Positive Definite Kernel Functions on Fuzzy Sets
FUZZ 2014

Jorge Guevara Díaz\textsuperscript{1}  Roberto Hirata Jr.\textsuperscript{1}  Stéphane Canu\textsuperscript{2}

\textsuperscript{1}Institute of Mathematics and Statistics
University of Sao Paulo-Brazil

\textsuperscript{2}Institut National de Sciences Appliquees
University of Normandy-France

July 08, 2014
Table of Contents

1 Motivation

2 Preliminares

3 Kernel on Fuzzy sets

4 Practical Applications

5 Conclusions
1 Motivation

2 Preliminares

3 Kernel on Fuzzy sets

4 Practical Applications

5 Conclusions
Data representation

Machine learning on vector spaces (feature space)
## Data representation

Machine learning on vector spaces (feature space)

Data representation: graphs, sets, distributions, logic terms, *fuzzy sets*
Data representation

Machine learning on vector spaces (feature space)
Data representation: graphs, sets, distributions, logic terms, fuzzy sets
Kernel methods: Data Representation $\rightarrow$ RKHS
Table of Contents

1 Motivation

2 Preliminares

3 Kernel on Fuzzy sets

4 Practical Applications

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

$$\mu_X : \Omega \rightarrow [0, 1] \quad (1)$$

$$x \mapsto \mu_X(x). \quad (2)$$

**Definition (Support of a fuzzy set)**

The support of a fuzzy set is the set

$$X_{>0} = \{x \in \Omega | \mu_X(x) > 0\}.$$
A triangular norm or T-norm is the function $T : [0, 1]^2 \to [0, 1]$, 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 $T(x, 1) = x$. 
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 $T(x, 1) = x$.

**a multiple-valued extension**

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

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

We will use $T$ to denote $T$ or $T_n$. 
A semi-ring of sets, $S$ on $\Omega$, is a subset of the power set $\mathcal{P}(\Omega)$, that is, a set of sets satisfying:

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

\[ A = \bigcup_{i=1}^{N} A_i. \]
A semi-ring of sets, $\mathcal{S}$ on $\Omega$, is a subset of the power set $\mathcal{P}(\Omega)$, that is, a set of sets satisfying:

1. $\emptyset \in \mathcal{S}$, $\emptyset$ 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 $A_2, A_3, \ldots A_N \subseteq \mathcal{S}$, such that

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

**Finite decomposition**

Condition 3 is called *finite decomposition of $A$.*/
Definition (Measure)

Let $\mathcal{S}$ be a semi-ring and let $\rho : \mathcal{S} \rightarrow [0, \infty]$ be a pre-measure, i.e., $\rho$ 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 Carathéodory’s extension theorem, $\rho$ is a measure on $\sigma(\mathcal{S})$, where $\sigma(\mathcal{S})$ is the smallest $\sigma$-algebra containing $\mathcal{S}$. 
A function

\[ k : E \times E \rightarrow \mathbb{R} \]
\[ (x, y) \mapsto k(x, t) \]  \hspace{1cm} (4)

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. \( \forall x \in E, \ \forall f \in \mathcal{H} \ \langle f, k(., x) \rangle_{\mathcal{H}} = f(x) \)
Definition (Reproducing kernel)

A function

\[ k : E \times E \rightarrow \mathbb{R} \]

\[ (x, y) \leftrightarrow k(x, t) \]  \hspace{1cm} (4)

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

Reproducing property

\[ \forall (x, y) \in E \times E, \ k(x, y) = \langle k(., x), k(., y) \rangle_{\mathcal{H}} \]  \hspace{1cm} (5)
Why kernels?

\[ k(x, x') = \langle k(., x), k(., x') \rangle \]
Definition (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.
Reproducing Kernel Hilbert Spaces

Definition (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.

Characterization
All the evaluation functionals are continuous on $\mathcal{H}$. :

\[ e_x : \mathcal{H} \to \mathbb{R} \]
\[ f \mapsto e_x(f) = f(x) \]
**Definition (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.

