Kernel Functions in Takagi-Sugeno-Kang Fuzzy System with Nonsingleton Fuzzy Input

Jorge Guevara, Roberto Hirata Jr, Stephane Canu

▶ **To cite this version:**

Jorge Guevara, Roberto Hirata Jr, Stephane Canu. Kernel Functions in Takagi-Sugeno-Kang Fuzzy System with Nonsingleton Fuzzy Input. 2013 IEEE International Conference on, Jul 2013, Hyderabad, India. pp.1-8, 10.1109/FUZZ-IEEE.2013.6622409 . hal-00788559v2

HAL Id: hal-00788559

https://hal.science/hal-00788559v2

Submitted on 15 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Kernel Functions in Takagi-Sugeno-Kang Fuzzy System with Nonsingleton Fuzzy Input

Jorge Guevara
Institute of Mathematics and Statistics
University of São Paulo
Sao Paulo, Brazil
Email: jorjasso@vision.ime.usp.br

Roberto Hirata Jr.
Institute of Mathematics and Statistics
University of São Paulo
Sao Paulo, Brazil
Email: hirata@ime.usp.br

Stéphane Canu
Normandie université
INSA de Rouen - LITIS
St Etienne du Rouvray
France
Email: scanu@insa-rouen.fr

***Abstract*—Algorithms for supervised classification problems usually does not consider imprecise data, e.g., observed data whose samples can be represented by a collection of intervals, histograms, list of values, fuzzy sets among others. Fuzzy theory is a naturally choice for imprecise data. On the other hand, the state of the art techniques such a kernel methods are still a natural choice for supervised classification problems because of its robustness. Under some assumptions, Takagi-Sugeno-Kang (TSK) fuzzy system rules are equivalent to positive definite kernels (PDK) but such relationship was given considering only singleton fuzzy sets to model crisp input. Imprecise data are better modeled by membership functions of nonsingleton fuzzy sets but the relationship among nonsingleton fuzzy input, TSK fuzzy systems and PDK was up to our knowledge unknown. In this work, we study such relationship. We formulate an extension of TSK fuzzy systems to deal with nonsingleton fuzzy input and then we show that a new class of PDK are derived from it. We give three examples of nonsingleton TSK kernels which are close related to Vapnik's vicinal kernels. Also, based on those TSK induced kernels and on the concept of distance substitution kernels, we formulate two PDK for interval data. Potential applications for the proposed kernels are pattern recognition problems with imprecise data. Experiments conducted with interval datasets show better performance than state of the art approaches.**

I. INTRODUCTION

Machine learning algorithms usually work only with crisp data or with crisp representation of imprecise data. Imprecise data is due to measurement errors, missing values, noise values, or it is usually added by instruments or by some preprocessing data transformations that involve sampling, randomness or numerical imprecision.

Some real examples of imprecise data are: *(1) gene-expression data* - because of the complexity of the biological experiments, almost all causes written before are present, in special, the final quantification of the expression for each gene that is due to a statistical model of the many sources of variation; *(2) some medical data* - in the diagnosis of dyslexia, for instance, patients solve graphical tests which are subjectively evaluated by experts, in the form of linguistic terms (e.g., "*near 10*", "*between 4 and 7*"). Clinical data is prune to several problems, in special, missing values; *(3) weather databases* - climate indicators along the day are better represented by intervals, for instance, temperature between 3 and 7 degrees, CO_2 between 394.1 and 394.2 parts per million, Wind between 7 and 10 miles per hour, Humidity between 10 and 16 percent and Sea level between 2.3 and 3.5 meters. *(4) economics* - some data is better represented by intervals or histograms among others, e.g., country's income distribution, histograms of people's systolic pressure, among others, are better described by density-like distributions or fuzzy sets. Some emergent areas such as *fuzzy statistics* [1] and Symbolic Data Analysis (SDA) [2] search to deal with that kind of data.

From this perspective, machine learning algorithms able to perform some tasks using imprecise data would have a pragmatic importance. However, imprecise data is frequently preprocessed to obtain crisp values. The drawback is that we loose important information about the problem.

Fuzzy theory is the naturally choice to deal with imprecise data. On the other hand, kernel methods such as Support Vector Machines (SVM) are the state of the art methods in Machine Learning. It is worth noticing that fuzzy rule-base classifiers can be trained using SVMs, introducing some of the SVM's advantages in fuzzy classifier design [3]–[10], such as the nonlinear mapping given by positive definite kernels.

A. kernel methods and fuzzy rules

The research done in machine learning using fuzzy set theory and positive definite kernels for data analysis can be grouped in two main groups. In the first one, fuzzy theories does not involve as a necessary consequence positive definite kernels (or vice versa), some specific fuzzy techniques and positive definite kernels are used to solve particular problems. Some works in clustering [11]–[13], classification problems with outliers or noises [14], feature extraction [15], discriminant analysis [16] among others are in this group.

In the second group, fuzzy set concepts always imply positive definite kernels (or vice versa), we list some of them

(1) Kernels and fuzzy rules: under a general assumption of T-norm operators and membership functions, positive definite kernels are equivalent to interaction between the fuzzy system input and fuzzy rules [5], [17]. Papers [3], [6]–[10] use this characteristic to train fuzzy classifiers with SVM.

(2) Kernels and fuzzy basis functions: using general fuzzy implication operators in fuzzy systems, and by constructing

vectors with fuzzy basis function as components, a positive definite linear kernel is obtained by computing the inner product of those vectors [18].

(3) Kernels and fuzzy equivalence relations: positive definite kernels that maps to the unit interval with constant one in the diagonal can be seen as fuzzy equivalence relations, and vice versa under some assumptions [19], [20]. A recent work [21] describe kernels that include some prior knowledge as fuzzy equivalence relations.

B. Contributions

All the precedent works relating positive definite kernels and fuzzy rules only consider crisp inputs in the form of singletons fuzzy sets in their analysis. Because imprecision in data is better modeled by some functions rather that crisp values, i.e., nonsingleton fuzzy sets, we will consider that situation in the context of positive definite kernel methods and Takagi-Sugeno-Kang fuzzy systems.

We claim the following contributions.

- *Nonsingleton TSK fuzzy logic system*. We present an extension of classical TSK fuzzy systems to deal with nonsingleton fuzzy input.
- *Nonsingleton TSK positive definite kernels*. We show that interaction between nonsingleton TSK fuzzy rules and nonsingleton fuzzy inputs induce a class of positive definite kernel functions.
- *Kernels for interval data*. Using nonsingleton TSK fuzzy kernels, and distance substitution kernels, we give a new general view to construct positive definite kernels for interval data.

Experiments are conducted for all the proposed kernels using a Soft C-SVM [22] in datasets with samples containing missing and interval values.

II. Fuzzy Systems

Fuzzy systems are universal approximators of functions [23]. They are composed of a set of *if-then* rules called the *rule-base*, a fuzzifier and, optionally, an inference algorithm and a defuzzifier.

