = n \times \max\{|y_i - \hat{y}_i|\}_{i=1}^{l} \Big/ \sum_{i=1}^{n} |y_i - \overline{y}| \qquad (19)$$

where $\hat{y}_i$ denotes the corresponding predicted value for observed value y_i, $\overline{y}$ denotes the mean of the observed values.

Generally speaking, 1) the larger the value of R^2, the more accurate the SVR; 2) the smaller the value of $RAAE$, the more accurate the SVR; 3) a small $RMAE$ is preferred and large $RMAE$ indicates large error in one region of the design space. However, it is not as important as R^2 and $RAAE$. Furthermore, the variance indicates the robustness of accuracy, i.e. the smaller the variance, the more robust the kernel function; more details see in [39], [44].

For the convenience of defining the fitness function in GA, a new measure, *Integration Precision* (IP), is introduced:

$$IP = \alpha(\beta R^2 + (1-\beta)/RAAE) + (1-\alpha)/RMAE \qquad (20)$$

where $\alpha, \beta \in [0, 1]$ are weights. In this paper, $\alpha = 0.9, \beta = 0.5$ to indicate that R^2 and $RAAE$ are more important than $RMAE$. It is obvious that the larger the IP, the more precise the SVR. Furthermore, the optimal results mentioned latter imply the computation result with the "best" parameters when IP is largest.

Furthermore, the larger ASV and MT, the more inefficient in SVR, where MT indicates the used training and validating time on existing data set, and ASV, according to (4), implies the predicting efficiency in new data.

VI. Simulation Results and Analysis

Based on the proposed schemes for comparative study, there are $36 * 10 * 1000 = 360000$ SVR models are trained for the eight synthetic problems (see Table I), where there are 12 test schemes for each three kernels, i.e. the *small scale* problems are trained only with *small data set* and others are trained under both sets as stated above. Moreover, 10 and 1000, which implies *population size* and *maximum generation*, are the parameters in GA,

A. Fitting Precision

To illustrate the performance of SVR based on PRK, MRK and RBF under different schemes, multiple bar-charts are shown. While the mean indicates the average performance of SVR, the variance illustrates the robustness of the performance. Henceforth, the performance of SVR based on a certain kernel is called that of the kernel for short, e.g. the accuracy of RBF.

1) Overall Performance: Illustrated in Figs. 1 and 2, the mean and variance of the precision metrics for all three kernels under all the test schemes, i.e. different problem scales, orders of nonlinearity, smoothness and sample sizes, are shown.

Fig. 1 shows that the average accuracies of R.Ks and RBF are close for all the test schemes, though more strictly speaking RBF is slightly superior to the R.Ks. However, in terms of the robustness for all the optimal results shown in Fig. 2, RBF is no longer the best kernel, because the variances of R^2 and $RAAE$ are distinctly larger than R.Ks. In other words, the features of test schemes have less impact on R.Ks than that on RBF. Moreover, the result that RPK possesses the better robustness than MRK is also shown in Fig. 2.

Overall, R.K is shown to be an equivalent or even better SV kernel than RBF, in terms of the average accuracy and robustness. Especially, it is shown that PRK possesses the best robustness, whereas there is a drawback in practice for PRK as the parameters have to be selected appropriately to ensure the robustness for given problems.

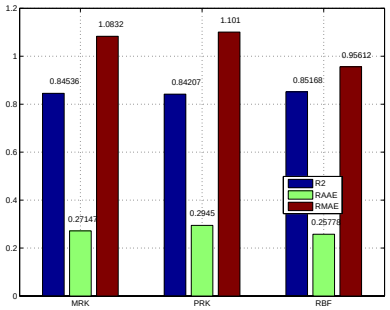

Fig. 1. Mean of Precision Metrics with Optimal Results

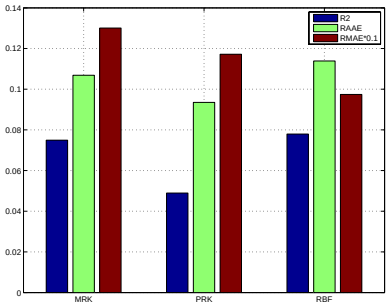

Fig. 2. Variance of Precision Metrics with Optimal Results

2) Performance for Different Types of Problems: Figs.3 and 4 illustrate the mean and variance bar-charts for different types of problems. In these figures, the labels for each subfigure are listed in Table I. The values in Figs.3 and 4 are derived based on the data from all sample sizes (small and large). It is noted that:

(i) Roughly speaking, the order of accuracy of all the three kernels based SVR is *small* $>$ *medium* $>$ *large*, and the higher the nonlinearity, the lower the accuracy.

