Positive Definite Kernel Functions on Fuzzy Sets

Jorge Guevara
Instituto de Matemática
e Estatística
Universidade de São Paulo
Sao Paulo, Brazil
Email: jorjasso@vision.ime.usp.br

Roberto Hirata Jr.
Instituto de Matemática
e Estatística
Universidade de 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—Embedding non-vectorial data into a vector space is very common in machine learning, aiming to perform tasks such as classification, regression or clustering. Fuzzy datasets or datasets whose observations are fuzzy sets, are examples of non-vectorial data and, several fuzzy pattern recognition algorithms analyze them in the space formed by the set of fuzzy sets. However, the analysis of fuzzy data in such space has the limitation of not being a vector space. To overcome such limitation, we propose the embedding of fuzzy data into a proper Hilbert space of functions called the Reproducing Kernel Hilbert Space (RKHS). This embedding is possible by using a positive definite kernel function on fuzzy sets. We present a formulation of a real-valued kernels on fuzzy sets, in particular, we define the intersection kernel and the cross product kernel on fuzzy sets giving some examples of them using T-norm operators. Also, we analyze the nonsingleton TSK fuzzy kernel and, finally, we give some examples of kernels on fuzzy sets that can be easily constructed from the previous ones.

Index Terms—Kernel on fuzzy sets, Reproducing Kernel Hilbert Space, positive definite kernel

I. INTRODUCTION

Several machine learning or pattern recognition applications contain datasets whose observations are fuzzy sets, i.e., datasets of the form $\{X_i\}_{i=1}^{N}$, where each X_i is a fuzzy set [1], [2]. Those datasets are a result of modeling impreciseness and vagueness in observations of real problems with fuzzy sets. For example, because of the uncertainty added by noise, or impreciseness due to measurement instruments, data from biological and astronomical problems can be modeled by fuzzy sets. Also, datasets with features given in the form of linguistic terms, words and intervals can be modeled by fuzzy sets [3]–[9].

In Machine Learning (ML), datasets are used to automatically construct models and algorithms that give some useful information to the user, for instance, to make future predictions from the actual data, to perform feature selection and other important tasks such as: clustering, regression, inference, density estimation [10], [11].

A commonly used method in ML is to perform an embedding of data into a proper Hilbert space of functions or *Reproducing Kernel Hilbert Space* (RKHS) [12]–[15]. To make this embedding possible, it is necessary to have a real-valued positive definite function called the *reproducing kernel*[1]. Methods

[1]The word *kernel* comes from the theory of integral operators and it should not be confused with the concept of kernel of a fuzzy set.

working in this way are called *kernel methods* [13], [14], for instance, the Support Vector Machine [16], Support Vector Data Description [17], Kernel PCA [13], Gaussian Process [18].

Kernel methods are attractive for data analysis because: 1) their domain of definition has not additionally requirements, allowing the analysis of non-vectorial data, such as graphs, sets, strings; 2) in a RKHS, the closeness of two functions implies closeness in their values; 3) to construct a RKHS it is only necessary to have a real-valued positive definite kernel $k : E \times E \to \mathbb{R}$; 4) computations in the RKHS are performed by knowing that kernel evaluations are equal to inner product of functions in the RKHS: $k(x,y) = \langle k(x,.), k(y,.) \rangle_{\mathcal{H}}$, where $x \in E \mapsto k(x,.) \in \mathcal{H}$ is the embedding of the data into the RHKS $\mathcal{H}$, and $k(x,.), k(y,.) \in \mathcal{H}$ are the representers of $x, y \in E$ in the RKHS; 5) $k(x,y)$ is a similarity measure between the objects $x, y \in E$ and, because the mapping $x \in E \mapsto k(x,.) \in \mathcal{H}$ is nonlinear, simple functions in the RKHS are useful to analyze complex input data; 6) kernel methods are modular, i.e., algorithms working in the RKHS are independent of the kernel that generates such space, meaning that, we can choose one among many kernels without changing the algorithm; 7) Many classical algorithms can be *kernelized* through the *kernel trick* [13].
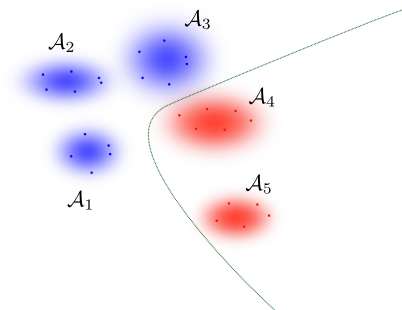

Fig. 1. Supervised classification of fuzzy data using support vector machines. $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_5$ are fuzzy sets [19]

In this paper, we give the theoretical basis to construct positive definite kernel on fuzzy sets, that is, we consider the set E as the set of all the fuzzy sets. This will allow us to use all the tools of kernel methods in datasets whose observations are given by fuzzy sets. As an example, Figure 1 shows a

nonlinear classifier obtained from a support vector machine using a dataset whose observations are fuzzy sets, using a positive definite kernel on fuzzy sets [19].

The literature reports some works using jointly fuzzy theory techniques and positive definite kernels to solve machine learning problems, particularly, in clustering [20]–[22], classification problems with outliers or noises [23], feature extraction [24] and discriminant analysis [25], without implying positive definite kernels on fuzzy sets, i.e., all the kernels are real valued functions defined on $\mathbb{R}^D \times \mathbb{R}^D$ ($\mathbb{R}$ is the set of real numbers and D is a positive integer) and fuzzy techniques and kernels are used in some steps of the algorithms.

