International Journal of Innovative Computing, Information and Control
Volume **8**, Number **7(A)**, July **2012**
ICIC International ©2012 ISSN 1349-4198
pp. 4811–4822

CONSTRUCTING KERNELS BY FUZZY RULES FOR SUPPORT VECTOR REGRESSIONS

Fengqiu Liu and Xiaoping Xue*

Department of Mathematics
Harbin Institute of Technology
No. 92, West Dazhi Street, Nangang District, Harbin 150001, P. R. China
*Corresponding author: xiaopingxue@263.net

Received April 2011; revised August 2011

Abstract. *This study focuses on designing a new class of kernels to incorporate the prior information into the training process of support vector regressions. The prior information in the form of fuzzy rules is considered for regression problems. First, the antecedent of each fuzzy rule is represented by some fuzzy equivalence relations. Moreover, the properties of kernels and pseudo-metrics are employed to discuss the conditions for fuzzy equivalence relations to be kernels. Then the kernels for each of the fuzzy rules are obtained by using the given additive generators and arbitrary pseudo-metrics as well as triangular norms. Furthermore, a class of kernels is obtained by linearly combining the kernels corresponding to each rule via fuzzy entropies for all the fuzzy rules. Finally, we apply this class of kernels to support vector regressions. The experimental results help quantify the performance of the proposed approach.*

Keywords: Kernel, Fuzzy rule, Fuzzy equivalence relation, Support vector regression

1. Introduction.

Support vector machines were first developed for pattern recognition by representing the decision boundary with support vectors. They minimize the upper bound of generalization error by using the structural-risk minimization principle [1-3]. Support vector machines have been a popular method in machine learning and are widely used in many areas, such as isolated handwritten digit recognition, text categorization, face detection and the control of dynamic systems [4, 5]. By introducing Vapnik's ε-insensitive loss function, support vector machines were generalized to the case of regression estimation, referred to as support vector regressions (SVRs). SVRs compute linear regression function in a high-dimensional feature space where the input data are mapped via a feature map.

1.1. Summary of related work and motivation.

A certain disadvantage of the standard SVR stems from the fact that prior information about the problems cannot be incorporated into the learning process. Since the prior information is capable of enhancing the performance of algorithms, extensive work in [6-13] has focused on modifying the standard SVR to make full use of prior information in the training process. Concluding the above mentioned work, there are three ways of prior information incorporation into SVRs: namely, prior information is transformed into constraints of the related optimization problem [6, 7, 12]; prior information is incorporated to weight the kernels or support vector regressors [8-11]; prior information is used to construct kernels [12, 13]. Furthermore, prior information takes different forms, such as information with certainty [6, 7, 12], probabilistic information [9-11] and information in the form of fuzzy rules [8, 13].

In particular, prior information in the form of fuzzy rules is usually obtained from field information and experts. Fuzzy rules are capable of dealing with uncertainty in learning problems in a way of human reasoning. Therefore, the incorporation of fuzzy rules and SVRs has been investigated. For example, some positive definite kernels are represented by fuzzy bi-implications based on fuzzy-logical concepts in [13]. Fuzzy rules are usually transformed into weights of training samples so that they can be incorporated into kernels or regression algorithms [8, 14, 15]. However, there is no consideration on how to use the kernels as well as the performance of kernel-based algorithms from theoretical aspects and applications in [13]. The work in [8, 14, 15] only transforms the prior information into the weights of training points and neglects the weights of test points. That is, they do not deal fairly with training and test points though the points are always assumed to be independent and identically distributed. Consequently, the corresponding SVRs fail to generalize from the training set to the test set. The generalization ability of SVRs decreases. Additionally, we note that the methods in [6, 7, 10-12] transform the prior information into the constraints of the related optimization problems. By these methods, the prior information is transformed into new training data before it is included in the learning procedure. Thus, this method cannot be used to ensure that the parameters determined by the derived optimization are suitable for all the data determined by the prior information. The algorithms might not generalize the training samples to the test samples corresponding to the prior knowledge. Therefore, it becomes important to develop a systematic and effective approach to incorporate fuzzy rules into SVRs.

1.2. **Main idea and contributions.** This study aims at developing an approach to integrate SVRs with fuzzy rules to improve the performance of SVRs. We construct a new class of kernels according to the antecedents of fuzzy rule by fuzzy equivalence relations and fuzzy entropy theory. By using the designed kernels, fuzzy rules are incorporated into the training process of SVRs.

The contribution of this work is three-fold. First, we provide a valid method to construct kernels. These kernels are more suitable than the positive definite kernels by free choice for the regression problem with fuzzy rules in applications. In fact, the classical techniques are concerned with constructing the positive definite kernels [1, 13]. To the best of our knowledge, few references focused on the construction of negative definite and conditionally positive definite kernels by some additive generators and arbitrary pseudo-metrics. Second, compared with the methods in [8, 14, 15], the proposed method is capable of dealing fairly with all the samples because the obtained kernels are symmetric. Additionally, the proposed approach integrates fuzzy rules into the training process by the designed kernels, which ensures that the optimal parameters are suitable for test points with the information. Thus, the SVRs based on the designed kernels are able to generalize from training data to test data in the applications. Finally, compared with the method of directly applying fuzzy rules to regression problems, such as Takagi-Sugeno models, Vapnik-style results are able to guarantee keeping the generalization error low for the kernel-based SVRs. However, to the best of our knowledge, it may not be the case for the methods of directly applying fuzzy rules to regression problems.

The remaining paper is structured as follows. Prerequisites are given in Section 2. The regression problem is stated in Section 3. Fuzzy rules are kernelized in Section 4. Applications of the proposed approach are given in Section 5. Finally, conclusions are drawn in Section 6.

2. **Prerequisites.**

We introduce some definitions and results in fuzzy mathematics and kernel theory which will be used in this study. Triangular norms and conorms are indispensable tools for inference and aggregation of fuzzy rules. We only use triangular norms (t-norms) in this study. The minimum $T_M(x, z) = \min\{x, z\}$, the product

$T_P(x,z) = xz$ and the Łukasiewicz $T_L(x,z) = \max\{x+z-1, 0\}$ are three basic t-norms. Let $f : [0,1] \to [0,\infty]$ be *an additive generator*, i.e., a strictly decreasing, which is right-continuous at 0 with $f(1) = 0$. We can construct Archimedean t-norms by using additive generators [16]. From [16], a t-norm $T : [0,1]^2 \to [0,1]$ is an Archimedean t-norm if and only if there exists an additive generator f such that for any $(x,z) \in [0,1]^2$

$$T(x,z) = f^{-1}\left(\min(f(x) + f(z), f(0))\right). \tag{1}$$

The concept of fuzzy equivalence relation was first introduced by Zadeh in [17] as a generalization of the concept of an equivalence relation. Fuzzy equivalence relations play important roles in different fields as a measure of similarity.