A. Fuzzy set, Rule base and T-norm

Let U the universe of discourse. A fuzzy set F is the set defined on U and with membership function (MF) $\mu_F : U \to [0,1]$.

A rule-base is a set of *If-Then* rules indexed by $l \in \{1,2,\dots,L\}$, of the form:

$$\textbf{If } \widetilde{u}_1 \text{ is } \widetilde{F}_1^l \text{ and}\dots\text{and } \textbf{If } \widetilde{u}_p \text{ is } \widetilde{F}_p^l \textbf{ Then } \widetilde{v} \text{ is } \widetilde{G}^l, \quad (1)$$

where $\{\widetilde{u}_j\}_{j=1}^p \cup \{\widetilde{v}\}$ are called *linguistic variables* and $\{\widetilde{F}_j^l\}_{j=1}^p \cup \widetilde{G}^l$ are their respective *linguistic values* [24].

In fuzzy systems, linguistic values are represented by fuzzy sets. Rules can be completely described by fuzzy sets. For example, the rule l given by Equation (1) can be described by the fuzzy sets $\{F_j^l\}_{j=1}^p \cup G^l$ defined in their respective universes of discourse $\{U_j\}_{j=1}^p \cup V$.

Fuzzy relations, i.e., fuzzy sets defined in the Cartesian product of several universes of discourses, are used to represent the *if* part of rules or antecedent part. The MF's of such fuzzy relations are obtained by T-norms operators [25] which are used to implement conjunction in fuzzy logic and consequently fuzzy set intersection.

A triangular norm or T-norm is the function $T : [0,1]^2 \to [0,1]$, that for all $x,y,z \in [0,1]$ satisfy:

- T1 commutativity: $T(x,y) = T(y,x)$;
- T2 associativity: $T(x,T(y,z)) = T(T(x,y),z)$;
- T3 monotonicity: $y \leq z \Rightarrow T(x,y) \leq T(x,z)$;
- T4 boundary condition $T(x,1) = x$.

Using $n \in \mathbb{N}$ and associativity, a multiple-valued extension $T_n : [0,1]^n \to [0,1]$ of a T-norm T is given by

$$T_n(x_1,x_2,\dots,x_n) = T(x_1, T_{n-1}(x_1,x_2,\dots,x_{n-1})). \quad (2)$$

We will use T to denote T or T_n.

Definition II.1 (Rule antecedent part representation by fuzzy relation and T-norm)**.** Let $\prod_{j=1}^p U_j$ the Cartesian product of universes of discourse. Let $\mathbf{x} = (x_1,x_2,\dots,x_p) \in \prod_{j=1}^p U_j$. Let $\{F_j^l \subset U_j\}_{j=1}^p$ fuzzy sets with their respective MF's $\{\mu_{F_j^l}\}_{j=1}^p$. The antecedent part of the l rule given by

$$\textbf{If } \widetilde{u}_1 \text{ is } \widetilde{F}_1^l \text{ and}\dots\text{and } \textbf{If } \widetilde{u}_p \text{ is } \widetilde{F}_p^l, \quad (3)$$

is represented by the fuzzy relation $A^l \subset \prod_{j=1}^p U_j$ with MF $\mu_{A^l} : \prod_{j=1}^p U_j \to [0,1]$ given by

$$\mu_{A^l}(\mathbf{x}) = T(\mu_{F_1^l}(x_1), \mu_{F_2^l}(x_2), \dots, \mu_{F_p^l}(x_p)), \quad (4)$$

Following Definition (II.1), fuzzy rules from Equation (1) can be characterized by fuzzy relations $R^l \subset \prod_{j=1}^p U_j \times V$ with MF $\mu_{R^l} : \prod_{j=1}^p U_j \times V \to [0,1]$ given by

$$\mu_{R^l}(\mathbf{x}, y) = I(\mu_{A^l}(\mathbf{x}), \mu_{G^l}(y)), \quad (5)$$

where I is some fuzzy implication operator (see [26]).

B. Fuzzifier and fuzzy input

Because all the operations performed in fuzzy systems are done using fuzzy sets, it is necessary to turn crisp input into fuzzy sets, this is done by the *fuzzifier*.

The fuzzifier transform the crisp input $(x_1,\dots,x_p) \in \prod_{j=1}^p U_j$ into fuzzy sets $X_1 \subset U_1, \dots, X_p \subset U_p$ with MF's $\{\mu_{X_j} : U_j \to [0,1]\}_{j=1}^p$[1].

The *singleton fuzzifier* converts a value $x_j \in \mathbb{R}$ in a *singleton fuzzy set* whose support is a single value. i.e., the membership function satisfy $\mu_{X_j}(x_j) = 1$ and $\mu_{X_j}(x) = 0$ for all $x \in \mathbb{R}$ with $x \neq x_j$.

On the other hand, the *nonsingleton fuzzifier* converts the value $x_j \in \mathbb{R}$ in a fuzzy set whose support are several values, for example, the Gaussian fuzzifier converts the point x_j in a fuzzy set with membership function $\mu_{X_j}(x) = \exp(-\gamma(x -$

[1]The fuzzifier can also be applied to transform sets into fuzzy sets or to increase the fuzziness of fuzzy sets, but we are only interested in the above definition.

$x_j)^2), \gamma \in \mathbb{R}^+$ and the triangular fuzzifier in the fuzzy set with membership function $\mu_{X_j}(x) = (1 - |x - x_j|/b_j), b_j \in \mathbb{R}^+$.

Definition II.2 (Fuzzy input)**.** Let $X_1 \subset U_1, \ldots, X_p \subset U_p$ be fuzzy sets with MF's $\{\mu_{X_j} : U_j \to [0,1]\}_{j=1}^p$ obtained after the fuzzification of $\mathbf{x} = (x_1, x_2, \ldots, x_p) \in \prod_{j=1}^p U_j$. The *fuzzy input* is the fuzzy relation $I \subset \prod_{j=1}^p U_j$ with MF $\mu_I : \prod_{j=1}^p U_j \to [0,1]$ given by

$$\mu_I(\mathbf{x}) = T(\mu_{X_1}(x_1), \mu_{X_2}(x_2), \ldots, \mu_{X_p}(x_p)). \quad (6)$$

We will call *singleton fuzzy input* if it was used the singleton fuzzifier or *nonsingleton fuzzy input* if it was used the nonsingleton fuzzifier.