(ii) It's shown that MRK and RBF perform closely in all types of problems, whereas PRK performs worse for S-L, S-H and M-L problems but best for others, i.e. M-H, L-L and L-H problems. The numerical results can be seen in Table II. Especially for M-H, L-L and L-H problems, RBF has worst accuracy, which implies that RBF is not the optimal choice for the cases that the dimensionality of input is relatively large ($>$5) and regression curve is rough, even though the SVR is trained with the best parameters.

(iii) All the three kernels have similar variance of R^2 and $RAAE$ for all problem types, which implies that they possess the similar global accuracy (R.Ks have smaller variances for large scale problems). However, the variances of $RMAE$ in R.K are relative larger than RBF for S-L, S-H and L-L problems, which means R.Ks can not fit as well as RBF in local areas, shown in Table III (all the values were multiplied by 1000).

Overall, the fact that all the kernels perform best for S-L, S-H, M-H and L-L problems (where their R^2 all close to 1 shown in Table II) indicates SVR is adapted to approximate the functions within these problem types. Moreover, R.K-based SVR performs better than RBF-based SVR in medium and large scale problems. It implies that RBF is just an acceptable rather than optimal kernel in SV algorithm at all time.

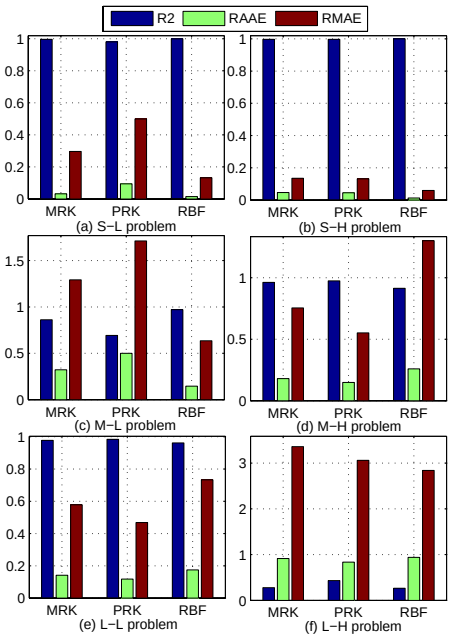

Fig. 3. Mean of Accuracy Metrics for Different Types of Problems

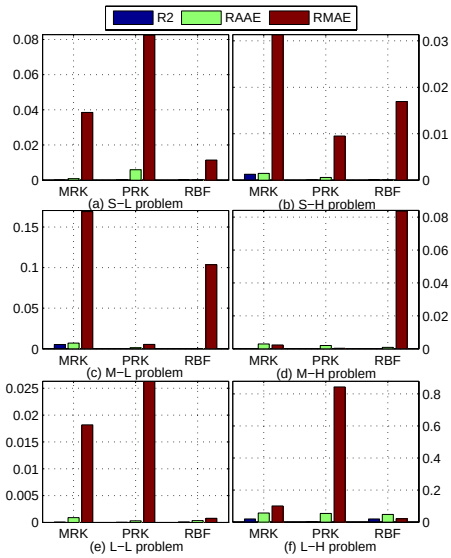

Fig. 4. Variance of Accuracy Metrics for Different Types of Problems

3) Performance under Different Sample Size: Fig 5 shows the mean of accuracy performance of the kernels under different sample sizes (i.e. S: *small set* and L: *large set*) respectively, where the labels for tick marks along the x axis denote the different problem types with different sample sizes, for example "P6 L" and "P6 S" denote the problem 6 listed in Table I are trained with large and small scale sample respectively.

It's shown that these three kernels can achieve good fitting precision for P5 L to P7 S, where R.Ks are better than RBF for