Definition 2.1. [17] *Let $\mathcal{G}$ be a non-empty set. $E : \mathcal{G}\times\mathcal{G} \to [0,1]$ is called a fuzzy relation. E is called a fuzzy equivalence relation with respect to the t-norm T, if the following conditions: (1) reflexivity: i.e., $E(x,x) = 1$, (2) symmetry: i.e., $E(x,y) = E(y,x)$, and (3) T – transitivity: i.e., $T(E(x,y), E(y,z)) \le E(x,z)$ are satisfied for any $x,\ y,\ z \in \mathcal{G}$.*

Let B^T be the transpose of a matrix B. 1_r stands for the $r \times r$ identity matrix. $diag(a_1, \cdots, a_n)$ denotes the diagonal matrix with entries given by $a_1, \cdots, a_n$. $\mathcal{E}$ is a nonempty set. For arbitrary $x,\ z \in \mathcal{E}$, $\langle x, x\rangle_{\mathcal{E}} := x^T I_{p,q} x$ is an inner product, where p and q are positive integers, and $I_{p,q} = diag(1_p, -1_q)$. $\| x \|_{\mathcal{E}}^2 = \langle x, x\rangle_{\mathcal{E}}$ and $d_{\mathcal{E}} : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ is defined as $d_{\mathcal{E}}^2(x,z) = \langle x-z, x-z\rangle_{\mathcal{E}}$ for arbitrary $x, z \in \mathcal{E}$. $d_{\mathcal{E}}$ is called *a pseudo-Euclidean distance*. A pseudo-Euclidean space is a set $\mathcal{E}$ equipped with d_{ε}, usually denoted by $\mathbb{R}^{(p,q)}$. Denote (p,q) as the signature of pseudo-Euclidean space $\mathcal{E}$. Unlike in a metric space, points in a pseudo-metric space need not be distinguishable; that is, one may have $d(x,z) = 0$ for $x \neq z$.

Lemma 2.1. [18] *Let $\mathcal{E}$ be a non-empty set, $d : \mathcal{E} \times \mathcal{E} \to [0,\infty)$ a pseudo-metric and $f : [0,1] \to [0,\infty]$ an additive generator. A binary function $E : \mathcal{E} \times \mathcal{E} \to [0,1]$ given by*

$$E(x,z) = f^{-1}(\min(d(x,z), f(0))) \tag{2}$$

is a fuzzy equivalence relation with respect to Archimedean t-norm T represented by (1).

Now, we introduce the concepts of kernels, positive definite (p.d.), conditionally positive definite (c.p.d.), negative definite (n.d.) kernels and indefinite (i.n.d.) kernels.

Definition 2.2. [1, 19] *Let $\mathcal{X}$ be a non-empty set and $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a real-valued and symmetric function. K is called a kernel if there exists an isometric embedding $\phi : \mathcal{X} \to \mathcal{E}$, such that $K(x,z) = \langle \phi(x), \phi(z)\rangle_{\mathcal{E}}$, where $\mathcal{E}$ is an inner product space equipped with inner product $\langle\cdot,\cdot\rangle_{\mathcal{E}}$. In particular,*

- *K is called a p.d. kernel if and only if it is p.d., that is, $\sum_{i,j=1}^n c_i c_j K(x_i, x_j) \ge 0$ for all $n \in \mathbb{N}$, $x_i \in \mathcal{X}$ and $c_i \in \mathbb{R}$.*
- *K is called a c.p.d. kernel if and only if it is c.p.d., that is, $\sum_{i,j=1}^n c_i c_j K(x_i, x_j) \ge 0$ for all $n \in \mathbb{N}$, $x_i \in \mathcal{X}$, $c_i \in \mathbb{R}$ and $\sum_{i=1}^n c_i = 0$.*
- *K is called an n.d. kernel if and only if it is n.d., that is, $\sum_{i,j=1}^n c_i c_j K(x_i, x_j) \le 0$ for all $n \in \mathbb{N}$, $x_i \in \mathcal{X}$, $c_i \in \mathbb{R}$ and $\sum_{i=1}^n c_i = 0$.*
- *K is called an i.n.d. kernel if and only if it is i.n.d., that is, if for some $x_1, x_2, \cdots, x_n \in \mathcal{X}$, $c_1, c_2, \cdots, c_n \in \mathbb{R}$ and $c_1', c_2', \cdots, c_n' \in \mathbb{R}$ exist such that $\sum_{i,j=1}^n c_i c_j K(x_i, x_j) \le 0$ and $\sum_{i,j=1}^n c_i' c_j' K(x_i, x_j) \ge 0$.*

3. Regression Problems with Fuzzy Rules.

In this study, a regression problem $\mathcal{P} = \{\mathcal{X}, \mathcal{Y}, \mathcal{S}, \mathcal{B}_{\mathcal{F}}\}$ is given as follows: $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}$ are two non-empty sets. $x = (x^1, x^2, \cdots, x^n) \in \mathcal{X}$ is the input-attribute vector; x^i is the ith input-attribute; $y \in \mathcal{Y}$ is the output-attribute. $\mathcal{S} = \{s_1, s_2, \cdots, s_m\}$ denotes the training set, where $s_i = (x_i, y_i) \in \mathcal{X} \times$

$\mathcal{Y}$ are independent and identically distributed, $x_i = (x_i^1, x_i^2, \cdots, x_i^n)$ is the ith observation value of input-attribute vector and y_i is the ith observation value output-attribute. $\mathcal{B}_\mathcal{F}$ represents the set of fuzzy rules for problem $\mathcal{P}$. We focus on a suitable approximation to the target function for problem $\mathcal{P}$ by incorporating $\mathcal{B}_\mathcal{F}$ into the SVR.

Let $\mathcal{D}_j \subset \mathbb{R}$ and $\mathcal{V} \subset \mathbb{R}$ be two given domains. Suppose that $\mathcal{F}_j$ and $\mathcal{Q}$ stand for the sets of normal fuzzy sets μ_{j,i_j} and ω_q respectively. Here, $\mathcal{F}_j = \{\mu_{j,i_j} | \mu_{j,i_j} : \mathcal{D}_j \to [0,1],\ i_j = 1, 2, \cdots, p_j\}$ for $j = 1, 2, \cdots, n$ and $\mathcal{Q} = \{\omega_q | \omega_q : \mathcal{V} \to [0,1],\ q = 1, 2, \cdots, s\}$. $\mathcal{B}_\mathcal{F}$ is a fuzzy rule base and consists of the following fuzzy rules:

$$IF\ x^1\ is\ \mu_{1,r}\ and\ \cdots\ and\ x^n\ is\ \mu_{n,r}\ THEN\ y\ is\ \omega_r,\ (r = 1, 2, \cdots, l). \tag{3}$$

We infer the antecedents of (3) by a t-norm T. Let $h_r(x) := T\left(\mu_{1,r}(x^1), \cdots, \mu_{n,r}(x^n)\right)$ and $I = \{1, 2, \cdots, l\}$. $h_r(x)$ is said to be *a fuzzy membership function of the antecedent* for $r \in I$.

This study aims at exploring the formulation of $\mathcal{B}_\mathcal{F}$ so that it can be incorporated into the SVR. On the one hand, we will construct new kernels by $\mathcal{B}_\mathcal{F}$, and then apply them to the SVR. Consequently, the SVR effectively integrates fuzzy rules based on their formulation.