A relationship between some fuzzy concepts and positive definite kernels, as for example, Takagi-Sugeno-Kang fuzzy systems, under some criteria, can be viewed as kernel evaluations [26]–[33]; some fuzzy basis functions can be used to construct positive definite kernels [34] and some positive definite kernels are fuzzy equivalence relations [35]–[37].

To the best of our knowledge, the first attempt to fill this gap is the work [19] giving a formulation to construct positive definite kernels on fuzzy sets and performing some experiments with those kernels using fuzzy and interval datasets.

To the best of our knowledge, there is no general formulation to define kernels on fuzzy sets. All previous works only consider kernels on $\mathbb{R}^D \times \mathbb{R}^D$ relating fuzzy concepts in the design of the kernel or as a step of some algorithm. This work has the following contributions:

- We give a general formulation of kernels on fuzzy sets, in particular, we define the intersection kernel and the cross product kernel on fuzzy sets. Also, we provide some examples of such kernels using different T-norm operators.
- We show that the kernel presented in [19] satisfy our definition of kernel on fuzzy sets and we prove that such kernel is a fuzzy equivalence relation and a fuzzy logic formula for fuzzy rules.
- We give several examples of new positive definite kernels on fuzzy sets; we present the Fuzzy Polynomial Kernel, the Fuzzy Gaussian Kernel and the Fuzzy Rational Quadratic Kernel and also, we present some conditionally positive kernels on fuzzy sets, such as: the Fuzzy Multiquadric Kernel and the Fuzzy Inverse Multiquadric Kernel.

II. THEORETICAL BACKGROUND

A. Reproducing Kernel Hilbert Spaces

A Reproducing Kernel Hilbert Space (RKHS) is a Hilbert space of functions with the nice property that the closeness of two functions, in the sense that the norm of their difference is small, implies closeness of their values. Such Hilbert spaces are widely used in machine learning for data analysis. Algorithms such as support vector machines [38], support vector data description [17] and kernel PCA [13] work with the embedding of the input data into some RKHS generated by a reproducing kernel.

In the sequel, E denotes a non empty set, $\mathcal{H}$ is the real RKHS of real valued functions on E. Notation $k(.,y)$ means the mapping $x \to k(x,y)$ with fixed y where k is a function on $E \times E$.

Definition II.1 (Reproducing kernel). A function[2]

$$
\begin{aligned}
k : E \times E &\to \mathbb{R} \\
(x,y) &\mapsto k(x,t)
\end{aligned}
\tag{1}
$$

is called a *reproducing kernel* of the Hilbert space $\mathcal{H}$ if and only if:

1) $\forall x \in E, \ k(.,x) \in \mathcal{H}$
2) $\forall x \in E, \ \forall f \in \mathcal{H} \ \langle f, k(.,x) \rangle_{\mathcal{H}} = f(x)$

Condition (2) is called *the reproducing property*. From the definition above follows that:

$$
\forall (x,y) \in E \times E, \ k(x,y) = \langle k(.,x), k(.,y) \rangle_{\mathcal{H}}
\tag{2}
$$

Definition II.2 (Real RKHS). A Hilbert Space of real valued functions on E, denoted by $\mathcal{H}$, with reproducing kernel is called a real Reproducing Kernel Hilbert Space or real RKHS.

A RKHS can be characterized as a Hilbert space of real valued functions on E where all the evaluation functionals:

$$
\begin{aligned}
e_x : \mathcal{H} &\to \mathbb{R} \tag{3} \\
f &\mapsto e_x(f) = f(x) \tag{4}
\end{aligned}
$$

are continuous on $\mathcal{H}$. By Riez representation theorem [39] and the reproducing property (see Definition II.1) it follows that $\forall x \in E, \ \forall f \in \mathcal{H}$:

$$
e_x(f) = f(x) = \langle f, k(.,x) \rangle_{\mathcal{H}} \leq \|f\|_{\mathcal{H}} \|k\|_{\mathcal{H}}
$$

As a result, in RKHS, a sequence converging in the norm also converges pointwise.

The following result shows the equivalence between reproducing kernels and positive functions [15].

Lemma 1. *Any reproducing kernel* $k : E \times E \to \mathbb{R}$ *is a symmetric positive definite function, that is, it satisfies:*

$$
\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0
\tag{5}
$$

$\forall N \in \mathbb{N}, \ \forall c_i, c_j \in \mathbb{R}$ *and* $k(x,y) = k(y,x), \ \forall x,y \in E$. *The converse is true.*

That is, positive definite kernels are reproducing kernels of some RKHS. The space spanned by $k(x,.)$ generates a RKHS or a Hilbert space with reproducing kernel k.

---

[2]We will consider only real valued kernels, because are the functions of more practical interest.

If k is a reproducing kernel, then

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j) &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \langle k(.,x_i), k(.,x_j)\rangle_{\mathcal{H}} \\
&= \langle \sum_{i=1}^{N} c_i k(.,x_i), \sum_{j=1}^{N} c_j k(.,x_j)\rangle_{\mathcal{H}} \\
&= \| \sum_{i=1}^{N} c_i k(.,x_i) \|_{\mathcal{H}}^2 \\
&= \ \geq 0
\end{aligned}
$$

That is, the elements of the RKHS are real-valued functions on E of the form $f(.) = \sum_{i=1}^{N} c_i k(.,x_i)$.