C. Inference Algorithm

The third part of a fuzzy system is an inference algorithm whose arguments are the fuzzy input and the rule-base. In Mamdani fuzzy systems, the inference algorithm uses fuzzy composition between the fuzzy input I (Definition II.2) and one element of the rule base given by R^l (Equation 5) to obtain the fuzzy set Y^l with membership function

$$\mu_{Y^l}(y) = \sup_{\mathbf{x} \in \mathbb{R}^p} T(\mu_I(\mathbf{x}), \mu_{R^l}(\mathbf{x}, y)).$$

The system output is a fuzzy set given by computing $\oplus_{l=1}^L Y^l$. where $\oplus$ is a T-conorm operator [25]. Optionally, it can be used a defuzzifier to convert the output fuzzy set to a crisp value.

D. Takagi-Sugeno-Kang fuzzy system

Takagi-Sugeno-Kang (TSK) fuzzy system does not have a fuzzy set in the consequent part of their rules, it has a function of the p antecedents values instead, then, it is not necessary either fuzzy composition or defuzzification step. In first order TSK fuzzy systems each rule is written as

$$\textbf{If } \widetilde{u}_1 \text{ is } \widetilde{F}_1^l \ldots \textbf{ If } \widetilde{u}_p \text{ is } \widetilde{F}_p^l \textbf{ Then } g^l(\mathbf{x}) = c_0^l + \sum_{j=1}^p c_j^l x_j. \quad (7)$$

The input space is divided into fuzzy regions by antecedents (*if*-part) and the system behavior in those regions is described by consequents in the form of functions $g^l(\mathbf{x}) : \mathbb{R}^p \to \mathbb{R}$ given by $g^l(\mathbf{x}) = c_0^l + \langle \mathbf{x}, \mathbf{c} \rangle$, $\mathbf{x}, \mathbf{c} \in \mathbb{R}^p$, $c_0^l \in \mathbb{R}$.

Following Definition (II.1) and setting all the universe of discourses equals to $\mathbb{R}$, each rule of a TSK fuzzy system define a function $f^l : \mathbb{R}^p \to \mathbb{R}$ given by

$$f^l(\mathbf{x}) = g^l(\mathbf{x}) \mu_{A^l}(\mathbf{x}), \quad (8)$$

The output of the first order TSK fuzzy system is given by the combination of the their L rules in the following manner

$$h(\mathbf{x}) = \frac{\sum_{l=1}^L f^l(\mathbf{x})}{\sum_{l=1}^L \mu_{A^l}(\mathbf{x})}. \quad (9)$$

III. Kernels and Fuzzy Rules

A Kernel is a function k from $\mathcal{X} \times \mathcal{X}$ to $\mathbb{R}$, where $\mathcal{X}$ is an arbitrary finite set. The function k is called **positive definite** if the $n \times n$ matrix $Q_{ij} = k(x_i, x_j)$ is positive semidefinite, that is, $\mathbf{c}^\top Q \mathbf{c} \geq 0$ for any choice of $\mathbf{c} \in \mathbb{R}^n$ and any choice of $x_i \in \mathcal{X}$ (see for instance ref. [27] or [28] for details). In machine learning this property is important because $k(x_i, x_j)$ defines implicitly an inner product $\langle \Phi(x_i), \Phi(x_j) \rangle$ in a *Reproducing Kernel Hilbert Space* $\mathcal{H}_k$, using the implicit map $\Phi : \mathcal{X} \to \mathcal{H}_k$. Also, positive definite kernels are useful to construct optimization algorithms for machine learning problems, ensuring that kernel algorithms converge to a relevant solution.

A. Takagi-Sugeno kernel

By the form of consequents, there are two positive definite kernels induced from interaction between TSK fuzzy rules and fuzzy inputs.

Lemma 1 (PDFC kernel [5])**.** *If all the MF's of the fuzzy sets of the antecedent part of the l rule (Definition (II.1)) are generated by translating the positive definite functions $u : \mathbb{R} \to [0,1]$ defined as:*

$$u(x) = \begin{cases} u(-x) & \textit{if } x \neq 0 \\ 1 & \textit{if } x = 0, \end{cases}, x \in \mathbb{R}, \quad (10)$$

such that $\mu_{F_j^l}(x_j) = u_j(x_j - z_j^l)$ for some parameter $z_j^l \in \mathbb{R}$, then the kernel $k : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ given by

$$k(\mathbf{x}, \mathbf{z}^l) = \prod_{j=1}^p u_j(x_j - z_j^l) = \prod_{j=1}^p \mu_{F_j^l}(x_j), \mathbf{x}, \mathbf{z} \in \mathbb{R}^p, \quad (11)$$

is a positive definite invariant translation kernel.

The proof can be found in reference [5] (Theorem 3.11). Lemma (1) assume algebraic product as T-norm operator. By setting $g(\mathbf{x})^l = c_0$, where c_0 is some constant value. Equation (8) can be rewritten as

$$f^l(\mathbf{x}) = c_0 k(\mathbf{x}, \mathbf{z}^l). \quad (12)$$

It is worth to noting that by using Gaussian functions in Equation (10), the gaussian RBF kernel is obtained, but in general PDFC kernels are not RBF kernels [17].

Lemma 2 (First order TSK-kernel [7], [10])**.** *If all the MF's of the fuzzy sets of the antecedent part are gaussian functions, and the consequent part is the function $g^l(\mathbf{x}) = \langle \mathbf{x}, \mathbf{z} \rangle$, then the kernel $k : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ given by*

$$k(\mathbf{x}, \mathbf{z}^l) = \langle \mathbf{x}, \mathbf{z}^l \rangle \exp(-\gamma ||\mathbf{x} - \mathbf{z}^l||^2), \quad (13)$$

is positive definite.

The proof is given in [7], [10]. Using Lemma (2), Equation (8) can be rewritten as

$$f^l(\mathbf{x}) = k(\mathbf{x}, \mathbf{z}^l). \quad (14)$$

It is worth to noting that a more general case can be derived using PDFC kernels for the antecedent part and the inhomogeneous polynomial kernel in the consequent part to represent TSK fuzzy rules, supported by the following fact.

Lemma 3. *Let $\mu_{A^l}(\mathbf{x})$ in Equation (8) equals to the PDFC kernel $k(\mathbf{x}, \mathbf{z}^l)$ given in Lema (1). Let the consequent function $g^l(\mathbf{x}) = (\langle \mathbf{x}, \mathbf{z}^l \rangle) + z_0)^d$ for $d = 0, 1, \dots, n$, then the TSK fuzzy rule (Equation (8)) defines the positive definite kernel*

$$f^l(\mathbf{x}) = \hat{k}(\mathbf{x}, \mathbf{z}^l) = (\langle \mathbf{x}, \mathbf{z}^l \rangle) + z_0)^d k(\mathbf{x}, \mathbf{z}^l). \tag{15}$$

Proof: The inhomogeneous polynomial kernel is a positive definite kernel, the PDFC is positive definite. The product of two positive definite kernels is also positive definite. ■

IV. Kernels and nonsingleton fuzzy input

Positive definite kernels related to TSK rules described in the previous section only consider singleton fuzzy inputs, In this section, we define an extension of TSK fuzzy systems for nonsingleton fuzzy inputs and then we relate the resulting fuzzy rules to positive definite kernel functions.