4. **Kernelizing Fuzzy Rules.** In this section, we concentrate on kernelizing $\mathcal{B}_\mathcal{F}$ for $\mathcal{P}$ by two classes of fuzzy equivalence relations and fuzzy entropies.

4.1. **Constructing kernels for each fuzzy rule.** First, we will use the fuzzy equivalence relations $K_r(x,z) : \mathcal{X} \times \mathcal{X} \to [0,1]$ given by

$$K_r(x,z) = \begin{cases} T(h_r(x), h_r(z)) & \text{if } x \neq z, \\ 1 & \text{if } x = z, \end{cases} \tag{4}$$

to kernelize $\mathcal{B}_\mathcal{F}$, where $h_r(x) : \mathcal{X} \to [0,1]$ for $r \in I$.

Theorem 4.1. *For $\mathcal{P} = \{\mathcal{X}, \mathcal{S}, \mathcal{Y}, \mathcal{B}_\mathcal{F}\}$, $h_r(x)$ is a fuzzy membership function of the antecedent for $r \in I$. Then, $K_r(x,z)$ expressed by (4) is p.d. if $T(h_r(x), h_r(z))$ is p.d..*

Proof: Let $u_i := h_r(x_i)$, $T_r(u_i, u_j) := T(h_r(x_i), h_r(x_j))$ and c_i be arbitrary constants for $i = 1, 2, \cdots, m$. Since $T_r(u_i, u_j)$ is p.d., we have $\sum_{i,j=1}^m c_i T_r(u_i, u_j) c_j \geq 0$. Accordingly,

$$\begin{aligned} \sum_{i,j=1}^m c_i K_r(x_i, x_j) c_j &= \sum_{i,j=1, i\neq j}^m c_i K_r(x_i, x_j) c_j + \sum_{i=1}^m c_i^2 \\ &\geq \sum_{i,j=1, i\neq j}^m c_i T_r(u_i, u_j) c_j + \sum_{i,j=1, i=j}^m c_i T_r(u_i, u_j) c_j \geq 0. \end{aligned}$$

This implies that $K_r(x,z)$ is p.d..

We separately replace t-norm T in (4) with $T_\mathcal{P}$ and $T_\mathcal{M}$. The corresponding fuzzy equivalence relations are denoted by $K_r^\mathcal{P}$ and $K_r^\mathcal{M}$. From Theorem 4.1, we have the following results.

Corollary 4.1. *For $\mathcal{P} = \{\mathcal{X}, \mathcal{S}, \mathcal{Y}, \mathcal{B}_\mathcal{F}\}$, $h_r(x)$ is a fuzzy membership function of the antecedent for $r \in I$. The following statements hold.*
(i) $K_r^\mathcal{P}$ is p.d.; (ii) $K_r^\mathcal{M}$ is p.d..

Proof: From Theorem 4.1, it is enough to show $T_\mathcal{P}$ and $T_\mathcal{M}$ are p.d..

(i) Since $\sum_{i,j=1}^m c_i c_j T_\mathcal{P}(h_r(x_i), h_r(x_j)) = \left(\sum_{i=1}^m c_i h_r(x_i)\right)\left(\sum_{j=1}^m c_j h_r(x_j)\right) \geq 0$ for arbitrary $h_r(x)$ and $c_1, \cdots, c_m \in \mathbb{R}$, it follows that $T_\mathcal{P}$ is p.d. from Definition 2.2.

(ii) Let $u_i := h_r(x_i)$. From Definition 2.2, to prove $T_{\mathcal{M}}$ is a p.d. kernel, it is clear that we have to show $\sum_{i,j=1}^{m} c_i T_{\mathcal{M}}(u_i, u_j) c_j = C^T \mathbf{T} C \geq 0$ for any $u_1, u_2, \cdots, u_m \in [0, 1]$. Here, $C = (c_1, \cdots, c_m)^T \in \mathbb{R}^m$; $\mathbf{T} = \big(T_{\mathcal{M}}(u_i, u_j)\big)_{m\times m}$.

Without loss of generality, assume that $m \geq 2$, $u_1 \neq 0$ and $u_i \leq u_j$ for $i \leq j$. We need to prove its kth order leading principal minors $\mathbf{T}^{(k)} \geq 0$ $(2 \leq k \leq m)$. By computing $\mathbf{T}^{(k)}$, we have

$$\mathbf{T}^{(k)} = \begin{vmatrix} u_1 & u_1 & \cdots & u_1 & u_1 \\ u_1 & u_2 & \cdots & u_2 & u_2 \\ \cdots & \cdots & \cdots & \cdots & \\ u_1 & u_2 & \cdots & u_{k-1} & u_{k-1} \\ u_1 & u_2 & \cdots & u_{k-1} & u_k \end{vmatrix} = u_1 \prod_{i=1}^{k-1} (u_{i+1} - u_i). \tag{5}$$

This implies that $T_{\mathcal{M}}$ is a p.d. kernel.

Next, we construct kernels for each fuzzy rule by additive generators and pseudo-metrics according to Lemma 2.1.

Theorem 4.2. *For $\mathcal{P} = \{\mathcal{X}, \mathcal{S}, \mathcal{Y}, \mathcal{B}_{\mathcal{F}}\}$, denote $a_r(x) = (\mu_{1,r}, \mu_{2,r}, \cdots, \mu_{n,r})$ for $r \in I$. Suppose that $f : [0, 1] \to [0, \infty]$ is an additive generator and $\rho : [0, 1]^n \times [0, 1]^n \to [0, \infty)$ is a pseudo-metric. Let $K_r(x, z) : \mathcal{X} \times \mathcal{X} \to [0, 1]$ and $K_r(x, z) = f^{-1}(\min(\rho(a_r(x), a_r(z)), f(0)))$. If f satisfies*

$$f^{-1}\left(\min\{f(u) + f(v), f(0)\}\right) \geq \max\{1 - \left((1-u)^{1/2} + (1-v)^{1/2}\right)^2, 0\} \tag{6}$$

for any $u, v \in [0, 1]$, then $K_r(x, z)$ is a kernel for arbitrary pseudo-metric ρ.