Another important result from probability theory is that if k is positive definite, then there is a family of zero-mean Gaussian random variables with k as their covariance function [12].

Examples of reproducing kernels or positive definite kernels on $\mathbb{R}^D \times \mathbb{R}^D$ widely used in machine learning community are:

- Linear kernel $k(x, y) = \langle x, y\rangle$
- Polynomial kernel $k(x, y) = (\langle x, y\rangle + 1)^D$
- Gaussian kernel $k(x, y) = \exp(-\|x - y\|^2/\sigma^2)$

More sophisticated examples are kernels on probability measures [40], kernels on measures [41], kernels on strings [42], and other kernels for non-vectorial data, such graphs, sets and logic terms [43].

As a comment, the space of square integrable functions L^2 is a Hilbert space and is isometric to the space of sequences ℓ^2 but is not a RKHS because it is a space of a equivalence class of functions rather than a function space. Then, L^2 does not have a reproducing kernel, note that the delta function has a reproducing property but does not belong to this space.

Reference [15] gives more details about the theory of RKHS and references [13], [14] give more details about kernel methods in machine learning.

Next, to introduce the concept of kernels on fuzzy sets we will review the concepts of fuzzy set, semi-ring of sets, measure and T-norm operator.

B. Fuzzy Set

Let Ω be a nonempty set. A fuzzy set on Ω, is a set $X \subseteq \Omega$ with membership function

$$
\begin{aligned}
\mu_X : \Omega &\rightarrow [0, 1] \quad &(6)\\
x &\mapsto \mu_X(x). \quad &(7)
\end{aligned}
$$

Definition II.3 (α-cut of a fuzzy set). The α-cut of a fuzzy set $X \subset \Omega$ is the set

$$X_\alpha = \{x \in \Omega | \mu_X(x) \geq \alpha, \alpha \in [0, 1]\}.$$

Definition II.4 (Support of a fuzzy set). The support of a fuzzy set is the set

$$X_{>0} = \{x \in \Omega | \mu_X(x) > 0\}.$$

A complete review of the theory of fuzzy sets and applications can be found in [44].

C. T-Norm

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

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 $n \in \mathbb{N}, n \geq 2$ and associativity, a multiple-valued extension $T_n : [0, 1]^n \rightarrow [0, 1]$ of a T-norm T is given by $T_2 = T$ and

$$T_n(x_1, x_2, \ldots, x_n) = T(x_1, T_{n-1}(x_2, x_3, \ldots, x_n)). \quad (8)$$

We will use T to denote T or T_n.

D. Semi-ring of Sets

A semi-ring of sets, $\mathcal{S}$ on Ω, is a subset of the power set $\mathcal{P}(\Omega)$, that is, a set of sets satisfying:

1   $\phi \in \mathcal{S}$, ϕ denotes the empty set;
2   $A, B \in \mathcal{S}, \implies A \cap B \in \mathcal{S}$;
3   for all $A, A_1 \in \mathcal{S}$ and $A_1 \subseteq A$, there exists a sequence of pairwise disjoint sets $A_2, A_3, \ldots A_N$, such

$$A = \bigcup_{i=1}^{N} A_i.$$

Condition 3 is called *finite decomposition of A*.

E. Measure

Definition II.5 (Measure). Let $\mathcal{S}$ be a semi-ring and let $\rho : \mathcal{S} \rightarrow [0, \infty]$ be a pre-measure, i.e., ρ satisfy:

1   $\rho(\phi) = 0$;
2   for a finite decomposition of $A \in \mathcal{S}$, $\rho(A) = \sum_{i=1}^{N} \rho(A_i)$;

by *Carathodory's extension theorem*, ρ is a measure on $\sigma(\mathcal{S})$, where $\sigma(S)$ is the smallest σ-algebra containing S [45].

From now on, capital letters A, B, C will denote sets and capital letters X, Y, Z will denote fuzzy sets. Notation $\mathcal{F}(\mathcal{S} \subset \Omega)$ stands for the set of all fuzzy sets over Ω whose support belongs to $\mathcal{S}$, i.e.,

$$\mathcal{F}(\mathcal{S} \subset \Omega) = \{X \subset \Omega | X_{>0} \in \mathcal{S}\}.$$

III. KERNELS ON FUZZY SETS

We define kernel functions on fuzzy sets as the mapping

$$
\begin{aligned}
k : \mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega) &\rightarrow \mathbb{R} \\
(X, Y) &\mapsto k(X, Y), \quad (9)
\end{aligned}
$$

where $\mathcal{S}$ is a semi-ring of sets on Ω and $\mathcal{F}(\mathcal{S} \subset \Omega)$ is the set of all fuzzy sets over Ω whose support belongs to $\mathcal{S}$. This is a kernel for non-vector input.

Because each fuzzy set X belongs to $\mathcal{F}(\mathcal{S} \subset \Omega)$, then the support $X_{>0}$ of X admits finite decomposition, that is,

$$X_{>0} = \bigcup_{i \in I} A_i, \ A_i \in \mathcal{S}$$

where $\mathcal{A} = \{A_1, A_2, \ldots, A_N\}$ are pairwise disjoint sets and I stand for an arbitrary index set.