A. TSK fuzzy system with nonsingleton input

Nonsingleton fuzzy logic system (NFLS) [29] is a Mamdani fuzzy system that accounts for the uncertainty in the input. NFLS has been applied successfully in engineering applications [29]–[33], with better results that singleton Mamdani fuzzy systems [34]. Its principal characteristic is the nonsingleton fuzzification step with the effect of noise suppression [29].

In TSK fuzzy systems only the antecedent part of the rules is represented by fuzzy relations (Definition (II.1)), because the consequent is a function. For TSK fuzzy system can to deal with nonsingleton fuzzy inputs we compute a representative value of the intersection between the fuzzy relations that represent the nonsingleton fuzzy input and rules.

Definition IV.1. Let $A^l, I \subset \prod_{j=1}^p U_j$ two fuzzy relations that represent the antecedent part of the l TSK rule and the nonsingleton fuzzy input with MF's given by Definition (II.1) and Definition (II.2) respectively. Let $g^l : \prod_{j=1}^p U_j \to \mathbb{R}$ the function that represent the consequent part of the TSK rule. The function $f^l : \prod_{j=1}^p U_j \to \mathbb{R}$ induced by A^l, I and g^l for the rule l is given by

$$f^l(\mathbf{x}) = g^l(\mathbf{x}) r(\mu_{A\cap B}(\mathbf{x})), \tag{16}$$

where $\mu_{A\cap B} : \prod_{j=1}^p U_j \to [0,1]$ is the MF of the fuzzy set $A \cap I$ given by

$$\mu_{A\cap B} = T(\mu_I(\mathbf{x}), \mu_{A^l}(\mathbf{x})), \tag{17}$$

and $r : [0,1] \to [0,1]$.

Using this definition, TSK fuzzy system can naturally extended to deal with nonsingleton fuzzy input.

Definition IV.2 (Output of nonsingleton TSK fuzzy system)**.** The output of TSK fuzzy system with nonsingleton fuzzy input is given by the function $\hat{h} : \prod_{j=1}^p U_j \to \mathbb{R}$ given by

$$\begin{aligned} \hat{h}(\mathbf{x}) &= \frac{\sum_{l=1}^L f^l(\mathbf{x})}{\sum_{l=1}^L r(\mu_{A\cap B}(\mathbf{x}))} \\ &= \frac{\sum_{l=1}^L g^l(\mathbf{x}) r(\mu_{A\cap B}(\mathbf{x}))}{\sum_{l=1}^L r(\mu_{A\cap B}(\mathbf{x}))}. \end{aligned} \tag{18}$$

In the same manner that classical TSK fuzzy systems, the output of the above formulation is a weighted combination of functions $\{g^l\}_{l=1}^M$ by values between $[0,1]$.

Lemma 4. *Let function r (Definition (IV.1)) be the* sup *function. If the fuzzifier is a singleton fuzzifier, then the output of nonsingleton TSK fuzzy system (Definition (IV.2)) reduces to output of TSK fuzzy system (Equation (9))*

Proof:

$$\begin{aligned} \hat{h}(\mathbf{x}) &= \frac{\sum_{l=1}^L g^l(\mathbf{x}) \sup_{\mathbf{x}\in \mathbf{U_1}\times \dots \mathbf{U_p}} \{T(\mu_I(\mathbf{x}), \mu_{A^l}(\mathbf{x}))\}}{\sum_{l=1}^L \sup_{\mathbf{x}\in \mathbf{U_1}\times \dots \mathbf{U_p}} \{T(\mu_I(\mathbf{x}), \mu_{A^l}(\mathbf{x}))\}} \\ &= \frac{\sum_{l=1}^L g^l(\mathbf{x}) \sup_{\mathbf{x}\in \mathbf{U_1}\times \dots \mathbf{U_p}} \{T(\mathbf{1}, \mu_{A^l}(\mathbf{x}))\}}{\sum_{l=1}^L \sup_{\mathbf{x}\in \mathbf{U_1}\times \dots \mathbf{U_p}} \{T(\mathbf{1}, \mu_{A^l}(\mathbf{x}))\}} \\ &= \frac{\sum_{l=1}^L g^l(\mathbf{x}) \mu_{A^l}(\mathbf{x})}{\sum_{l=1}^L \mu_{A^l}(\mathbf{x})} \end{aligned}$$

Then $\hat{h} = h$. ■

B. Relation between TSK rules for nonsingleton fuzzy inputs and positive definite kernels

Lemma 5 (Positivity of the nonsingleton TSK fuzzy kernel)**.** *Let $U = \prod_{j=1}^p U_j$ the Cartesian product of universes of discourse. Let the set of MF's of normal fuzzy sets.*

$$\mathcal{X} = \{\mu | \mu : U \to [0,1] \quad \text{and} \quad \exists \mathbf{x} \in U : \sup_{\mathbf{x}\in U}(\mu(\mathbf{x})) = 1\}.$$

Let $\mathcal{I}$ a nonempty and finite set of indices. Let $\mu_i, \mu_j \in \mathcal{X}$ for $i, j \in \mathcal{I}$. Then, the kernel $k : \mathcal{X} \times \mathcal{X} \to [0,1]$ given by

$$k(\mu_i, \mu_j) = \sup_{\mathbf{x}\in U} T(\mu_i(\mathbf{x}), \mu_j(\mathbf{x})), \tag{19}$$

is positive definite.

Proof: By commutativity property of T-norms k is symmetric. Noting that

$$\sum_{i,j\in\mathcal{I}} c_i c_j k(\mu_i, \mu_j) = \sum_{i\in\mathcal{I}} c_i^2 k(\mu_i, \mu_i) + 2 \sum_{i>j, i,j\in\mathcal{I}} c_i c_j k(\mu_i, \mu_j)$$

and $\sup_{\mathbf{x}\in U} T(\mu_i(\mathbf{x}), \mu_i(\mathbf{x})) = 1$, $\forall i \in \mathcal{I}$ then

$$\sum_{i,j\in\mathcal{I}} c_i c_j k(\mu_i, \mu_j) = \sum_{i\in\mathcal{I}} c_i^2 + 2 \sum_{i>j, i,j\in\mathcal{I}} c_i c_j k(\mu_i, \mu_j)$$