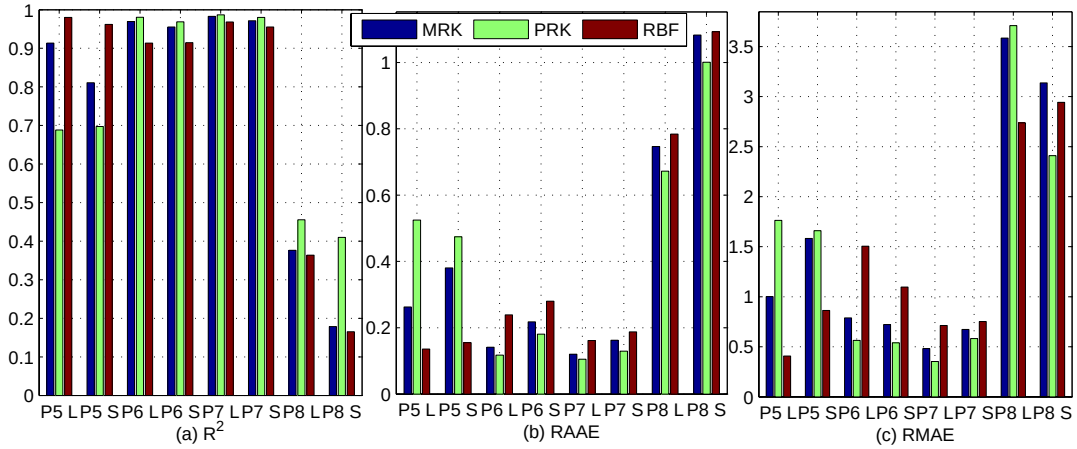

Fig. 5. Mean of Accuracy Metrics under Different Sample Scales ((a) R^2, (b) $RAAE$ and (c) $RMAE$)

TABLE II
SUMMARY OF MEAN OF ACCURACY METRICS FOR PROBLEM TYPES

		S-L	S-H	M-L	M-H	L-L	L-H
R^2	MRK	0.99622	0.99720	0.86217	0.96237	0.97727	0.27742
	PRK	0.98034	0.99694	0.69269	0.97461	**0.98355**	**0.43261**
	RBF	**0.99925**	**0.99974**	**0.97103**	0.91422	0.96165	0.26443
	Best	**RBF**	**RBF**	**RBF**	**PRK**	**PRK**	**PRK**
$RAAE$	MRK	0.03201	0.04637	0.32181	0.17964	0.14137	0.91483
	PRK	0.09454	0.04504	0.49973	**0.14901**	**0.11734**	**0.83658**
	RBF	**0.01476**	**0.01227**	**0.14554**	0.25954	0.17472	0.93863
	Best	**RBF**	**RBF**	**RBF**	**PRK**	**PRK**	**PRK**
$RMAE$	MRK	0.29695	0.13526	1.29190	0.75520	0.57797	3.36105
	PRK	0.49971	0.13141	1.71138	**0.55182**	**0.46775**	3.06000
	RBF	**0.13214**	**0.05927**	**0.63484**	1.30120	0.73280	**2.84003**
	Best	**RBF**	**RBF**	**RBF**	**PRK**	**PRK**	**RBF**

TABLE III
SUMMARY OF VARIANCE OF ACCURACY METRICS FOR PROBLEM TYPES

		S-L	S-H	M-L	M-H	L-L	L-H
R^2	MRK	0.02500	4.28711	5.28699	0.10294	0.06394	19.5798
	PRK	0.31821	0.24991	**0.04231**	0.06704	**0.02336**	**1.05634**
	RBF	**0.00111**	**0.02100**	0.16647	**0.00081**	0.08341	19.7319
	Best	**RBF**	**RBF**	**PRK**	**RBF**	**PRK**	**PRK**
$RAAE$	MRK	0.86582	4.94504	6.93542	2.91023	0.89302	56.5760
	PRK	5.93907	4.21719	1.28327	2.00939	**0.29934**	54.0109
	RBF	**0.12273**	**0.00807**	**0.18631**	**0.84272**	0.33971	**47.6868**
	Best	**RBF**	**RBF**	**RBF**	**RBF**	**PRK**	**RBF**
$RMAE$	MRK	38.5129	106.280	168.962	2.28069	18.1802	100.672
	PRK	82.4435	69.7875	**5.39769**	**0.32046**	26.2748	843.211
	RBF	**11.4125**	**11.6976**	103.529	83.7204	**0.77053**	**21.1269**
	Best	**RBF**	**RBF**	**PRK**	**PRK**	**RBF**	**RBF**

all the problem types except for P5. In addition, the impacts of sample size on average accuracy of all the kernels are relatively smaller for P6 and P7. It's also observed that, the smaller the sample size, the lower the accuracy.

B. Efficiency

The efficiency of each kernel-based SVR is measured by the time used for SVR training and new predictions The former, referred to as MT, which includes two parts, i.e. the time for training SVR with the given training data set and the time for validating with test data set, depends on the problem scale and the sample size. And the time used for a new prediction just depends on the amount of SVs and kernel type. In this paper, the MT is recorded on Matlab7.5 workstation with its "stopwatch timer" function.