In the following, we will derive some kernels on fuzzy sets, based on the intersection of fuzzy sets and the cross product between their elements.

A. Intersection kernel on Fuzzy Sets

The intersection of two fuzzy sets $X, Y \in \mathcal{F}(\mathcal{S} \subset \Omega)$ is the fuzzy set $X \cap Y \in \mathcal{F}(\mathcal{S} \subset \Omega)$ with membership function

$$\mu_{X \cap Y} : \Omega \rightarrow [0,1] \tag{10}$$
$$x \mapsto \mu_{X \cap Y} = T(\mu_X(x), \mu_Y(x)) \tag{11}$$

where T is a T-norm operator. Using this fact, we define the intersection kernel on fuzzy sets as follows:

Definition III.1 (Intersection Kernel on Fuzzy Sets). Let X, Y be two fuzzy sets in $\mathcal{F}(\mathcal{S} \subset \Omega)$, the intersection kernel on fuzzy sets is the function

$$k : \mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega) \rightarrow \mathbb{R}$$
$$(X \quad , \quad Y) \mapsto k(X,Y) = g(X \cap Y),$$

where g is the mapping

$$g : \mathcal{F}(\mathcal{S} \subset \Omega) \rightarrow [0, \infty]$$
$$X \mapsto g(X)$$

The mapping g plays an important role assigning real values to the intersection fuzzy set $X \cap Y$. We can think about this function as a similarity measure between two fuzzy sets and its design will be highly dependent of the problem and the data.

For instance, our first choice for g uses the fact that the support of $X \cap Y$, has finite decomposition, that is,

$$(X \cap Y)_{>0} = \bigcup_{i \in I} A_i, \ A_i \in \mathcal{S},$$

of pairwise disjoint sets $\{A_1, A_2, \ldots, A_N\}$. We can measure its support using the measure $\rho : \mathcal{S} \rightarrow [0, \infty]$ as follows:

$$\rho((X \cap Y)_{>0}) = \rho(\bigcup_{i \in I} A_i) = \sum_{i \in I} \rho(A_i),$$

The idea to include fuzziness is to weight each $\rho(A_i)$ by a value given by the contribution of the membership function on all the elements of the set A_i.

Next, we give a definition of a intersection kernel on fuzzy sets using the concept of measure and membership function.

Definition III.2 (Intersection Kernel on Fuzzy Sets with measure ρ). Let $\bigcup_{i \in I} A_i, \ A_i \in \mathcal{S}$, a finite decomposition of the support of the intersection fuzzy set $X \cap Y \in \mathcal{F}(\mathcal{S} \subset \Omega)$ as defined before. Let g be the function

$$g : \mathcal{F}(\mathcal{S} \subset \Omega) \rightarrow [0, \infty]$$
$$X \cap Y \mapsto g(X \cap Y) = \sum_{i \in I} \mu_{X \cap Y}(A_i) \rho(A_i)$$

where

$$\mu_{X \cap Y}(A_i) = \sum_{x \in A_i} \mu_{X \cap Y}(x)$$

and ρ is a measure according to Definition (II.5). We define the *Intersection Kernel on Fuzzy Sets with measure ρ* as:

$$k(X,Y) = g(X \cap Y)$$
$$= \sum_{i \in I} \mu_{X \cap Y}(A_i) \rho(A_i) \tag{12}$$

Using the T-norm operator, the intersection kernel on fuzzy sets with measure ρ given by (12) can be written as:

$$k(X,Y) = \sum_{i \in I} \mu_{X \cap Y}(A_i) \rho(A_i)$$
$$= \sum_{i \in I} \sum_{x \in A_i} \mu_{X \cap Y}(x) \rho(A_i)$$
$$= \sum_{i \in I} \sum_{x \in A_i} T(\mu_X(x), \mu_Y(x)) \rho(A_i)$$

Table I shows several kernels examples for different T-norm operators. Function Z in Table I is defined as

$$Z(\mu_X(x), \mu_Y(x)) = \begin{cases} \mu_X(x), & \mu_y(x) = 1 \\ \mu_Y(x), & \mu_X(x) = 1 \\ 0, & otherwise \end{cases}$$

Intersection kernels on fuzzy sets with measure ρ	
$k_{\min}(X,Y)$	$\sum_{i \in I} \sum_{x \in A_i} \min(\mu_X(x), \mu_Y(x)) \rho(A_i)$
$k_P(X,Y)$	$\sum_{i \in I} \sum_{x \in A_i} \mu_X(x) \mu_Y(x) \rho(A_i)$
$k_{\max}(X,Y)$	$\sum_{i \in I} \sum_{x \in A_i} \max(\mu_X(x) + \mu_Y(x) - 1, 0) \rho(A_i)$
$k_Z(X,Y)$	$\sum_{i \in I} \sum_{x \in A_i} Z(\mu_X(x), \mu_Y(x)) \rho(A_i)$

TABLE I
KERNELS ON FUZZY SETS.

More examples can be obtained by setting specific measures, for example, the probability measure $\mathbb{P}$.

The next step is to determine which intersection kernels on fuzzy sets with measure ρ are positive definite, that is, which intersection kernels are reproducing kernels of some RKHS.