Proof: Define $d_r(x, z) := \rho(a_r(x), a_r(z))$. Clearly, $d_r(x, z)$ is a pseudo-metric from $\mathcal{X} \times \mathcal{X}$ to $[0, \infty)$. From Lemma 2.1, the left of (6) is a t-norm denoted by $T(u, v)$; $K_r(x, z)$ is a fuzzy equivalence relation with respect to t-norm T. Thus, for any $x, y, z \in \mathcal{X}$, we obtain

$$K_r(x, z) \geq T(K_r(x, y), K_r(y, z)) \geq \max\{(1-(1-K_r(x, y))^{1/2}+(1-K_r(y, z))^{1/2})^2, 0\}. \tag{7}$$

Define $\Gamma =: 1 - [(1 - K_r(x, y))^{1/2} + (1 - K_r(y, z))^{1/2}]^2$. If $\Gamma < 0$, then $2[(1 - K_r(x, y))^{1/2} + (1 - K_r(y, z))^{1/2}]^2 > 2$. From $(2 - 2K_r(x, z)) \leq 2$, it follows that

$$(2 - 2K_r(x, z))^{\frac{1}{2}} < (2 - 2K_r(x, y))^{1/2} + (2 - 2K_r(y, z)\ )^{1/2}.$$

If $\Gamma \geq 0$, from (7), it follows that

$$K_r(x, z) \geq \Gamma = 1 - (1 - K_r(x, y)) - (1 - K_r(y, z)) - 2\big[(1 - K_r(x, y))(1 - K_r(y, z))\big]^{1/2}.$$

Thus, we have

$$2K_r(x, z) \geq 2 - (2 - 2K_r(x, y)) - (2 - 2K_r(y, z)) - 2\big[(2 - 2K_r(x, y))(2 - 2K_r(y, z))\big]^{1/2}.$$

That is, $(2 - 2K_r(x, z))^{1/2} \leq (2 - 2K_r(x, y))^{1/2} + (2 - 2K_r(y, z))^{1/2}$. Let $D_r(x, z) = (2 - 2K_r(x, z))^{1/2}$. It is obvious that $D_r(x, z) \leq D_r(x, y) + D_r(y, z)$. Additionally, $D_r(x, x) = (2 - 2K_r(x, x))^{1/2} = 0$ and $D_r(x, z) = D_r(z, x)$. Thus, $D_r(x, z)$ is a pseudo-metric.

On the other hand, for $\{x_i\}_{i=1}^{m} \in \mathcal{X}^m$, $K_r(x, z)$ is allowed to define the matrix $\mathbf{K} = (K_r(x_i, x_j))_{i,j=1}^{m}$. The eigen-decomposition of $\mathbf{K}$ is performed as $\mathbf{K} = \Upsilon\Lambda\Upsilon^T$, where Υ is orthogonal, Λ is diagonal starting with p positive eigenvalues followed by q negative ones as well as $m - p - q$ zeros, where $m - p - q \geq 0$. Assume that the eigenvalue λ_i of matrix $\mathbf{K}$ corresponds to the orthonormal eigenvector $v_i = (v_{i1}, v_{i2}, \cdots, v_{im})^T$. That is,

$$K_r(x_i, x_j) = \lambda_1 v_{1,i} v_{1,j} + \cdots + \lambda_p v_{p,i} v_{p,j} - \ |\lambda_{p+1}| v_{p+1,i} v_{p+1,j} - \cdots - |\lambda_{p+q}| v_{p+q,i} v_{p+q,j}$$

for any $i, j = 1, 2, \cdots, m$. Denote $\phi(x_i) = (|\lambda_1|^{1/2}v_{1,i}, \cdots, |\lambda_{p+q}|^{1/2}v_{p+q,i})^T$. It follows that

$$K_r(x_i, x_j) = \phi(x_i)^T I_{p,q}\phi(x_j) = \langle \phi(x_i), \phi(x_j)\rangle_{\mathcal{E}} \tag{8}$$

where $\mathcal{E} = \mathbb{R}^{(p,q)}$ and ϕ is a map from $\mathcal{X}$ to $\mathcal{E}$. By (8) and $K_r(x, x) = 1$, we have

$$\begin{aligned} D_r^2(x,z) &= 2 - 2K_r(x,z) = 2 - 2\langle \phi(x), \phi(z)\rangle_{\mathcal{E}} \\ &= \langle \phi(x), \phi(x)\rangle_{\mathcal{E}} + \langle \phi(z), \phi(z)\rangle_{\mathcal{E}} - 2\langle \phi(x), \phi(z)\rangle_{\mathcal{E}} \\ &= \|\phi(x) - \phi(z)\|_{\mathcal{E}}^2 = d_{\mathcal{E}}^2(\phi(x), \phi(z)). \end{aligned}$$

This means that $\phi : \mathcal{X} \to \mathbb{R}^{(p,q)}$ is an isometric embedding such that $K_r(x, z) = \langle\phi(x), \phi(z)\rangle_{\mathcal{E}}$, which completes the proof.

Theorem 4.2 shows that a class of kernels can be obtained by selecting proper additive generators for arbitrary pseudo-metrics. For example, $f_{\mathcal{P}}(s) = -\ln s$ and $f_{\mathcal{F}}(s) = \ln \frac{\lambda-1}{\lambda^s-1}$ ($\lambda \in (0, 1) \cup (1, \infty)$) are additive generators. Clearly, they satisfy the condition (6). Consequently, according to Theorem 4.2, we can construct the kernels by using these additive generators and any pseudo-metrics.

Corollary 4.2. *(i) For $f_{\mathcal{P}}(s) = -\ln s$ and any pseudo-metric $\rho : [0, 1]^n \times [0, 1]^n \to [0, \infty)$, the corresponding fuzzy equivalence relation $K_r^{\mathcal{P}}(x, z) = e^{-\rho(a_r(x),a_r(z))}$ is a kernel.*

(ii) For $f_{\mathcal{F}}(s) = \ln \frac{\lambda-1}{\lambda^s-1}$ and any pseudo-metric $\rho : [0, 1]^n \times [0, 1]^n \to [0, \infty)$, the corresponding fuzzy equivalence relation $K_r^{\mathcal{F}}(x, z) = \log_\lambda(1 + (\lambda - 1))e^{-\rho(a_r(x),a_r(z))}$ is a kernel for $\lambda \in (0, 1) \cup (1, \infty)$.

4.2. **Kernels for fuzzy rule base.** A usual way to generate new kernels from multiple kernels is to linearly combine multiple kernels. From Theorems 4.1 and 4.2, we have known that every fuzzy rule in $\mathcal{B}_{\mathcal{F}}$ corresponds to a structure of kernel $K_r(x, z)$. In order to apply $K_r(x, z)$ to the SVR, we linearly combine these kernels. Namely, $K(x, z) = \sum_{r=1}^{l} \alpha_r K_r(x, z)$, where α_r ($r \in I$) are the weights of $K_r(x, z)$.

In this study, we determine weights α_r according to the certainty degree of every fuzzy rule. The concept of fuzzy entropy means the fuzziness degree of fuzzy sets. We calculate fuzzy entropy H_r of each fuzzy membership function of the antecedent as follows [21]: $H_r = \frac{1}{m \ln 2}\sum_{i=1}^{m} s(h_r(x_i))$, where $x_i \in \mathcal{S}$ for $i = 1, 2, \cdots, m$ and

$$s(x) = \begin{cases} -x\ln x - (1-x)\ln(1-x), & x \in (0,1) \\ 0, & x = 0, 1 \end{cases}$$

is the Shannon function. Let $\alpha_r = 1 - H_r$. It follows that

$$K(x, z) = \sum_{r=1}^{l}(1 - H_r)K_r(x, z) \tag{9}$$

That is, we use the certainty degree of every fuzzy rule as the weight of $K_r(x, z)$.