Using $(\sum_{i\in\mathcal{I}} c_i)^2 = \sum_{i\in\mathcal{I}} c_i^2 + 2\sum_{i>j,i,j\in\mathcal{I}} c_i c_j \geq 0$ and by the fact that $k(\mu_i, \mu_j) \in [0,1]$, we have

a) If $k(\mu_i,\mu_j)=0, \forall i,j\in\mathcal{I}: i>j$, then

$$\sum_{i,j\in\mathcal{I}} c_ic_jk(\mu_i,\mu_j)=\sum_{i\in\mathcal{I}} c_i^2\geq 0$$

b) If $k(\mu_i,\mu_j)=1, \forall i,j\in\mathcal{I}: i>j$, then

$$\begin{aligned}\sum_{i,j\in\mathcal{I}} c_ic_jk(\mu_i,\mu_j) &= \sum_{i\in\mathcal{I}} c_i^2+2\sum_{i,j\in\mathcal{I},i\neq j.i>j} c_ic_j\\ &= (\sum_{i\in\mathcal{I}} c_i)^2\geq 0\end{aligned}$$

Thus k is positive definite. ■

Lemma 6. *Function induced by nonsingleton TSK fuzzy rule (Equation* (16)*) can be written as*

$$f^l(\mathbf{x})=g^l(\mathbf{x})k(\mu_I,\mu_A^l) \tag{20}$$

Proof: Let function r (Definition (IV.1)) be the sup function, substituting $\mu_I=\mu_i, \mu_A^l=\mu_j$ in Equation (17) and by Lemma (5) completes the proof ■

By basic properties of kernels, if $g^l(\mathbf{x})=c_0$ where c_0 is some positive constant then $f^l(\mathbf{x})=c_0k(\mu_I,\mu_A^l)$ is positive definite, also if If $g^l(\mathbf{x})$ is a positive definite function then $f^l(\mathbf{x})=g^l(\mathbf{x})k(\mu_I,\mu_A^l)$ is positive definite

The above results mean that we could use whatever T-norm operator and MF's of normal fuzzy sets, we could obtain a positive definite kernel k.

Also, kernel given by Lemma (5) is a fuzzy equivalence relation [19] with respect to a given T-norm and can be represented by a fuzzy bi-implication formula [20], because every positive definite kernel that maps to the unit interval with constant one in the diagonal fullfill these requeriments. Details are omitted due to space constraints.

C. Positive definite TSK gaussian kernels

We follows with a classical result of NFLS's [29].

Lemma 7. *Let $\mu_{X_1},\mu_{X_2},\dots,\mu_{X_p}$ be the MF's of the fuzzy sets of the fuzzy input with parameters given by $(m_1,\dots,m_p)^\top\in\mathbb{R}^p$, $(\sigma_1,\dots,\sigma_p)^\top\in\mathbb{R}^p$ and let $\mu_{F_1}^l,\mu_{F_2}^l,\dots,\mu_{F_p}^l$ be the MF's of the fuzzy sets of the antecedent part of the rule l with parameters given by $(m_1^l,\dots,m_p^l)^\top\in\mathbb{R}^p$, $(\sigma_1^l,\dots,\sigma_p^l)^\top\in\mathbb{R}^p$ such that for $j=1,2\dots p$*

$$\begin{aligned}\mu_{X_j}(x_j) &= \exp\left(-0.5(x_j-m_j)^2/\sigma_j^2\right),\\ \mu_{F_j^l}(x_j) &= \exp\left(-0.5(x_j-m_j^l)^2/(\sigma_j^l)^2\right),\end{aligned}$$

Let μ_A^l and μ_I given by Definition (II.1) *and* (II.2) *respectively. If is used algebraic product as T-norm operator, then*

$$\sup_{\mathbf{x}\in\mathbb{R}^p} T(\mu_I(\mathbf{x}),\mu_{A^l}(\mathbf{x}))=\prod_{j=1}^p\exp\left(-\frac{1}{2}\frac{(m_j-m_j^l)^2}{\sigma_j^2+(\sigma_j^l)^2}\right), \tag{21}$$

The proof is in [35].

Lemma 8. *Let the set of MF's*

$$\mathcal{X}=\{\mu|\mu:\mathbb{R}^p\to[0,1] \quad \textit{and} \quad \exists\mathbf{x}\in\mathbb{R}^p: \sup_{\mathbf{x}\in\mathbb{R}^p}(\mu(\mathbf{x}))=1\}.$$

Let $\mu_i,\mu_j\in\mathcal{X}$ for $i,j\in\mathcal{I}=\{1,2,\dots,N\}$. The kernel $k:\mathcal{X}\times\mathcal{X}\to[0,1]$ given by

$$\begin{aligned}k(\mu_i,\mu_j) &= \sup_{\mathbf{x}\in U} T(\mu_i(\mathbf{x}),\mu_j(\mathbf{x}))\\ &= \prod_{j=1}^p\exp\left(-\frac{1}{2}\frac{(m_j-m_j^l)^2}{\sigma_j^2+(\sigma_j^l)^2}\right),\end{aligned} \tag{22}$$

is positive definite.

Proof: By Lemma (5), k is a positive definite kernel. ■

Lemma 9. *Let k given in Lemma* (8)*. If g^l given in Definition* (IV.1) *is a positive definite function or g^l is an affine function, then the function f^l given in Definition* (IV.1) *can be written as the positive definite kernel $k:\mathbb{R}^p\times\mathbb{R}^p\to[0,1]$ given by*

$$k(\mathbf{x},\mathbf{x}^l)=g(\mathbf{x}-\mathbf{x}^\mathbf{l})k(\mu_I,\mu_{A^l}). \tag{23}$$

Proof: Kernel g is positive by hypothesis and k through the previous lemma. The results product kernel is then positive since the product of two positive kernels is positive, then following Lemma (6) completes the proof. ■

Note that the kernel $k_\gamma:\mathcal{X}\times\mathcal{X}\to[0,1]$ given by

$$k_\gamma(\mu_I,\mu_{A^l})=\prod_{j=1}^p\exp\left(-\frac{1}{2}\frac{(m_j-m_j^l)^2}{\sigma_j^2+(\sigma_j^l)^2+\gamma}\right), \tag{24}$$

is also positive definite by the change of variables $\sigma'^2=\sigma^2+\gamma$.

Parameter γ plays the same regularization role as it does in the RBF kernel. Kernels given by Equations (22) and (24) are closed related to Gaussian vicinal kernels defined by Vapnik in [28]. These kernels can be viewed as vicinal kernels where the vicinities are defined by the volume given by the spread of each feature.

D. kernels for interval data

Let's assume that imprecise data is given as the interval $[\underline{\mathbf{x}},\overline{\mathbf{x}}]=[(\underline{x}_1,\underline{x}_2,\dots,\underline{x}_p)^\top,(\overline{x}_1,\overline{x}_2,\dots,\overline{x}_p)^\top]\subset\mathbb{R}^p$, where each feature $\{[\underline{x}_j,\overline{x}_j]\}_{j=1}^p$ is a closed interval in $\mathbb{R}$ with $\{\underline{x}_j\leq\overline{x}_j\}_{j=1}^p$. In this context, all the induced TSK kernels with nonsingleton fuzzy input can be used, e.g., it is possible to construct a fuzzy data set by setting one fuzzy set X_j with MF μ_{X_j} for each interval $[\underline{x}_j,\overline{x}_j]$.