1) Variations of Modeling Time: Fig. 6 shows the mean of MT for different problem types and sample sizes. Some conclusions can be summarized as follows:

(i) The MT increases with (a) the *problem scale* and (b) the *order of nonlinearity*;

(ii) The larger the sample data set for training SVR, the larger the MT;

(iii) The MTs of PRK and RBF are close for all test schemes, whereas that of MRK are distinct smallest.

It's obvious that MRK is the most efficient kernel for

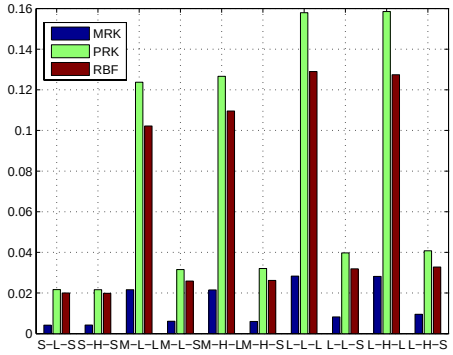

Fig. 6. Statistics of MT under Different Test Schemes

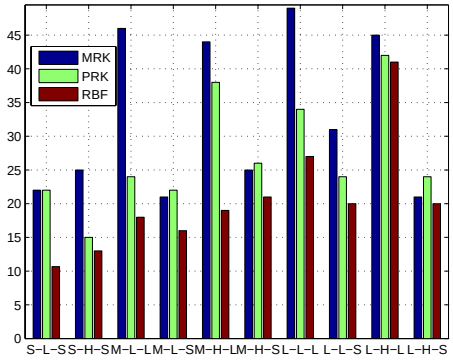

Fig. 7. Statistics of ASV under Different Test Schemes

training SVR models. It's found that, relatively speaking, RBF is somewhat superior to PRK for large scale problems, however greatly inferior to MRK for all the test schemes. Moreover, the sample size has more dramatic impact on PRK and RBF than MRK, which implies R.K is more efficient than RBF.

2) Variations of the Amount of Support Vector: Fig. 7 illustrates the mean of the amount of SVs (ASV) under all the test schemes. It's noted that:

(i) The ASV generally increases with (a) the problem scale and (b) the order of nonlinearity. However, strictly speaking, the ASV is largest for the large scale problem with higher nonlinearity for a large sample set, because the potential SVs, which carry the most information, concentrate in the rough regions of a curve.

(ii) The larger the sample data set for training SVR, the larger the ASV;

(iii) MRK generates more SVs than RBF, though the MT of MRK is smallest. Furthermore, RBF and PRK have the similar ASV for all the test schemes, which indicates that they possess the same generalization ability. Whereas, strictly speaking, the ASV of RBF is smallest, which may explain that why RBF is usually employed in SV algorithms for problems without any prior information. Overall, the efficiency of R.K is close or sometimes superior to that of RBF, especially when there are large number of training samples and problem dimensions.

VII. Conclusion

The systematic comparative study presented in this paper has provided insightful observations into the performance of the SVR based on the conventional RBF and two general R.Ks.

In terms of the accuracy and efficiency of SVR based on the three kernels for different all the test schemes, it's noted that conventional RBF can only perform good for small scale problems and medium scale problem with lower nonlinearity, whereas the R.Ks are superior to RBF for others (shown in Table IV). It implies that R.K can be used as an equivalent or even better SV kernel than RBF for more complex regression problems.

TABLE IV
Summary of Best SV Kernel for Accuracy

	Small scale	Medium scale	Large scale
Low-order nonlinearity	RBF	RBF	PRK
High-order nonlinearity	RBF	PRK	PRK
Overall	RBF	RBF/PRK	PRK

It's believed that the SVR is a superior method for approximating complex functions, e.g. a function with higher dimensionality (>5) and nonlinearity (i.e. L-H problem), to other black-box modeling approaches, such as polynomial. It's also observed that the sample size and distribution as well as the underlying noise have significant impact on the performance of the SVR. Therefore, appropriate amount of training data of good quality are needed to ensure a fast and accurate approximation and prediction. In other words, an effective data sampling technique (such as sequential design or adaptive sampling strategy) is regarded as an important step before training a SVR model.

Finally, more synthetic problems are to be considered in the future work. Moreover, the incorporation of prior knowledge (e.g. some useful structural or numeric knowledge) in SVR can improve the quality of the regression model, which is of interests to further explore.