From Theorems 4.1 and 4.2, for an arbitrary choice of membership functions $\mu_{1,r}(x^1)$, $\mu_{2,r}(x^2), \cdots, \mu_{n,r}(x^n)$, we obtain the kernels $K_r(x, z)$. Furthermore, the linear combinations of p.d., c.p.d. and n.d. $K_r(x, z)$ are still p.d., c.p.d., and n.d. respectively [20]. For i.n.d kernels, we are unable to know whether their linear combinations are i.n.d. For example, let

$$\rho(a_r(x), a_r(z)) = \begin{cases} \displaystyle\sum_{i=1}^{n} |a_r(x^i) - a_r(z^i)|^q, & 0 < q < 1 \\ \left[\displaystyle\sum_{i=1}^{n} |a_r(x^i) - a_r(z^i)|^q\right]^{1/q}, & q \geq 1 \end{cases}$$

and $d_{rq}(x,z) := \rho(a_r(x), a_r(z))$ for $r = 1, 2, \cdots, l$. From Corollary 4.2, we obtain a great number of kernels K_r, which are given for different parameters q in Table 1 Column 3. The corresponding kernel combinations are given in Table 1 Column 4.

Remark 4.1. *For K_r in Table 1, according to* [25] *when $0 < q \le 2$, $K^{\mathcal{P}}_{d_{rq}}(x,z) := e^{-d_{rq}(x,z)}$ is p.d.. Additionally, $1 + (\lambda - 1)e^{-d_{rq}(x,z)}$ is n.d. when $\lambda \in (0,1)$ and $0 < q \le 2$; $1 + (\lambda - 1)e^{-d_{rq}(x,z)}$ is p.d. when $\lambda \in (1,\infty)$ and $0 < q \le 2$. Furthermore, according to* [20]*, if $K : \mathcal{X} \times \mathcal{X} \to (0,\infty)$ is n.d., then $\ln K(x,z)$ is n.d.; If kernel $K : \mathcal{X} \times \mathcal{X} \to (0,\infty)$ is p.d., then $-\ln K(x,z)$ is n.d. Let $K^{\mathcal{F}}_{d_{rq}}(x,z) := \ln(1 + (\lambda - 1)e^{-d_{rq}(x,z)})$. It follows that $K^{\mathcal{F}}_{d_{rq}}(x,z)$ is n.d. for $\lambda \in (0,1)$ and $-K^{\mathcal{F}}_{r}(x,z)$ is n.d. for $\lambda \in (1,\infty)$ when $0 < q \le 2$. Thus $K^{\mathcal{F}}_{d_{rq}}(x,z)$ is c.p.d. for $\lambda \in (1,\infty)$ and $0 < q \le 2$.*

Table 1. Several kernels represented by additive generators and pseudo-metrics

Additive generators	Pseudo-metrics	K_r	Kernel combinations
$f_{\mathcal{P}}$	$d_{rq}(x,z)$ for $0 < q \le 2$	$K^{\mathcal{P}}_{d_{rq}}$ (p.d.)	$K^{\mathcal{P}}_{d_q}$ (p.d.)
	$d_{rq}(x,z)$ for $q > 2$	$K^{\mathcal{P}}_{d_{rq}}$ $(-)$	$K^{\mathcal{P}}_{d_q}$ $(-)$
$f_{\mathcal{F}}$ for $\lambda \in (0,1)$	$d_{rq}(x,z)$ for $0 < q \le 2$	$K^{\mathcal{F}}_{d_{rq}}$ (n.d.)	$K^{\mathcal{F}}_{d_q}$ (c.p.d.)
$f_{\mathcal{F}}$ for $\lambda \in (1,\infty)$	$d_{rq}(x,z)$ for $0 < q \le 2$	$K^{\mathcal{F}}_{d_{rq}}$ (p.d.)	$K^{\mathcal{F}}_{d_q}$ (p.d.)
$f_{\mathcal{F}}$ for $\lambda \in (0,1) \cup (1,\infty)$	$d_{rq}(x,z)$ for $q > 2$	$K^{\mathcal{F}}_{d_{rq}}$ $(-)$	$K^{\mathcal{F}}_{d_q}$ $(-)$

Finally, the proposed approach is summarized as follows.

- Step 1) Represent fuzzy rules $\mathcal{B}_{\mathcal{F}}$. If we lack the field knowledge, we use the fuzzy c-mean (FCM) clustering algorithm to get $\mathcal{B}_{\mathcal{F}}$. Otherwise, we represent $\mathcal{B}_{\mathcal{F}}$ by using the field knowledge.
- Step 2) Given either t-norm T or the additive generator f and pseudo-metric ρ, the kernel K_r for each fuzzy rule is represented by Theorems 3.1 or 3.2 respectively.
- Step 3) Calculate the entropy H_r for each fuzzy rule and linearly combine K_r represented by (9).
- Step 4) Apply the kernels represented by (9) to SVRs.

5. **Applications.** In this section, three experiments are performed to validate the proposed approach according to the four steps above. The following terms are used in the experiments.

- **Kernels:** The constructed kernels $K^{\mathcal{P}}$ and $K^{\mathcal{M}}$ are obtained by substituting $K^{\mathcal{P}}_r$ and $K^{\mathcal{M}}_r$ into (9) respectively. $K^{\mathcal{P}}_{d_2}$, $K^{\mathcal{P}}_{d_3}$, $K^{\mathcal{F}}_{d_2}$ and $K^{\mathcal{F}}_{d_3}$ stand for the kernels in Table 1, where the parameters $\lambda = 0.5$ in $K^{\mathcal{P}}_{d_2}$ and $K^{\mathcal{F}}_{d_2}$, $\lambda = 2$ in $K^{\mathcal{P}}_{d_3}$ and $K^{\mathcal{F}}_{d_3}$. K^{rbf} stands for Gaussian kernel used in the experiments.
- **Three regression algorithms:** SVR, fuzzy weight SVR (FWSVR) [8] and Takagi-Sugeno model (TS). Here, all the kernels mentioned above are used in SVR; K^{rbf} is used in FWSVR.
- **Parameters:** c and γ stand for the regularization parameter and Gaussian kernel parameter respectively. The parameter ε is in the ε-insensitive loss function. N_c and η represent the number of clusters and the overlap parameter in the FCM clustering algorithm [8] respectively. c and γ are chosen using a grid search method without special explanation, where $c \in \{2^{-5}, 2^{-4}, \cdots, 2^{16}\}$ and $\gamma \in \{2^{-5}, 2^{-4}, \cdots, 2^{5}\}$.

- **Two performance indices (PIs):** M_{se} and N_{sv}. Here, M_{se} denotes the minimal mean squared error on the validation set. $N_{sv} = N/m$, where N stands for the mean number of support vectors and m the number of training data. In this study, N_{sv} is given for the minimal M_{se}.