Lemma 2. $k_{\min}(X,Y)$ *is positive definite*

Proof: We first, define a function $\mathbf{1}_{[0,a]}$ as:

$$\mathbf{1}_{[0,a]} : [0,1] \rightarrow \{0,1\}$$
$$t \mapsto \mathbf{1}_{[0,a]}(t) = \begin{cases} 1, & t \in [0,a], \ a \in [0,1] \\ 0, & otherwise \end{cases}$$

then function $\min : [0,1] \times [0,1] \rightarrow [0,1]$ can be written as, $\forall a, b \in [0,1]$

$$\min(a,b) = \int_{[0,1]} \mathbf{1}_{[0,a]}(t) \mathbf{1}_{[0,b]}(t) d\lambda(t)$$
$$= \langle \mathbf{1}_{[0,a]}(.), \mathbf{1}_{[0,b]}(.) \rangle_{\mathcal{H}}$$

where $\mathcal{H} = L^2([0,1], \lambda)$, i.e., the space of square integrable functions on $[0,1]$ with respect to the Lebesgue measure λ. By

Lemma (1) and Equation (2), it follows that min is positive definite. That is, for a fixed $x \in \Omega$,

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \min(\mu_{X_i}(x), \mu_{X_j}(x)) \geq 0,$$

$\forall N \in \mathbb{N}$, $\forall c_i, c_j \in \mathbb{R}$, $\forall X_i, X_j \in \mathcal{F}(\mathcal{S} \subset \Omega)$.

Next, we show that k_{min} is positive definite, that is:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \sum_{l \in I}\sum_{x \in A_l} \min(\mu_{X_i}(x), \mu_{X_j}(x)) \rho(A_l) \geq 0,$$

$\forall N \in \mathbb{N}$, $\forall c_i, c_j \in \mathbb{R}$, $\forall X_i, X_j \in \mathcal{F}(\mathcal{S} \subset \Omega), \forall x \in \Omega$, and I stands for an arbitrary index set.

Note that

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \sum_{l \in I}\sum_{x \in A_l} \min(\mu_{X_i}(x), \mu_{X_j}(x)) \rho(A_l)$$
$$= \sum_{l \in I}\sum_{x \in A_l} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \min(\mu_{X_i}(x), \mu_{X_j}(x)) \right) \rho(A_l)$$
$$\geq 0$$

∎

Lemma 3. $k_P(X, Y)$ *is positive definite.*

Proof: Note that for a fixed $x \in \Omega$,

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \mu_{X_i}(x) \mu_{X_j}(x) =$$
$$\left(\sum_{i=1}^{N} c_i \mu_{X_i}(x) \right)^2 \geq 0$$

Next, we show that k_P is positive definite, that is

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \sum_{l \in I}\sum_{x \in A_l} \mu_{X_i}(x) \mu_{X_j}(x) \rho(A_l) \geq 0,$$

$\forall N \in \mathbb{N}$, $\forall c_i, c_j \in \mathbb{R}$, $\forall X_i, X_j \in \mathcal{F}(\mathcal{S} \subset \Omega), \forall x \in \Omega$, and L stands for an arbitrary index set. Note that:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \sum_{l \in I}\sum_{x \in A_l} \mu_{X_i}(x) \mu_{X_j}(x) \rho(A_l)$$
$$= \sum_{l \in I}\sum_{x \in A_l} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \mu_{X_i}(x), \mu_{X_j}(x) \right) \rho(A_l)$$
$$\geq 0$$

∎

It is worth to note that, if the σ-algebra is a Borel algebra of subsets of $\mathbb{R}^D$, then the intersection kernel with measure ρ can be written as

$$k(X, Y) = \int_{\mathbb{R}^D} T(\mu_X(x), \mu_Y(x)) d\rho(x)$$

as for example k_{min} and k_P can be written as

$$k_{min}(X, Y) = \int_{\mathbb{R}^D} \min(\mu_X(x), \mu_Y(x)) d\rho(x) \quad (13)$$

$$k_P = \int_{\mathbb{R}^D} \mu_X(x) \mu_Y(x) d\rho(x) \quad (14)$$

$$(15)$$

B. Cross product kernel between fuzzy sets

Definition III.3. Let $k : \Omega \times \Omega \to \mathbb{R}$ be a positive definite kernel. The cross product kernel between fuzzy sets $X, Y \in \mathcal{F}(\mathcal{S} \subset \Omega)$ is the real valued function $k_\times$ defined on $\mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega)$ as

$$k_\times(X, Y) = \sum_{x \in X}\sum_{y \in Y} k(x, y) \mu_X(x) \mu_Y(y) \quad (16)$$

Lemma 4. *kernel $k_\times$ is positive definite*

Proof: By Definition (III.3)

$$k_\times(X, Y) = \sum_{x \in X}\sum_{y \in Y} k(x, y) \mu_X(x) \mu_Y(y)$$
$$= \langle \sum_{x \in X} k(., x) \mu_X(x), \sum_{y \in Y} k(., y) \mu_Y(y) \rangle$$

∎

IV. Nonsingleton TSK Fuzzy Kernel

Another type of intersection kernel is the *nonsingleton TSK fuzzy kernel* presented in [19]. We will see that this kernel satisfy our definition of intersection kernel (Definition (III.1)) and we will review their properties and study the link with fuzzy equivalence relations.