It is possible to derive new positive definite kernels for fuzzy data using the concept of *distance substitution* kernels [36], for example the *distance substitution Gaussian kernel*:

$$k_d^{rbf}(\mathbf{x},\mathbf{x}')=\exp(-\gamma d(\mathbf{x},\mathbf{x}')^2),\gamma\in\mathbb{R}^+,\mathbf{x},\mathbf{x}'\in\mathcal{X} \tag{25}$$

only needs to construct a specific distance measure.

Lemma 10. *Let $\mu=(\mu_{X_1},\mu_{X_2},\dots,\mu_{X_p})^\top$ and $\mu'=(\mu'_{X_1},\mu'_{X_2},\dots,\mu'_{X_p})^\top$ two vectors of MF's with parameters: $(m_1,\dots,m_p)^\top\in\mathbb{R}^p$, $(\sigma_1,\dots,\sigma_p)^\top\in\mathbb{R}^p$ and $(m'_1,\dots,m'_p)^\top\in\mathbb{R}^p$, $(\sigma'_1,\dots,\sigma'_p)^\top\in\mathbb{R}^p$ respectively,* *constructed for the intervals* $[\underline{\mathbf{x}}, \overline{\mathbf{x}}] \subset \mathbb{R}^p$ *and* $[\underline{\mathbf{x}}', \overline{\mathbf{x}}'] \subset \mathbb{R}^p$ *respectively. Functions*

$$d_1(\mu, \mu') = \frac{\prod_{j=1}^{p}(m_j - m'_j)^2}{\sum_j^p \sigma_j^2 + \sum_j^p (\sigma'_j)^2}$$

$$d_2(\mu, \mu') = \frac{\prod_{j=1}^{p}(m_j - m'_j)^2}{\prod(\sigma_j^2 + (\sigma'_j)^2)}$$

have zero diagonal and are pseudometrics on $\mathcal{X}$.

Proof: d_1 and d_2 are symmetric and have zero diagonal by construction. They are positive as the product of positive function and subadditive. Thus both functions are pseudometrics distances since they can be set to zero for some distinct variances values. ■

Lemma 11. *The kernels*

$$k_\gamma(\mu, \mu') = \exp(-\gamma d_1(\mu, \mu')^2), \gamma \in \mathbb{R}^+ \quad (26)$$

$$k_\gamma(\mu, \mu') = \exp(-\gamma d_2(\mu, \mu')^2), \gamma \in \mathbb{R}^+ \quad (27)$$

are positive definite

Proof: By definition of substitution kernel both kernels are positive definite ■

Of course, many other kernels can be derived using the this concept and this is question of further research.

V. Experiments

A. Data and Implementation

Four low quality datasets from the KEEL-dataset repository [37] were used. *Low quality data* [38] refers to data with uncertainty about the actual value of a feature. These KEEL datasets contain samples with missing values and interval features. Table I contains the summary of these datasets.

TABLE I
Summary of low quality datasets

Dataset	Samples	Classes	Features	Missing Values
Long-4	25	2	4	No
100mlI-4	52	2	4	No
100mlP-4	52	2	4	No
dyslexic-12-4	65	4	12	Yes

1) Description of the data:

- *Dataset Long-4* is an athletic performance dataset, it contains 4 features. The two first features are determined by a coach in form of linguistic values, intervals or numbers; the other two indicators are measured three times producing uncertainty information represent as intervals.
- *Dataset 100mlI-4* is an athletic performance dataset used to classify whether or not a mark in a 100 meters race is being achieved. The features are given by the weight to height ratio, the reaction time, the starting speed, and 40m. speed. Measurements were obtained by three different observers.
- *Dataset 100mlP-4* is the same as *Dataset 100mlI-4* but measurements were obtained by a subjective judgment of the coach, in linguistic terms as " reaction time is low".
- *Dataset Dyslexic-12-4* is a dataset with twelve features and four classes { dyslexia, no dyslexia, control, other} that contain missing values.

Table II shows two examples of *Long-4* dataset where each feature is an interval. A detailed description of these datasets can be found in [38], [39].

TABLE II
Two samples of the low quality dataset *Long-4*

x_1	x_2	x_3	x_4	y
$[8.7, 10.1]$	$[45, 47]$	$[2, 2.15]$	$[5, 5.1]$	$\{1\}$
$[9.5, 10]$	$[60, 64]$	$[2.21, 2.23]$	$[5.33, 5.4]$	$\{0, 1\}$

2) Scaling and fuzzification: Some samples of these datasets belong to two different classes at the same time. Because it is not the purpose of this work, they were removed. Next, intervals were scaled (by a linear transformation) to be in the unit square.

For each (feature) interval $[\underline{x}_j, \overline{x}_j]$ of the four datasets, a fuzzy set with Gaussian MF μ_{X_j} with $m_j = (\underline{x}_j + \overline{x}_j)/2$ and $\sigma_j^i = (\underline{x}_j - \overline{x}_j)/(2\sqrt{2}\ln(2))$ has been built[2]. That is, for each sample $[\underline{\mathbf{x}}, \overline{\mathbf{x}}] = [(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_p)^\top, (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_p)^\top] \subset \mathbb{R}^p$, we have the fuzzy sets $X_1, \ldots, X_p$ with MF's $\mu_{X_1}, \ldots, \mu_{X_p}$ with the following parameter vectors $(m_1, m_2, \ldots, m_p)^\top$ and $(\sigma_1, \sigma_2, \ldots, \sigma_p)^\top$ obtained directly from the data.

In the case of missing values, we constructed an interval that spans the whole range of the variable.

B. Kernels Setting

Were used the kernels given by Equations (22), (23), (24), (26) and (27). For comparison purposes, it was set the RBF kernel as baseline.

Table III show three different setting for experiments. The first one consider model selection (see Section V-B2 for details) over the γ parameter. Kernels with $id \in \{0, 1, 2, 3, 4\}$ are in this group.

The second group is composed by one kernel $id = 5$ and it does not have model selection.

The last one group is composed of kernels $id \in \{6, 7, 8, 9\}$ and experiments in these kernels consider model selection over γ and multiple kernel setting for fuzzy rules.

1) Multiple kernel setting for fuzzy rules: By considering each fuzzy rule as combination of two (or more) rules and, using positive definite kernels to represent it. The resulting kernel is given by $k(\mathbf{x}, \mathbf{x}') = \sum_r \beta_r k_s(\mathbf{x}, \mathbf{x}')$, for $\beta_r > 0$, and $\sum_r \beta_r = 1$. That permits to use more general approach as the multiple kernel learning [40] to optimize the β_r parameters. For the third group of kernels $id = \{6, 7, 8, 9\}$ from Table III it was set arbitrarily $\beta = 0.25$.