**Characterization**

All the evaluation functionals are continuous on $\mathcal{H}$.

\[
e_x : \mathcal{H} \rightarrow \mathbb{R}  \\
\text{such that } e_x(f) = f(x) \quad (7)
\]

A sequence converging in the norm also converges pointwise

By Riez representation theorem and the reproducing property it follows that $\forall x \in E, \forall f \in \mathcal{H}$:

\[
e_x(f) = f(x) = \langle f, k(., x) \rangle_{\mathcal{H}} \leq \|f\|_{\mathcal{H}} \|k\|_{\mathcal{H}}
\]
**Lemma**

Any reproducing kernel \( k : E \times E \to \mathbb{R} \) is a symmetric positive definite function, that is, it satisfies:

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

\( \forall N \in \mathbb{N}, \forall c_i, c_j \in \mathbb{R} \) and \( k(x, y) = k(y, x), \forall x, y \in E \). The converse is true.
Any reproducing kernel \( k : E \times E \to \mathbb{R} \) is a symmetric positive definite function, that is, it satisfies:

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0 \tag{8}
\]

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

Consequently, Kernels \( k(.,.) \) are reproducing kernels of some RKHS. The space spanned by \( k(x,.) \) generates a RKHS or a Hilbert space with reproducing kernel \( k \).
If $k$ is a reproducing kernel, then

\[
\begin{align*}
\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(x_i, x_j) &= \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j \langle k(\cdot, x_i), k(\cdot, x_j) \rangle_{\mathcal{H}} \\
&= \langle \sum_{i=1}^{N} c_i k(\cdot, x_i), \sum_{j=1}^{N} c_j k(\cdot, x_j) \rangle_{\mathcal{H}} \\
&= \| \sum_{i=1}^{N} c_i k(\cdot, x_i) \|_{\mathcal{H}}^2 \\
&\geq 0
\end{align*}
\]
If $k$ is a reproducing kernel, then

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(x_i, x_j) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j \langle k(., x_i), k(., x_j) \rangle_{\mathcal{H}}$$

$$= \langle \sum_{i=1}^{N} c_i k(., x_i), \sum_{j=1}^{N} c_j k(., x_j) \rangle_{\mathcal{H}}$$

$$= \| \sum_{i=1}^{N} c_i k(., x_i) \|_{\mathcal{H}}^2$$

$$\geq 0$$

That is

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

Guevara, Hirata, Canu, Fuzz IEEE 2014 (Un)Positive Definite Kernel Functions on Fuzzy Sets
Examples of reproducing kernels or positive definite kernels

- Linear kernel $k(x, y) = \langle x, y \rangle$ $x, y \in \mathbb{R}^D$
Examples of reproducing kernels or positive definite kernels

- Linear kernel \( k(x, y) = \langle x, y \rangle \) \( x, y \in \mathbb{R}^D \)
- Polynomial kernel \( k(x, y) = (\langle x, y \rangle + 1)^D \) \( x, y \in \mathbb{R}^D \)
Examples of reproducing kernels or positive definite kernels

- Linear kernel $k(x, y) = \langle x, y \rangle \ x, y \in \mathbb{R}^D$
- Polynomial kernel $k(x, y) = (\langle x, y \rangle + 1)^D \ x, y \in \mathbb{R}^D$
- Gaussian kernel $k(x, y) = \exp\left(-\|x - y\|^2/\sigma^2\right) \ x, y \in \mathbb{R}^D$
Examples of reproducing kernels or positive definite kernels

- **Linear kernel** \( k(x, y) = \langle x, y \rangle \quad x, y \in \mathbb{R}^D \)
- **Polynomial kernel** \( k(x, y) = (\langle x, y \rangle + 1)^D \quad x, y \in \mathbb{R}^D \)
- **Gaussian kernel** \( k(x, y) = \exp(-\|x - y\|^2 / \sigma^2) \quad x, y \in \mathbb{R}^D \)
- **Probability product kernel** \( \tilde{k}(\mathbb{P}, \mathbb{Q}) = \int_X \mathbb{P}(x)\rho \mathbb{Q}(x)^\rho dx \)
Examples of reproducing kernels or positive definite kernels