Definition IV.1 (Nonsingleton TSK Fuzzy Kernel)**.** Let $X \cap Y$ be a fuzzy set given by Definition (III.1) and let g be the function:

$$g : \mathcal{F}(\mathcal{S} \subset \Omega) \to [0, \infty]$$
$$X \cap Y \mapsto g(X \cap Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x),$$

then the *Nonsingleton TSK Fuzzy Kernel* is given by :

$$k_{tks}(X, Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x) \quad (17)$$

Using T-norm operators, this kernel can be written as:

$$k_{tks}(X, Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x)$$
$$= \sup_{x \in \Omega} T(\mu_X(x), \mu_Y(x))$$

Note that the definition of the nonsingleton TSK Fuzzy kernel satisfy the definition of intersection kernel on fuzzy sets (Definition III.1) for the particular setting of $g(X \cap Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x)$.

Lemma 5. *The* Nonsingleton TSK Fuzzy Kernel *is positive definite, that is:*

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k_{tks}(X_i, X_j) \geq 0,$$

$\forall N \in \mathbb{N}, \ \forall c_i, c_j \in \mathbb{R}, \ \forall X_i, X_j \in \mathcal{F}(\mathcal{S} \subset \Omega).$

Proof: Denote k_{tks} by k for clear notation. Let I an arbitrary index set. By commutativity property of T-norms, k is symmetric. Note that:

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

and $\sup_{x \in \Omega} T(\mu_{X_i}(x), \mu_{X_i}(x)) = 1, \ \forall i \in I$ then,

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

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

a)   If $k(X_i, X_j) = 0, \forall i, j \in I : i > j$, then

$$\sum_{i,j \in I} c_i c_j k(X_i, X_j) = \sum_{i \in I} c_i^2 \geq 0$$

b)   If $k(X_i, X_j) = 1, \forall i, j \in I : i > j$, then

$$\sum_{i,j \in I} c_i c_j k(X_i, X_j) = \sum_{i \in I} c_i^2 + 2 \sum_{i,j \in I, i \neq j, i > j} c_i c_j$$
$$= (\sum_{i \in I} c_i)^2 \geq 0$$

$\blacksquare$

Some examples of this kernel are given in [19]

We now review two results from [35] (Corollary 6) and [36] (Theorem 9). The first one shows that every positive definite kernel mapping to the unit interval with constant one in the diagonal is a fuzzy equivalence relation with respect to a given T-norm. The second one shows that such kernels can be viewed as a fuzzy logic formula used to represent fuzzy rule bases. Then we show that the Nonsingleton TSK Fuzzy Kernel satisfy such results.

Definition IV.2 (Fuzzy Equivalence Relation). A function $\mathcal{E} : \mathcal{X} \times \mathcal{X} \to [0, 1]$ is called a fuzzy equivalence relation with respect to the T-norm T if

1) $\forall x \in \mathcal{X}, \ \mathcal{E}(x, x) = 1$;
2) $\forall x, y \in \mathcal{X}, \ \mathcal{E}(x, y) = \mathcal{E}(y, x)$;
3) $\forall x, y, z \in \mathcal{X}, \ T(\mathcal{E}(x, y), \mathcal{E}(y, z)) \leq \mathcal{E}(x, z)$.

The value $\mathcal{E}(x, y)$ can be interpreted as *"x is equal to y.* Condition 3 is called *T-transitivity* and can be regarded as the statement *"If x and y are similar, and y and z are similar then x and z are also similar."* [35].

Lemma 6 (kernels are at least T_{cos} transitive [35]). *Let the nonempty set $\mathcal{X}$. Let $k : \mathcal{X} \times \mathcal{X} \to [0, 1]$ a positive definite kernel such that $\forall x \in \mathcal{X} : k(x, x) = 1$; then $\forall x, y, z \in \mathcal{X}$, kernel k satisfy T_{cos}-transitivity:*

$$T_{cos}(k(x, y), k(y, z)) \leq k(x, z) \qquad (18)$$

where

$$T_{cos}(a, b) = max(ab - \sqrt{1 - a^2}\sqrt{1 - b^2}, 0), \qquad (19)$$

is a Archimedean T-norm and it is the greatest T-norm with this property [35].

Lemma 7 (kernels as fuzzy logic formula for fuzzy rules [36]). *Let the nonempty set $\mathcal{X}$. Let $k : \mathcal{X} \times \mathcal{X} \to [0, 1]$ a positive definite kernel such that $\forall x \in \mathcal{X} : k(x, x) = 1$; then $\forall x, y, z \in \mathcal{X}$, there is a family of membership functions $\mu_{i \in I} : \mathcal{X} \to [0, 1]$, where I is a nonempty index set such that*

$$\forall x, y \in \mathcal{X} : k(x, y) = \inf_{i \in I} \overleftrightarrow{T_M}(\mu_i(x), \mu_i(y)) \qquad (20)$$

where $\overleftrightarrow{T_M} = \min(\overrightarrow{T}(x, y), \overrightarrow{T}(y, x))$ is its induced bi-implication operator and $\overrightarrow{T}(x, y) = \sup\{t \in [0, 1] | T(x, t) \leq y\}$ is a implication function generated from a T-norm T [36].