Kernel arguments in Table III denoted by $\mathbf{x}$ and $\mathbf{x}'$ correspond to crisp values, we set those values to be the means of the intervals, i.e., $\mathbf{x} = (m_1, m_2, \ldots, m_p)^\top$.

[2]This measurement for σ parameter is called the *full width at half maximum.*

TABLE III
GAUSSIAN FUZZY KERNELS FOR INTERVAL DATA

Model selection over γ	
id	**kernel**
0	$k_\gamma^{\text{rbf}}(\mathbf{x}, \mathbf{x}')$
1	$k_\gamma^{\text{eq(24)}}(\mu_X, \mu_{X'})$
2[a]	$k^{\text{eq(23)}}(\mathbf{x}, \mathbf{x}')$
3[b]	$\beta k_\gamma^{\text{eq(26)}}(\mu_X, \mu_{X'})$
4[b]	$\beta k_\gamma^{\text{eq(27)}}(\mu_X, \mu_{X'})$
No model selection	
5[c]	$k^{\text{eq(23)}}(\mathbf{x}, \mathbf{x}')$
Multiple kernel setting and model selection over γ	
6	$\beta k^{\text{eq(22)}}(\mu_X, \mu_{X'}) + (1-\beta) k_\gamma^{\text{rbf}}(\mathbf{x}, \mathbf{x}')$
7[a]	$\beta k^{\text{eq(23)}}(\mathbf{x}, \mathbf{x}') + (1-\beta) k_\gamma^{\text{rbf}}(\mathbf{x}, \mathbf{x}')$
8[b]	$\beta k_{\gamma'}^{\text{eq(26)}}(\mu_X, \mu_{X'}) + (1-\beta) k_\gamma^{\text{rbf}}(\mathbf{x}, \mathbf{x}')$
9[b]	$\beta k_{\gamma'}^{\text{eq(27)}}(\mu_X, \mu_{X'}) + (1-\beta) k_\gamma^{\text{rbf}}(\mathbf{x}, \mathbf{x}')$

[a] Function g was setting to the rbf kernel.
[b] The γ' parameter in $k^{\text{eq(26)}}$ was 1.
[c] Function g was setting to the one-degree polinomial kernel.

TABLE IV
RESULTS EXPERIMENTS I

Dataset Long-4				
Kernel	Acc.	svs	C	γ
0	73.33	15.1	2^0	2^{-3}
1	73.33	15.1	2^0	2^2
2	43.33	16.2	2^{-5}	2^5
3	**85**	13.2	2^1	2^{-3}
4	**73.33**	9.8	$2^1 1$	2^-7
5	46.67	16.2	2^3	2^5
6	68.33	16.2	2^2	2^{-7}
7	68.33	16.2	2^2	2^{-7}
8	**76.67**	15	2^1	2^{-6}
9	**78.33**	16	2^1	2^4
Dataset 100mlI-4				
Kernel	Acc.	svs	C	γ
0	96	11.8	2^{15}	2^{-4}
1	**98**	31	2^1	2^{-5}
2	**98**	37.1	2^2	2^{-4}
3	58.33	41.4	2^{-2}	2^5
4	78.67	38.1	2^{-1}	2^5
5	**98**	41.5	2^0	2^5
6	**98**	37.1	2^3	2^{-8}
7	**98**	37.1	2^3	2^{-8}
8	96	12.7	2^{16}	2^2
9	94.33	29.5	2^2	2^{32}
Dataset 100mlP-4				
Kernel	Acc.	svs	C	γ
0	88	28.6	2^3	2^2
1	**88**	**18.9**	2^4	2^{-2}
2	84.67	42.4	2^1	2^{-1}
3	65.17	16.2	2^{11}	2^{-5}
4	83.67	16.4	2^{15}	2^5
5	82.67	42.3	2^3	2^5
6	88	37.2	2^4	2^{-3}
7	88	37.3	2^4	2^{-3}
8	**88**	**26.4**	$2^1 6$	2^4
9	**88**	**26**	2^2	2^8
Dataset Dyslexic-12-4				
Kernel	Acc.	svs	C	γ
0	36	33	2^5	2^{-8}
1	36	33.3	2^3	2^5
2	21.119	45	2^{-5}	2^5
3	21.12	40.5	2^{-5}	2^5
4	33.12	37	2^2	2^5
5	**38.45**	45	2^1	2^5
6	36	38.1	2^1	2^{-4}
7	36	38.1	2^1	2^{-4}
8	36	33.8	2^2	2^{-4}
9	**44.10**	30.6	$2^1 6$	2^{-6}

2) Model Selection: It was performed model selection in the kernel parameter γ with values $\{2^4, \ldots, 2^{-10}\}$ and in the parameter C of soft C-SVM with values $\{2^{-1}, \ldots, 2^{14}\}$ using grid search. For each point in the grid, it was performed 10-fold cross-validation using partitions from *KEEL-dataset repository* [37]. Finally, we reported the pair (C, γ) with best cross validation accuracy.

C. Results

Table IV shows the cross validation accuracy, number of support vectors and the best parameters C and γ from the model selection step. A soft C-SVM [22] were used to test all the proposed kernels. The Results show that using the proposed kernels,in most of the cases, outperform the RBF kernel for those datasets. Results in the datasets differ, showing that the kernel choice is a crucial part of the supervised classification problem.

VI. CONCLUSION

Real world applications of supervised classifiers involve imprecise data. This is due to the fact that no measurement instrument can produce an exact result, usually data is gathered by subjective opinions, and is better expressed by linguistic, interval or fuzzy values. In kernels methods, the choice of a kernel function is crucial because it constitutes prior knowledge about a task. No free lunch theorem states that learning algorithms with better generalisation capabilities are obtained by using some prior information of the domain.

In this work, we show that the relationship among TSK fuzzy systems, nonsingleton fuzzy input and positive definite kernels a new class of positive definite kernels can be derived. Also we give a formulation to construct kernels for interval data using the concept of distance substitution kernels. Experiments performed in interval datasets show promising results.

Acknowledgment

The authors would like to thank CNPq and CAPES (Brazilian Agencies) for their financial support.

References

[1] J. Buckley, *Fuzzy statistics*. Springer, 2004, vol. 149.

[2] L. Billard and E. Diday, *Symbolic data analysis: conceptual statistics and data mining*. Wiley, 2012, vol. 654.

[3] J.-T. Jeng and T.-T. Lee, "Support vector machines for the fuzzy neural networks," in *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on*, vol. 6, 1999, pp. 115 –120 vol.6.

[4] S. Sohn and C. Dagli, "Advantages of using fuzzy class memberships in self-organizing map and support vector machines," in *International Joint Conference on Neural Networks, 2001. Proceedings. IJCNN '01.*, vol. 3, 2001, pp. 1886 –1890 vol.3.

[5] Y. Chen and J. Wang, "Support vector learning for fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 6, pp. 716 – 728, dec. 2003.