- Linear kernel: $k(x, y) = \langle x, y \rangle$, $x, y \in \mathbb{R}^D$
- Polynomial kernel: $k(x, y) = (\langle x, y \rangle + 1)^D$, $x, y \in \mathbb{R}^D$
- Gaussian kernel: $k(x, y) = \exp\left(-\|x - y\|^2 / \sigma^2\right)$, $x, y \in \mathbb{R}^D$
- Probability product kernel: $\tilde{k}(\mathbb{P}, \mathbb{Q}) = \int_X \mathbb{P}(x)^\rho \mathbb{Q}(x)^\rho \, dx$
- Kernel on probability measures: $X \sim \mathbb{P}, X' \sim \mathbb{Q}$, $\tilde{k}(\mathbb{P}, \mathbb{Q}) = \langle \mathbb{E}_{\mathbb{P}}[k(X.,.)], \mathbb{E}_{\mathbb{Q}}[k(X'.,.)] \rangle_{\mathcal{H}}$
\[ \mathcal{X} \rightarrow \mathcal{H} \]
\[ x \mapsto k(., x) \]

\[ k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} \]
Now the problem is different -fuzzy sets for imprecise data -PD kernels on fuzzy sets

$k((\cdot, (\cdot))$

Uncertain and imprecise data
Observation

Capital letters $A$, $B$, $C$ denote sets and $X$, $Y$, $Z$ denote fuzzy sets.

Remark

Notation

$F(S \subset \Omega)$ stands for the set of all fuzzy sets over $\Omega$ whose support belongs to $S$, i.e., $F(S \subset \Omega) = \{X \subset \Omega | X > 0 \in S\}$.

where $S$ is a semi-ring of sets on $\Omega$

Example

If $X \cap Y \in F(S \subset \Omega)$ then satisfy (finite decomposition):

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

where $\{A_1, A_2, ..., A_N\}$ are pairwise disjoint sets
Kernel on fuzzy sets

Observation
Capital letters $A, B, C$ denote sets and $X, Y, Z$ denote fuzzy sets.

Remark
Notation $\mathcal{F}(S \subset \Omega)$ stands for the set of all fuzzy sets over $\Omega$ whose support belongs to $S$, i.e.,

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

where $S$ is a semi-ring of sets on $\Omega$
Kernel on fuzzy sets

Observation
Capital letters $A, B, C$ denote sets and $X, Y, Z$ denote fuzzy sets.

Remark
Notation $\mathcal{F}(S \subset \Omega)$ stands for the set of all fuzzy sets over $\Omega$ whose support belongs to $S$, i.e.,

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

where $S$ is a semi-ring of sets on $\Omega$

Example
If $X \cap Y \in \mathcal{F}(S \subset \Omega)$ then satisfy (finite decomposition):

$$(X \cap Y)_{>0} = \bigcup_{i \in I} A_i, \ A_i \in S,$$
We can measure \((X \cap Y)_{>0} = \bigcup_{i \in I} A_i, A_i \in S\) using the measure \(\rho : S \to [0, \infty]\) as follows:

\[
\rho((X \cap Y)_{>0}) = \rho(\bigcup_{i \in I} A_i) = \sum_{i \in I} \rho(A_i),
\]
Kernel on fuzzy sets

Definition (Kernels on Fuzzy Sets)

A kernel on fuzzy sets is a real valued-function

\[ k : \mathcal{F}(S \subseteq \Omega) \times \mathcal{F}(S \subseteq \Omega) \rightarrow \mathbb{R} \]

\[ (X, Y) \mapsto k(X, Y), \quad (9) \]

Observation

Because each fuzzy set \( X \) belongs to \( \mathcal{F}(S \subseteq \Omega) \), then the support \( X \geq 0 \) of \( X \) admits a finite decomposition, that is,

\[ X \geq 0 = \bigcup_{i \in I} A_i, \quad A_i \in S \]

where \( \{ A_1, A_2, \ldots, A_N \} \) are pairwise disjoint sets and \( I \) stand for an arbitrary index set.
Kernel on fuzzy sets

**Definition (Kernels on Fuzzy Sets)**

A kernel on fuzzy sets is a real valued-function

\[
    k : \mathcal{F}(S \subset \Omega) \times \mathcal{F}(S \subset \Omega) \to \mathbb{R}
    