Appendix A

P1: $f(x) = 4x_1 - 5x_2$

P2: $f(x) = 0.5x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2$

P3: $f(x) = 0.5x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2 + Noise(rand)$ where $Noise(rand) \sim N(0, 0.5)$ denotes adding some $N(0, 0.5)$ noise at some random points.

P4: $f(x) = \exp(-x_1^2 - x_2^2) + 3\sin(x_1)$

P5: $f(x) = (x_1 - 1)^2 + 0.5(x_2 - x_3)^2 + 0.5(0.5x_4 + 0.5x_5 - x_6)^2$

P6: $f(x) = \sum_{j=1}^{6} \left\{ \exp(x_j) \left(c_j + x_j - \ln \left(\sum_{k=1}^{6} \exp(x_k) \right) \right) \right\}$, $c_j = -6.09, -17.16, -24.05, -15.91, -24.72, -14.99$

P7: $f(x) = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 7)^2 - 4(x_4 - 5)^2 + (x_5 - 3)^2 - 2(x_6 - 1)^2 + 5x_7^2 - 7(x_8 - 5)^2 + 2(x_9 - 10)^2 - (x_{10} - 7)^2 + 45$

P8: $f(x) = \sum_{j=1}^{10} [(\ln(x_j - 2))^2 + (\ln(10 - x_j))^2]$, $x_j \in [2.1, 9.9]$

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the financial support of the National Defense Pre-Research Foundation of China (Grant No. 9140C640505), the National Natural Science Foundation of China (No. 60974073 and No. 60974074) .