5.1. **A nonlinear dynamic system approximation.** The nonlinear dynamic system in [8] is considered:

$$\begin{aligned} y(t+1) &= \frac{y(t)y(t-1)[y(t)+2.5]}{1+y^2(t)+y^2(t-1)} + u(t) + Noise, \\ y(0) &= y(1) = 0, \ u(t) = \sin(2\pi t/50). \end{aligned} \tag{10}$$

It is shown in Figure 1 for $t \in [1, 100]$ with a Gaussian noise $[0, 0.25]$. In this experiment, 501 training points are generated from (10) with a Gaussian noise $[0, 0.25]$ for $t \in [1, 50]$. The number of test points is 1001. We use the FCM clustering algorithm to obtain $\mathcal{B}_{\mathcal{F}}$.

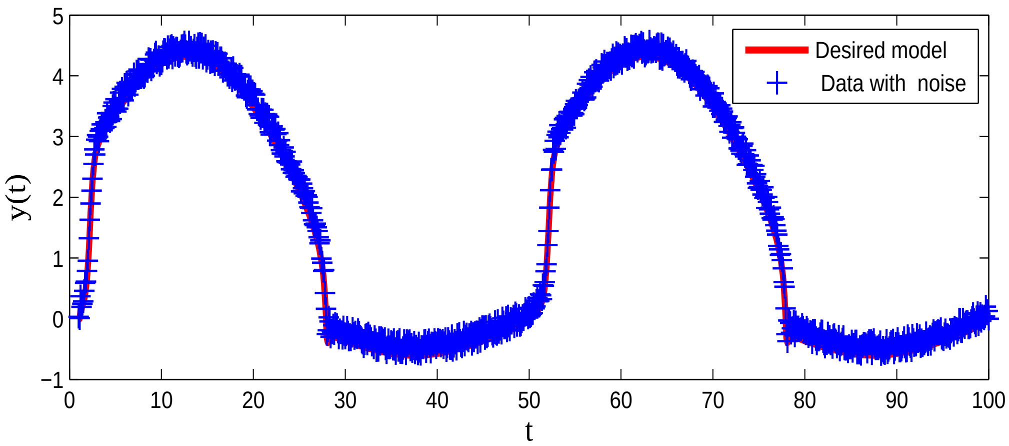

FIGURE 1. A nonlinear dynamic system with a Gaussian noise

For comparisons the proposed approach with FWSVR and TS, the membership functions in $\mathcal{B}_{\mathcal{F}}$ are the same with those in [8], where $N_c = 3, 5, 7, 9$ and $\eta = 2.5$. Additionally, the values of parameters for K^{rbf} and FWSVR are the same as those in [8] (i.e., $\sigma = 1$, $c = 1000$ and $\varepsilon = 0.2$). The numerical results are summarized in Table 2.

From Table 2, N_{sv} appears stable with the increase of N_c for SVRs based on $K^{\mathcal{P}}$, $K^{\mathcal{P}}_{d_2}$ and $K^{\mathcal{F}}_{d_2}$. However, it is not the case for SVRs based on $K^{\mathcal{M}}$, $K^{\mathcal{P}}_{d_3}$ and $K^{\mathcal{F}}_{d_3}$ when N_c is small. On the other hand, M_{se} decreases with N_c for FWSVR, TS and SVRs based on the designed kernels. Furthermore, M_{se} for our kernels are less than those for FWSVR, TS and SVR based on the Gaussian kernels when $N_c = 5, 7, 9$. Therefore, the SVR based on our kernels achieves better performance than that based on the Gaussian kernels. In particular, i.n.d kernels $K^{\mathcal{P}}_{d_3}$ and $K^{\mathcal{F}}_{d_3}$ perform better than p.d K^{rbf}.

5.2. **A gas oven model identification.** A real data set concerning an identification of a gas oven model is used [22]. The data set consists of 296 pairs of input-output pairs $(u(t), y(t))(t = 1, 2, \cdots, 296)$ shown in Figure 2 that can be downloaded from [22]. We use $(u(t), u(t-1), y(t-1), y(t-2))$ as input data and $y(t)$ as output data, where $t = 3, 4, \cdots, 296$. $\mathcal{B}_{\mathcal{F}}$ is obtained by the FCM clustering algorithm. Moreover, the values of parameters in these experiments are the same as those in [8] for SVR based on K^{rbf}

TABLE 2. Numerical results for the dynamic system approximation

N_c	PIs	$K^{\mathcal{P}}$	$K^{\mathcal{M}}$	$K^{\mathcal{P}}_{d_2}$	$K^{\mathcal{P}}_{d_3}$	$K^{\mathcal{F}}_{d_2}$	$K^{\mathcal{F}}_{d_3}$	K^{rbf}	FWSVR	TS
3	N_{sv}	5.3%	4.9%	**5.1%**	87.1%	6.1%	44.3%	11.0%	–	–
	M_{se}	0.2047	0.2263	**0.1232**	0.3419	0.1742	0.7276	0.2276	0.2832	0.6798
5	N_{sv}	24.7%	64.2%	**5.3%**	6.1%	6.3%	7.0%	–	–	–
	M_{se}	**0.1348**	0.2003	0.1551	0.1551	0.1623	0.1864	–	0.2762	0.4779
7	N_{sv}	9.2%	5.7%	4.9%	**4.9%**	5.1%	5.9%	–	–	–
	M_{se}	0.2235	0.1989	**0.1516**	0.1537	0.1536	0.1846	–	0.2645	0.2780
9	N_{sv}	6.8%	4.9%	**4.7%**	5.0%	4.8%	**4.7%**	–	–	–
	M_{se}	0.2164	0.1965	**0.1501**	0.1522	0.1529	0.1521	–	0.2583	0.2560

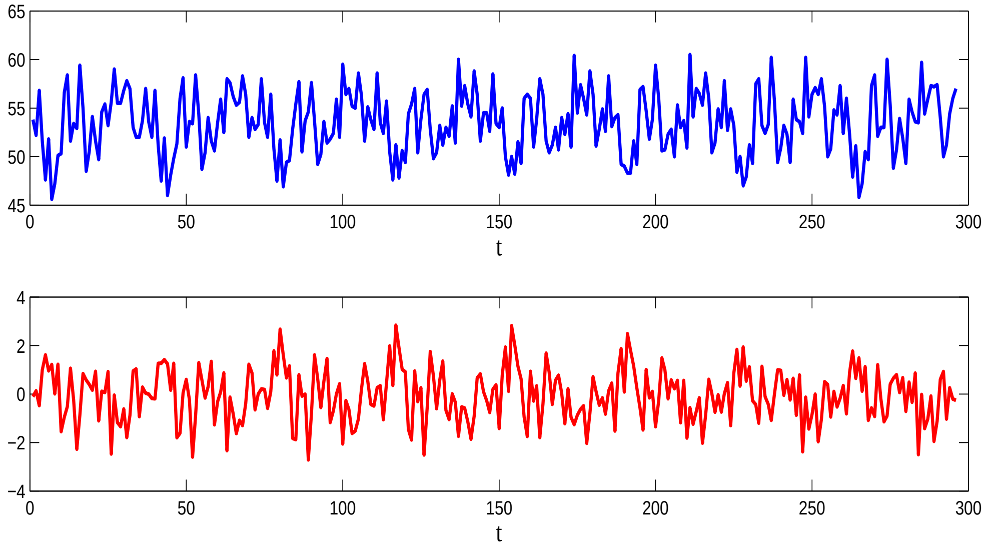

FIGURE 2. A gas oven model

and FWSVR (i.e., $\sigma = 1.5$, $c = 1000$ and $\varepsilon = 0.1$; $N_c = 4, 8$ and $\eta = 3.5$). 5-folds cross-validation is employed to evaluate the performance. The numerical results are reported in Table 3.