Lemma 8. *The Nonsingleton TSK Fuzzy Kernel is T_{cos} transitive (Lemma (6)) and admits the representation given by Lemma (7).*

Proof: By construction, the Nonsingleton TSK Fuzzy Kernel is a positive definite kernel such that $\forall X \in \mathcal{F}(\mathcal{S} \subset \Omega) : k_{tsk}(X, X) = 1$ and also k_{tsk} has values in the interval $[0, 1]$. By Lemma (6) k_{tsk} is T_{cos} transitive and by Lemma (7), k_{tsk} admits representation given by Lemma (7). $\blacksquare$

V. MORE KERNELS ON FUZZY SETS

It is easy to construct new kernels on fuzzy sets from the previously defined kernels. For example, if $k_1(., .)$ and $k_2(., .)$ are positive definite kernels on fuzzy sets, by closure properties of kernels [14], also are positive definite kernels on fuzzy sets:

1) $k_1(X, Y) + k_2(X, Y)$;
2) $\alpha k_1(X, Y), \quad \alpha \in \mathbb{R}^+$;
3) $k_1(X, Y) k_2(X, Y)$;
4) $f(X) f(Y), \quad f : \mathcal{F}(\mathcal{S} \subset \Omega) \to \mathbb{R}$;
5) $k_1(f(X), f(Y)), \quad f : \mathcal{F}(\mathcal{S} \subset \Omega) \to \mathcal{F}(\mathcal{S} \subset \Omega)$;
6) $\exp(k_1(X, Y))$;
7) $p(k_1(X, Y)), \quad p$ is a polynomial with positive coefficients.

More kernels on fuzzy sets could be obtained using the nonlinear mapping

$$\phi : \mathcal{F}(\mathcal{S} \subset \Omega) \rightarrow \mathcal{H}$$
$$X \mapsto \phi(X),$$

and using the fact that $k(X, Y) = \langle \phi(X), \phi(Y) \rangle_{\mathcal{H}}$ and

$$D(X, Y) \stackrel{\text{def}}{=} \|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2$$
$$= k(X, Y) - 2k(X, Y) + k(Y, Y),$$

we have the following positive definite kernels on fuzzy sets.

- **Fuzzy Polynomial kernel** $\alpha \geq 0, \beta \in \mathbb{N}$

$$k_{pol}(X, Y) = (\langle \phi(X), \phi(Y) \rangle_{\mathcal{H}} + \alpha)^{\beta}$$
$$= (k(X, Y) + \alpha)^{\beta}.$$

- **Fuzzy Gaussian kernel** $\gamma > 0$

$$k_{gauss}(X, Y) = \exp(-\gamma \|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2)$$
$$= \exp(-\gamma D(X, Y)).$$

- **Fuzzy Rational Quadratic kernel** $\alpha, \beta > 0$

$$
\begin{aligned}
k_{ratio}(X, Y) &= (1 + \frac{\|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2}{\alpha\beta^2})^{-\alpha} \\
&= (1 + \frac{D(X, Y)}{\alpha\beta^2})^{-\alpha}.
\end{aligned}
$$

A. Conditionally Positive Definite Kernels on Fuzzy Sets

We can construct another class of fuzzy kernels using the concept of *conditionally positive definite kernels*, which are kernels satisfying Lemma 1 but, additionally it is required that $\sum_{i=1}^{N} c_i = 0$. Examples of Conditionally Positive Definite Fuzzy Kernels with this property are:

- **Fuzzy Multiquadric kernel**

$$
\begin{aligned}
k_{multi}(X, Y) &= -\sqrt{\|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2 + \alpha^2} \\
&= -\sqrt{D(X, Y) + \alpha^2}.
\end{aligned}
$$

- **Fuzzy Inverse Multiquadric kernel**

$$
\begin{aligned}
k_{invmult}(X, Y) &= (\sqrt{\|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2 + \alpha^2})^{-1} \\
&= (\sqrt{D(X, Y) + \alpha^2})^{-1}.
\end{aligned}
$$

It is easy to construct positive definite kernels from CPD kernels by doing $\exp(tk_{multi}(X, Y))$ and $\exp(tk_{invmult}(X, Y))$ for $t > 0$, because a kernel k is conditionally positive definite if and only if $\exp(tk)$ is positive definite for all $t > 0$ [13]. See Proposition 2.22 of [13] for more details on how to construct positive definite kernels from conditionally positive definite kernels.

VI. COMMENTS ON PRACTICAL APPLICATIONS

All the proposed kernels on fuzzy sets have potential applications in fuzzy data, that is, data sets of the form $\{X_i\}_{i=1}^{N}$, where each X_i is fuzzy set. Kernels on fuzzy sets can be modularly used in several machine learning formulations such as: support vector machines, support vector data description, support vector regression, kernel PCA, Gaussian process, and so on, which only require a positive definite Gram matrix induced by some positive definite kernel, allowing to perform machine learning on fuzzy data with the tools of kernel methods.

If a set of clusters: $s_1, \ldots, s_i, \ldots, s_N$ are given as a dataset, where each cluster s_i contains a set of points $\{x_j\}_{j=1}^{M_i}$ with some degree of membership, for example, points far away the mean of cluster s_i have lower membership than points near the mean of such cluster, then, it is possible to use the *Cross product kernel between fuzzy sets*, by representing each cluster s_i as a fuzzy set. Also. the same reasoning can be used to perform machine learning on big data by means of *data squashing* [46].

As a next step of our research, we are going to perform experiments with the proposed kernels on several machine learning problems. Besides the applications mentioned in the introductory part of this paper, we are particularly interested in apply those kernels in datasets whose observations are clusters,

prototypes or groups of samples, an example of those datasets is the *Sloan Sky Survey Data Set*[3].