[6] C.-T. Lin, C.-M. Yeh, S.-F. Liang, J.-F. Chung, and N. Kumar, "Support-vector-based fuzzy neural network for pattern classification," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 1, pp. 31 – 41, feb. 2006.

[7] C.-F. Juang, S.-H. Chiu, and S.-W. Chang, "A self-organizing ts-type fuzzy network with support vector learning and its application to classification problems," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 5, pp. 998 –1008, oct. 2007.

[8] C. T. Wee and T. W. Wan, "Efsvmrobus-fcm: Evolutionary fuzzy rule-based support vector machines classifier with fcm clustering," in *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, june 2008, pp. 606 –612.

[9] S.-M. Zhou and J. Gan, "Constructing l2-svm-based fuzzy classifiers in high-dimensional space with automatic model selection and fuzzy rule ranking," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 3, pp. 398 –409, june 2007.

[10] W.-Y. Cheng and C.-F. Juang, "An incremental support vector machine-trained ts-type fuzzy system for online classification problems," *Fuzzy Sets Syst.*, vol. 163, pp. 24–44, January 2011.

[11] Y. Kanzawa, Y. Endo, and S. Miyamoto, "On kernel fuzzy c-means for data with tolerance using explicit mapping for kernel data analysis," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–6.

[12] H.-C. Huang, Y.-Y. Chuang, and C.-S. Chen, "Multiple kernel fuzzy clustering," *Fuzzy Systems, IEEE Transactions on*, vol. 20, no. 1, pp. 120 –134, feb. 2012.

[13] M. Raza and F.-H. Rhee, "Interval type-2 approach to kernel possibilistic c-means clustering," in *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, june 2012, pp. 1 –7.

[14] X. Yang, G. Zhang, J. Lu, and J. Ma, "A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises," *Fuzzy Systems, IEEE Transactions on*, vol. 19, no. 1, pp. 105 –115, feb. 2011.

[15] J. Wang, J. Hua, and J. Guo, "Fuzzy maximum scatter discriminant analysis with kernel methods," in *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on*, vol. 2, aug. 2010, pp. 560 –564.

[16] G. Heo and P. Gader, "Robust kernel discriminant analysis using fuzzy memberships," *Pattern Recogn.*, vol. 44, no. 3, pp. 716–723, Mar. 2011.

[17] Y. Chen, "Support vector machines and fuzzy systems," in *Soft Computing for Knowledge Discovery and Data Mining*, O. Maimon and L. Rokach, Eds. Springer US, 2008, pp. 205–223.

[18] J. M. Lski, "On support vector regression machines with linguistic interpretation of the kernel matrix," *Fuzzy Sets Syst.*, vol. 157, no. 8, pp. 1092–1113, Apr. 2006.

[19] B. Moser, "On the t-transitivity of kernels," *Fuzzy Sets and Systems*, vol. 157, no. 13, pp. 1787 – 1796, 2006.

[20] ——, "On representing and generating kernels by fuzzy equivalence relations," *J. Mach. Learn. Res.*, vol. 7, pp. 2603–2620, Dec. 2006.

[21] F. Liu and X. Xue, "Design of natural classification kernels using prior knowledge," *Trans. Fuz Sys.*, vol. 20, no. 1, pp. 135–152, Feb. 2012. [Online]. Available: http://dx.doi.org/10.1109/TFUZZ.2011.2170428

[22] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[23] B. Kosko, "Fuzzy systems as universal approximators," *Computers, IEEE Transactions on*, vol. 43, no. 11, pp. 1329 –1333, nov 1994.

[24] L. A. Zadeh, "The concept of a linguistic variable and its applicationto approximate reasoning - ii," *Inf. Sci.*, vol. 8, no. 4, pp. 301–357, 1975.

[25] E. P. Klement, R. Mesiar, and E. Pap, "Triangular norms. position paper i: basic analytical and algebraic properties," *Fuzzy Sets and Systems*, vol. 143, no. 1, pp. 5 – 26, 2004.

[26] Y. Shi, B. V. Gasse, D. Ruan, and E. Kerre, "On dependencies and independencies of fuzzy implication axioms," *Fuzzy Sets and Systems*, vol. 161, no. 10, pp. 1388 – 1405, 2010.

[27] B. Schölkopf and A. J. Smola, *Learning with kernels : support vector machines, regularization, optimization, and beyond*, ser. Adaptive computation and machine learning. MIT Press, 2002.

[28] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[29] G. Mouzouris and J. Mendel, "Nonsingleton fuzzy logic systems: theory and application," *IEEE Transactions on Fuzzy Systems*, vol. 5, no. 1, pp. 56 –71, Feb. 1997.

[30] A. Simon and G. T. Flowers, "Non-singleton fuzzy sets for disturbance attenuation," *International Journal*, no. September, pp. 36 849–36 849, 2007.

[31] Z. Hakan Akpolat, "Non-singleton fuzzy logic control of a dc motor," *Journal of Applied Sciences*, vol. 5, pp. 887–891, 2005.

[32] T. W. Chua and W. W. Tan, "Non-singleton genetic fuzzy logic system for arrhythmias classification," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 2, pp. 251–259, 2011.

[33] D. Kim, S.-h. Huh, and G.-t. Park, "Modeling corrupted time series data via nonsingleton fuzzy logic system 2 nonsingleton fuzzy logic system," *Time*, pp. 1298–1303, 2004.

[34] A. Cara, I. Rojas, H. Pomares, C. Wagner, and H. Hagras, "On comparing non-singleton type-1 and singleton type-2 fuzzy controllers for a nonlinear servo system," in *Advances in Type-2 Fuzzy Logic Systems (T2FUZZ), 2011 IEEE Symposium on*, april 2011, pp. 126 – 133.

[35] J. M. Mendel, *Uncertain rule-based fuzzy logic system: introduction and new directions*. Prentice–Hall PTR, 2001.

[36] B. Haasdonk and C. Bahlmann, "Learning with distance substitution kernels," *Pattern Recognition*, pp. 220–227, 2004.

[37] J. Alcal-Fdez, A. Fernndez, J. Luengo, J. Derrac, and S. Garca, "Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework." *Multiple-Valued Logic and Soft Computing*, vol. 17, no. 2-3, pp. 255–287, 2011.

[38] A. Palacios, L. Sánchez, and I. Couso, "Extending a simple genetic cooperative-competitive learning fuzzy classifier to low quality datasets," *Evolutionary Intelligence*, vol. 2, no. 1-2, pp. 73–84, 2009.

[39] A. M. Palacios, L. Sánchez, and I. Couso, "Future performance modeling in athletism with low quality data-based genetic fuzzy systems," *Multiple-Valued Logic and Soft Computing*, vol. 17, no. 2-3, pp. 207–228, 2011.

[40] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, Dec. 2004.