TABLE 3. Numerical results for the gas oven model identification

N_c	PIs	$K^{\mathcal{P}}$	$K^{\mathcal{M}}$	$K^{\mathcal{P}}_{d_2}$	$K^{\mathcal{P}}_{d_3}$	$K^{\mathcal{F}}_{d_2}$	$K^{\mathcal{F}}_{d_3}$	K^{rbf}	FWSVR	TS
4	N_{sv}	95.1%	96.2%	97.1%	96.5%	96.6%	98.3%	97.4%	–	–
	M_{se}	0.5141	0.7020	0.4959	0.5211	**0.5005**	0.6276	0.8242	0.5163	0.7576
8	N_{sv}	95.4%	96.4%	94.0%	89.5%	92.3%	93.1%	–	–	–
	M_{se}	0.4203	0.4824	0.3970	0.4534	**0.3888**	0.4799	–	0.4861	0.4976

From Table 3, all the N_{sv} are about 95 percent. M_{se} is decreasing with N_c for FWSVR, TS and SVRs based on the designed kernels. Furthermore, M_{se} for our kernels are less than those for FWSVR and TS when $N_c = 8$. Similar to the first experiment, the SVR based on our kernels achieves better performance than that based on the Gaussian kernels, FWSVR and TS.

5.3. **Forest-fire prediction.** The forest-fire data set originates from the University of California at Irvine (UCI) machine learning database, which were created by P. Cortez and A. Morais [23]. Our aim is to predict the burned area of forest fires by using these data.

5.3.1. *Data sets.* The forest-fire data set includes twelve input-attributes, one output-attributes and 517 instances. According to the related field knowledge, the input-attributes may be correlated, thus it is enough to apply some of them to predict the burned area of forest fires. In fact, the SVR tends to produce the best predictions via the given feature selection [23]. Referring to the results in [23], we perform the simulation with input-attributes (i.e., four weather variables) and output-attribute (i.e., *area*), which are listed in Table 4. As stated in [23], all entries denote fire occurrences and zero value means that areas y lower than 1ha/100 = 100m^2 was burned. Thus, the logarithm function $\psi = \ln(y+1)$ is used to improve the regression results for the right-skewed targets [24].

Let $\mathcal{X}_1 = [0, 40]$ be the interval of x_1, $\mathcal{X}_2 = [10, 100]$ the interval of x_2, $\mathcal{X}_3 = [0, 10]$ the interval of x_3, $\mathcal{X}_4 = [0, 7]$ the interval of x_4 and $\mathcal{Y} = [0, 1100]$ the interval of y. Obviously, $\mathcal{X} = \prod_{i=1}^{4} \mathcal{X}_i$.

TABLE 4. Attribute description for forest-fires data set

Attribute Information	Notation
temp – Temperature in Celsius degrees: 2.2 to 33.30	x^1
RH – Relative humidity in %: 15.0 to 100	x^2
wind – Wind speed in km/h: 0.40 to 9.40	x^3
rain – Outside rain in mm/m^2: 0.0 to 6.4	x^4
area – the burned area of the forest in ha (hectares (ha)): 0.00 to 1090.84	y

5.3.2. *Fuzzy rule formulation.* It is known that temperature, air humidity, wind speed and rain considered in this study affect fire occurrence. Based on the related field knowledge and the present data set, x_1 is described by three fuzzy sets: *high temp*, *median temp* and *low temp* whose membership functions are μ_{1,i_1} $(i_1 = 1, 2, 3)$; x_2 is described by three fuzzy sets: *high RH*, *median RH* and *low RH* whose membership functions are μ_{2,i_2} $(i_2 = 1, 2, 3)$; x_3 is described by two fuzzy sets: *high wind* and *low wind* whose membership functions are μ_{3,i_3} $(i_3 = 1, 2)$; x_4 is described by *median rain*; output-attribute y is described by three fuzzy sets: *large area*, *median area* and *small area*, whose membership functions are ω_k $(k = 1, 2, 3)$. In most cases, if *temp* is *high*, *RH* is *low*, *wind* is *large* and *rain* is *median*, then *area* is *larger*; if *temp* is *low*, *RH* is *high*, *wind* is *low* and *rain* is *median*, then *area* is *small*; if *temp* is *median*, *RH* is *high*, *wind* is *low* and *rain* is *median*, then *area* is *small*. However, we are unable to know the other cases for the lack of field knowledge based on the given data set. Thus, $\mathcal{B}_\mathcal{F}$ consists of the following 18 fuzzy rules:

$$
\begin{aligned}
&\text{If } x^1 \text{ is } \mu_{1,1} \text{ and } x^2 \text{ is } \mu_{2,3} \text{ and } x^3 \text{ is } \mu_{3,1} \text{ and } x^4 \text{ is } \mu_{4,1} \text{ then } y \text{ is } \omega_1,\\
&\text{If } x^1 \text{ is } \mu_{1,3} \text{ and } x^2 \text{ is } \mu_{2,1} \text{ and } x^3 \text{ is } \mu_{3,2} \text{ and } x^4 \text{ is } \mu_{4,1} \text{ then } y \text{ is } \omega_3,\\
&\text{If } x^1 \text{ is } \mu_{1,2} \text{ and } x^2 \text{ is } \mu_{2,1} \text{ and } x^3 \text{ is } \mu_{3,2} \text{ and } x^4 \text{ is } \mu_{4,1} \text{ then } y \text{ is } \omega_3,\\
&\text{If } x^1 \text{ is } \mu_{1,i_1} \text{ and } x^2 \text{ is } \mu_{2,i_2} \text{ and } x^3 \text{ is } \mu_{3,i_3} \text{ and } x^4 \text{ is } \mu_{4,1} \text{ then } y \text{ is unknown},
\end{aligned} \tag{11}
$$

where $(i_1,\ i_2,\ i_3) \in \{(1,3,2),(1,2,1),(1,2,2),(1,1,1),(1,1,2),(3,2,2),\ (3,3,2),(3,1,1),(3,1,3),(3,2,1),(2,1,1),(2,2,1),(2,2,2),(2,3,1),(2,3,2)\}$ and $i_4 = 1$ for the fourth fuzzy rule in (11). In general, membership functions closely depend upon the field knowledge and the statistical information of the training-samples. Their shapes are determined by the semantics of the corresponding fuzzy sets. The parameters in them are obtained by the statistical information of the data set. In this study, μ_{j,i_j} $(j = 1,\ i_1 = 1;\ j = 2,\ i_2 = 1;\ j = 3,\ i_3 = 1)$ and ω_1 are S-shaped; μ_{j,i_j} $(j = 1,\ i_1 = 2;\ j = 2,\ i_2 = 2)$ and ω_2 are Gaussian. μ_{j,i_j} $(j = 1,\ i_1 = 3;\ j = 2,\ i_2 = 3;\ j = 3,\ i_3 = 3,\ j = 4,\ i_4 = 1)$ and ω_3 are Z-shaped. For S-shaped and Z-shaped membership functions, their parameters locate the extremes of the sloped portion of the curves. These parameters are determined by the minimal and maximal values of the data. The parameters in Gaussian membership functions are determined by the standard deviation and the mean of the data.