REFERENCES

[1] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[2] G. Bloch, F. Lauer, G. Colin, and Y. Chamaillard, "Support vector regression from simulation data and few experimental samples," *Information Sciences*, vol. 178, pp. 3813–3827, 2008.
[3] J.-B. Gao, S. R. Gunn, and C. J. Harris, "Mean field method for the support vector machine regression," *Neurocomputing*, vol. 50, pp. 391–405, 2003.
[4] M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, "Support vector machines for wind speed prediction," *Renewable Energy*, vol. 29, no. 6, pp. 939–947, 2004.
[5] W.-W. He, Z.-Z. Wang, and H. Jiang, "Model optimizing and feature selecting for support vector regression in time series forecasting," *Neurocomputing*, vol. 73, no. 3, pp. 600–611, 2008.
[6] F. Pan, P. Zhu, and Y. Zhang, "Metamodel-based lightweight design of b-pillar with twb structure via support vector regression," *Computers and Structures*, vol. 88, pp. 36–44, 2010.
[7] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
[8] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
[9] J. Mercer, Ed., *Functions of positive and negative type and their connection with the theory of integral equations*, ser. Philosophical Transactions of the Royal Society, London, 1909, vol. A, 209.
[10] B. E. Boser, I. M. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, D. Haussler, Ed. Pittsburgh, PA: ACM Press, 1992, pp. 144–152.
[11] B. Schpölkopf, "The kernel trick for distances," *Neural Information Process. Systems (NIPS)*, vol. 13, 2000.
[12] B. Schölkopf, K.-K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with gaussian kernels to radial basis function classifiers," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2758–2765, 1997.
[13] D. Anguita and G. Bozza, "The effect of quantization on support vector machines with gaussian kernel," in *Proceedings of International Joint Conference on Neural Networks*, Montreal, Canada, 2005, pp. 681–684.
[14] X.-Y. Zhang and Y.-C. Liu, "Performance analysis of support vector machines with gauss kernel," *Computer Engineering*, vol. 29, no. 8, pp. 22–25, 2003.
[15] Y. Tan and J. Wang, "A support vector machine with a hybrid kernel and minimal vapnik-chervonenkis dimension," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, pp. 385–395, 2004.
[16] J.-X. Liu, J. Li, and Y.-J. Tan, "An empirical assessment on the robustness of support vector regression with different kernels," in *Proceedings of the 4th International Conference on Machine Learning and Cybernetics*, vol. 7, Guangzhou, China, 2005.
[17] R. Opfer, "Multiscale kernels," *Advances in Computational Mathematics*, vol. 25, pp. 357–380, 2006.
[18] L. Zhang, W.-D. Zhou, and L.-C. Jiao, "Wavelet support vector machine," *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, vol. 34, pp. 34–39, 2004.
[19] X.-G. Zhang, D. Gao, X.-G. Zhang, and S.-J. Ren, "Robust wavelat support machines for regression estimation," *International Journal Information Technology*, vol. 11, no. 9, pp. 35–46, 2005.
[20] A. J. Smola, B. Schölkopf, and K.-R. Müller, "The connection between regularization operators and support vector kernels," *Neural Networks*, vol. 11, pp. 637–649, 1998.
[21] G. Wahba, "Support vector machines,reproducing kernel hilbert spaces and randomized gacv," in *Advances in Kernel Methods - Support Vector Learning*, B. Sch?lkopf, C. J. Burges, and A. J. Smola, Eds. Cambridge, England: MIT Press, 1999, pp. 69–88.
[22] L.-M. Ma and Z.-M. Wu, "Kernel based approximation in sobolev spaces with radial basis functions," *Applied Mathematics and Computation*, vol. 215, pp. 2229–2237, 2009.
[23] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.
[24] J. Gao, C. J. Harris, and S. R. Gunn, "Support vector kernel based on frames in function hilbert spaces," *Neural Computation*, vol. 13, pp. 1975–1994, 2001.
[25] R. Schaback, "A unified theory of radial basis functions native hilbert spaces for radial basis functions ii," *Journal of Computational and Applied Mathematics*, vol. 121, pp. 165–177, 2000.
[26] R. Schaback and H. Wendland, "Approximation by positive definite kernels," in *Advanced Problems in Constructive Approximation*, M. D. Buhmann and D. H. Mache, Eds. Birkh?user, Basel: Verlag, 2002, pp. 203–221.
[27] A. J. Smola, B. Schölkopf, and G. Rätsch, "Linear programs for automatic accuracy control in regression," in *Proceedings of the 9th international conference on artificial neural networks*, vol. 2, Edinburgh, UK., 1999.
[28] O. L. Mangasarian and D. R. Musicant, "Large scale kernel regression via linear programming," *Machine Learning*, vol. 46, no. 1-3, pp. 255–269, 2002.
[29] A. J. Smola, B. Schölkopf, and K.-R. Mller, "General cost functions for support vector regression," in *Proceedings of Ninth Australian Conf. on Neural Networks*, Brisbane, Australia, University of Queensland, 1998, pp. 79–83.
[30] S. Bochner, *Lectures on Fourier Integral*. Princeton, New Jersey: Princeton University Press, 1959.
[31] A. J. Smola, Z. L. Óvri, and R. C. Williamson, "Regularization with dot-product kernels," in *Advances in Neural Information Processing Systems*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds., vol. 13. MIT Press, 2001, pp. 308–314.
[32] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadephia, 1990.
[33] J. Stöckler, "Multivariate bernoulli splines and the periodic interpolation problem," *Constr. Approx.*, vol. 7, pp. 105–120, 1991.
[34] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Boston, Dordrecht, London: Kluwer Academic Publishers Group, 2003.
[35] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, U.K: Cambridge University Press, 2000.
[36] B. Schölkopf, R. Herbrich, A. J. Smola, and R. C. Williamson, "A generalized representer theorem," Tech. Rep. NeuroCOLT Technical Report 2000-81, 2000.
[37] W. Zhang, "The construction of reproducing kernel and some approximating problems in the reproducing kernel spaces," Ph.D. dissertation, National University of Defense Technology, 2005.
[38] R. A. Adams, *Sobolev Spaces*. New York: Academic Press, 1975.
[39] R. Jin, W. Chen, and T. W. Simpson, "Comparative studies of metamodeling techniques under multiple modeling criteria," in *Proceedings of the 8th AIIA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, 2000.
[40] Y.-J. Park, "Application of genetic algorithms in response surface optimization problems," Doctor of Philosophy, Arizona State University, December 2003.
[41] S. S. Chaudhry and W. Luo, "Application of genetic algorithms in production and operations management: A review," *International Journal of Production Research*, vol. 43, pp. 4083–4101, 2005.
[42] C. R. Houck, J. A. Joines, and M. G. Kay, "A genetic algorithm for function optimization: A matlab implementation," Tech. Rep. NCSU-IE TR 95-09, 1995.
[43] W. Zhang, H. Wu, J. Liu, Y.-F. Zhu, and Q. Li, "A study of exploratory analysis experimental design supporting robust decision-making," *Journal of System Simulation*, vol. 21, no. 14, pp. 4461–4466, 2009.
[44] J. P. C. Kleijnen and R. G. Sargent, "A methodology for fitting and validating metamodels in simulation," *European Journal of Operational Research*, vol. 120, no. 1, pp. 14–29, 2000.