Another important applications to be explored are in the areas of big data and large scale machine learning. Because hardware and software requirements, large datasets are difficult to analyze, possible solutions are to construct a summary of data or transform large datasets into a smaller ones (*data squashing* [46]), where all the information of the features are preserved but the size of the dataset is decreased with the hope that the analysis of the smaller dataset gives approximately the same information that the large dataset. Modeling groups of samples as fuzzy sets jointly with the kernels on fuzzy sets in supervised classification problems is an interesting topic to be investigated. Another important topic to be investigated is the performance of another formulations for kernels on fuzzy sets such as the multiple kernel learning approach [47] and non positive kernels [48] on fuzzy sets.

VII. CONCLUSIONS AND FURTHER RESEARCH

It is possible to work with kernel methods for datasets whose observations are fuzzy sets. The set of all fuzzy sets is a example of non-vectorial data, we can embed this data into a vectorial space as the RKHS, using a positive definite kernel defined on the set of all the fuzzy sets. Using some basic concepts of set theory such as semi-ring of sets, we presented a novel formulation for kernels on fuzzy sets. As instances of our general definition, we gave a definition of the intersection kernel on fuzzy sets and the cross product kernel on fuzzy sets. Also we gave some examples of positive definite functions for those kernels. Moreover, we showed that the kernel on fuzzy sets presented in [19] is a fuzzy equivalence relation, admits some special representation as fuzzy logic formula for fuzzy rules and is a special case of the intersection kernel on fuzzy sets. Furthermore, we gave some examples of positive definite kernels on fuzzy sets and conditionally positive definite kernels on fuzzy sets.

---

[3]http://www.sdss.org/

x—

REFERENCES

[1] C. Bertoluzza, M.-Á. Gil, and D. A. Ralescu, *Statistical modeling, analysis and management of fuzzy data*. Physica-Verlag, 2002.

[2] R. Viertl, *Statistical Methods for Fuzzy Data*, ser. Wiley Series in Probability and Statistics. Wiley, 2011. [Online]. Available: http://books.google.com.br/books?id=R5RN3akq5vwC

[3] A. Palacios and J. Alcala-Fdez, "Mining fuzzy association rules from low quality data," *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, pp. 0–0, 2012.

[4] A. Palacios, L. Sánchez, and I. Couso, "Diagnosis of dyslexia with low quality data with genetic fuzzy systems," *International Journal of Approximate Reasoning*, vol. 51, no. 8, pp. 993–1009, 2010.

[5] 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.

[6] L. Sánchez, I. Couso, and J. Casillas, "Genetic learning of fuzzy rules based on low quality data," *Fuzzy Sets and Systems*, vol. 160, no. 17, pp. 2524–2552, 2009.

[7] L. Zadeh, "Soft computing and fuzzy logic," *Software, IEEE*, vol. 11, no. 6, pp. 48 –56, nov. 1994.

[8] P. P. Wang, *Computing with words*. John Wiley & Sons, Inc., 2001.

[9] L. A. Zadeh, "Fuzzy logic= computing with words," *Fuzzy Systems, IEEE Transactions on*, vol. 4, no. 2, pp. 103–111, 1996.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.

[11] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. springer New York, 2006, vol. 1.

[12] G. Wahba, *Spline models for observational data*. Siam, 1990, no. 59.

[13] 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.

[14] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge University Press, 2004.

[15] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*. Kluwer Academic Boston, 2004, vol. 3.

[16] N. Cristianini and J. Shawe-Taylor, *An introduction to support Vector Machines: and other kernel-based learning methods*. New York, NY, USA: Cambridge University Press, 2000.

[17] D. M. Tax and R. P. Duin, "Support vector data description," *Machine learning*, vol. 54, no. 1, pp. 45–66, 2004.

[18] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[19] J. Guevara, R. Hirata, and S. Canu, "Kernel functions in takagi-sugeno-kang fuzzy system with nonsingleton fuzzy input," in *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, 2013, pp. 1–8.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

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

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

[37] 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.

[38] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[39] L. Debnath and P. Mikusiński, *Hilbert Spaces with Applications*. Elsevier Academic Press, 2005.

[40] T. Jebara, R. Kondor, and A. Howard, "Probability product kernels," *The Journal of Machine Learning Research*, vol. 5, pp. 819–844, 2004.

[41] M. Cuturi, K. Fukumizu, and J.-P. Vert, "Semigroup kernels on measures," in *Journal of Machine Learning Research*, 2005, pp. 1169–1198.

[42] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins, "Text classification using string kernels," *The Journal of Machine Learning Research*, vol. 2, pp. 419–444, 2002.

[43] T. Gartner and W. Scientific., *Kernels for structured data*. Hackensack, N.J.: World Scientific, 2008.

[44] W. Pedrycz and F. A. C. Gomide, *Fuzzy Systems Engineering - Toward Human-Centric Computing*. Wiley, 2007.

[45] R. B. Ash and C. A. Doléans-Dade, *Probability and measure theory*. Access Online via Elsevier, 2000.

[46] J. Abello, P. M. Pardalos, and M. G. Resende, *Handbook of massive data sets*. Springer, 2002, vol. 4.

[47] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "Simplemkl," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[48] C. Ong, X. Mary, S. Canu, and A. Smola, "Learning with non-positive kernels," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 81.