5.3.3. *Simulation procedure and numerical results.* Parameter tuning is performed by training SVR on 2/3 of the training set and choosing the minimal M_{se} on the remaining 1/3. Set $\varepsilon = 0.5$. The corresponding results are summarized in Table 5.

TABLE 5. Numerical results for the forest-fire prediction

	$K^{\mathcal{P}}$	$K^{\mathcal{M}}$	$K^{\mathcal{P}}_{d_2}$	$K^{\mathcal{P}}_{d_3}$	$K^{\mathcal{F}}_{d_2}$	$K^{\mathcal{F}}_{d_3}$	K^{rbf}	FWSVR	TS
N_{sv}	13.07%	27.15%	18.71%	21.06%	13.62%	22.90%	38.26%	–	–
M_{se}	2.1270	2.1903	1.9477	1.8906	**1.6186**	1.6946	2.6998	2.2850	3.955

By comparing these results in Table 5 for our proposed kernels with those for K^{rbf}, M_{se} and N_{sv} for our kernels are less than those for FWSVR and SVR based on the Gaussian kernels. Therefore, the SVR based on our kernels has better generalization ability than that based on the Gaussian kernels. These prediction results are better than those in [23].

6. **Conclusion.** We have presented a method for the incorporation fuzzy rules into the SVR. The proposed method is also suitable to other kernel-based algorithms with the information in the form of fuzzy rules, e.g., classification. The construction of kernels is the key to the incorporation. Notice that some of these kernels are i.n.d.. Thus, exploring the new algorithms to solve the corresponding optimization problems is a further research direction. Additionally, p.d. and c.p.d. kernels are still attractive because of the obtained convex optimization problems for support vector algorithms. In order to avoid the nonconvex optimization problems, exploring which t-norms and additive generators can be used to construct p.d. and c.p.d. kernels by fuzzy rules is a further research direction.

Acknowledgment. The work was supported by the NSF (No. 10971043), Heilongjiang Province foundation for distinguished young scholars (No. JC200810), the National Scientific Research Innovation Foundation in Harbin Institute of Technology (HIT. NSRIF. 2009 048), and the fundamental research funds for the central universities (No. HIT. KLOE 2010047).

REFERENCES

[1] B. Schölkopf and A. J. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.
[2] B. Schölkopf, P. Simard, A. J. Smola and V. N. Vapnik, *Prior Knowledge in Support Vector Kernels*, MIT Press, Cambridge, MA, 1998.
[3] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[4] L. Jiang, M. Deng and A. Inoue, Obstacle avoidance and motion control of a two wheeled mobile robot using SVR technique, *International Journal of Innovative Computing, Information and Control*, vol.5, no.2, pp.253-262, 2009.
[5] M. Kobayashi, Y. Konishi and H. Ishigaki, A lazy learning control method using support vector regression, *International Journal of Innovative Computing, Information and Control*, vol.3, no.6(B), pp.1511-1523, 2007.
[6] F. Lauer and G. Bloch, Incorporating prior knowledge in support vector regression, *Mach. Learn.*, vol.70, no.1, pp.89-118, 2008.
[7] O. L. Mangasarian and E. W. Wild, Nonlinear knowledge in kernel approximation, *IEEE Trans. Neural Netw.*, vol.18, no.1, pp.300-306, 2007.
[8] C. C. Chuang, Fuzzy weighted support vector regression with a fuzzy partition, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol.37, no.3, pp.630-640, 2007.
[9] S. Chakraborty, M. Ghosh and B. Mallick, *Bayesian Non-Linear Regression for Large p Small n Problems*, http://www.stat.uc.edu/schakrab/svmregression.pdf, 2005.
[10] P. Sollich, Bayesian methods for support vector machines: Evidence and predictive class probabilities, *Mach. Learn.*, vol.46, no.1, pp.21-52, 2002.
[11] R. F. MacLehose and D. B. Dunson, *Nonparametric Bayes Kernel-Based Priors for Functional Data Analysis*, http://ftp.isds.duke.edu/WorkingPapers/07-23.pdf.
[12] T. Farooq, A. Guergachi and S. Krishnan, Knowledge-based Green's kernel for support vector regression, *Mathematical Problems in Engineering*, vol.2010, pp.1-17, 2010.
[13] B. Moser, On representing and generating kernels by fuzzy equivalence relations, *J. Mach. Learn. Res.*, vol.7, pp.2603-2620, 2006.
[14] C. F. Juang and D. H. Cheng, TS-fuzzy system-based support vector regression, *Fuzzy Sets Syst.*, vol.160, no.17, pp.2486-2504, 2009.
[15] I. Y. Seo, S. W. Yim, H. R. Kim and O. B. Hyun, Condition based monitoring of superconducting fault current limiter using fuzzy support vector regression, *IEEE Trans. on Appl. Supercon.*, vol.21, no.3, pp.1229-1232, 2011.
[16] E. P. Klement, R. Mesiar and E. Pap, Triangular norms. Position paper II: General constructions and parameterized families, *Fuzzy Sets Syst.*, vol.145, no.3, pp.411-438, 2004.
[17] L. Zadeh, Similarity relations and fuzzy orderings, *Inf. Sci.*, vol.3, no.2, pp.177-200, 1971.
[18] B. Baets and R. Mesiar, Metrics and T-equalities, *J. Math. Anal. Appl.*, vol.267, no.2, pp.531-547, 2002.
[19] M. Hein and O. Bousquet, Hilbertian metrics and positive definite kernels on probability measures, *Proc. of AISTATS*, pp.136-143, 2005.
[20] C. Berg, J. P. R. Christensen and P. Ressel, *Harmonic Analysis on Semigroups*, Springer-Verlag, New York, 1984.
[21] S. Al-sharhan, F. Karray, W. Gueaieb and O. Basir, Fuzzy entropy: A brief survey, *The 10th IEEE International Conference on Fuzzy Systems*, vol.3, pp.1135-1139, 2001.
[22] *http://www-personal.buseco.monash.edu.au/hyndman/TSDL/Chemistry.*
[23] P. Cortez and A. Morais, A data mining approach to predict forest fires using meteorological data, *Proc. of the 13th Portuguese Conference on Artificial Intelligence*, Guimaraes, Portugal, pp.512-523, 2007.
[24] S. Menard, *Applied Logistic Regression Analysis*, 2nd Edition, Sage Publications, Inc., 2001.
[25] B. Schölkopf, The kernel trick for distances, *Technical Report, MSR-TR-2000-51, Microsoft Research*, pp.1-9, 1993.