    (X, Y) \mapsto k(X, Y),
\]

\( (9) \)

**Observation**

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

\[
    X_{>0} = \bigcup_{i \in I} A_i, \ A_i \in S
\]

where \( \{A_1, A_2, \ldots, A_N\} \) are pairwise disjoint sets and \( I \) stand for an arbitrary index set.
Definition (Intersection Kernel on Fuzzy Sets)

Let $X, Y$ in $\mathcal{F}(S \subset \Omega)$, the intersection kernel on fuzzy sets is

$$k : \mathcal{F}(S \subset \Omega) \times \mathcal{F}(S \subset \Omega) \rightarrow \mathbb{R}$$

$$k(X, Y) = g(X \cap Y),$$

Where $g : \mathcal{F}(S \subset \Omega) \rightarrow [0, \infty]$ and the FS $X \cap Y \in \mathcal{F}(S \subset \Omega)$ has M.F.

$$\mu_{X \cap Y} : \Omega \rightarrow [0, 1],$$

$$x \mapsto \mu_{X \cap Y}(x) = \min(\mu_X(x), \mu_Y(x)).$$
Intersection Kernel on Fuzzy Sets

Definition (Intersection Kernel on Fuzzy Sets)

Let $X, Y$ in $\mathcal{F}(S \subset \Omega)$, the intersection kernel on fuzzy sets is

$$k : \mathcal{F}(S \subset \Omega) \times \mathcal{F}(S \subset \Omega) \rightarrow \mathbb{R}$$

$$(X, Y) \mapsto k(X, Y) = g(X \cap Y),$$

Where

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

$X \mapsto g(X)$$
Definition (Intersection Kernel on Fuzzy Sets)

Let $X, Y$ in $\mathcal{F}(S \subset \Omega)$, the intersection kernel on fuzzy sets is

$$k : \mathcal{F}(S \subset \Omega) \times \mathcal{F}(S \subset \Omega) \rightarrow \mathbb{R}$$

$$\quad (X, Y) \mapsto k(X, Y) = g(X \cap Y),$$

Where

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

$$X \mapsto g(X)$$

and the FS $X \cap Y \in \mathcal{F}(S \subset \Omega)$ has M.F.

$$\mu_{X \cap Y} : \Omega \rightarrow [0, 1]$$

$$x \mapsto \mu_{X \cap Y} = T(\mu_X(x), \mu_Y(x)) \quad (10)$$

$$\mu_X(x) = \mu_Y(x) \quad (11)$$
From previous example.

We can measure \((X \cap Y)_{>0} = \bigcup_{i \in I} A_i, \; A_i \in S\) using the measure \(\rho : S \rightarrow [0, \infty]\) as follows:

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

Adding fuzziness

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 \(A_i\).
From previous example.

We can measure \((X \cap Y)_{>0} = \bigcup_{i \in I} A_i, \ A_i \in S\) using the measure \(\rho : S \to [0, \infty]\) as follows:

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

Adding fuzziness

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 \(A_i\).
Intersection Kernel on Fuzzy Sets

Definition (Intersection Kernel on Fuzzy Sets with measure $\rho$)

Using $(X \cap Y)_{>0} = \bigcup_{i \in I} A_i, A_i \in S$. Let $g$ be the function

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

$$X \cap Y \mapsto g(X \cap Y) = \sum_{i \in I} \mu_{X \cap Y}(A_i) \rho(A_i)$$

where

$$\mu_{X \cap Y}(A_i) = \sum_{x \in A_i} \mu_{X \cap Y}(x).$$
Intersection Kernel on Fuzzy Sets

**Definition (Intersection Kernel on Fuzzy Sets with measure $\rho$)**

Using $(X \cap Y) > 0 = \bigcup_{i \in I} A_i$, $A_i \in S$. Let $g$ be the function

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

$$X \cap Y \mapsto g(X \cap Y) = \sum_{i \in I} \mu_{X \cap Y}(A_i) \rho(A_i)$$

where

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

**Intersection Kernel on Fuzzy Sets with measure $\rho$**

We define the *Intersection Kernel on Fuzzy Sets with measure $\rho$* as:

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

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

\[
k(X, Y) = \sum_{i \in I} \mu_{X \cap Y}(A_i) \rho(A_i)
\]

\[
= \sum_{i \in I} \sum_{x \in A_i} \mu_{X \cap Y}(x) \rho(A_i)
\]

\[
= \sum_{i \in I} \sum_{x \in A_i} T(\mu_X(x), \mu_Y(x)) \rho(A_i)
\]
Some kernel examples for different T-norm operators

<table>
<thead>
<tr>
<th>Intersection kernels on fuzzy sets with measure $\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{\min}(X, Y)$</td>
</tr>
<tr>
<td>$\sum_{i \in I} \sum_{x \in A_i} \min(\mu_X(x), \mu_Y(x))\rho(A_i)$</td>
</tr>
<tr>
<td>$k_P(X, Y)$</td>
</tr>
<tr>
<td>$\sum_{i \in I} \sum_{x \in A_i} \mu_X(x)\mu_Y(x)\rho(A_i)$</td>
</tr>
<tr>
<td>$k_{\max}(X, Y)$</td>
</tr>
<tr>
<td>$\sum_{i \in I} \sum_{x \in A_i} \max(\mu_X(x) + \mu_Y(x) - 1, 0)\rho(A_i)$</td>
</tr>
<tr>
<td>$k_Z(X, Y)$</td>
</tr>
<tr>
<td>$\sum_{i \in I} \sum_{x \in A_i} Z(\mu_X(x), \mu_Y(x))\rho(A_i)$</td>
</tr>
</tbody>
</table>

Table: kernels on fuzzy sets.

Function $Z$ is defined as

$$Z(\mu_X(x), \mu_Y(x)) = \begin{cases} 
\mu_X(x), & \text{if } \mu_Y(x) = 1 \\
\mu_Y(x), & \text{if } \mu_X(x) = 1 \\
0, & \text{otherwise}
\end{cases}$$
**Lema**

\[ k_{\text{min}}(X, Y) = \sum_{i \in I} \sum_{x \in A_i} \min(\mu_X(x), \mu_Y(x)) \rho(A_i) \]

is positive definite

**Lema**

\[ k_P(X, Y) = \sum_{i \in I} \sum_{x \in A_i} \mu_X(x) \mu_Y(x) \rho(A_i) \]

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

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

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

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

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

$$ k_P = \int_{\mathbb{R}^D} T(\mu_X(x), \mu_Y(x)) d\rho(x) \quad (15) $$
Definition

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

$$k_\times(X, Y) = \sum_{x \in X} \sum_{y \in Y} k(x, y) \mu_X(x) \mu_Y(y)$$  \hspace{1cm} (16)

Lema

kernel $k_\times$ is positive definite
Definition (Nonsingleton TSK Fuzzy Kernel)

Let $X \cap Y$ be a fuzzy set given by Definition (3.2) and let $g$ be the function:

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

$$X \cap Y \mapsto g(X \cap Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x),$$

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:

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

$$= \sup_{x \in \Omega} T(\mu_{X}(x), \mu_{Y}(x))$$
The Nonsingleton TSK Fuzzy Kernel is positive definite, that is:

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

$$\forall N \in \mathbb{N}, \forall c_i, c_j \in \mathbb{R}, \forall X_i, X_j \in \mathcal{F}(S \subset \Omega).$$
Examples of nonsingleton TSK fuzzy kernels

Gaussian MF

\[ k_{tks}(X, Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x) \]
\[ = \sup_{x \in \Omega} T(\mu_X(x), \mu_Y(x)) \]
\[ = \exp \left\{ -\frac{1}{2} \sum_{j=1}^{p} \frac{(m_j - m_j^l)^2}{\sigma_j^2 + (\sigma_j^l)^2} \right\} \]
Examples of nonsingleton TSK fuzzy kernels

**Gaussian MF**

\[
k_{tks}(X, Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x) \\
= \sup_{x \in \Omega} T(\mu_X(x), \mu_Y(x)) \\
= \exp \left\{ -\frac{1}{2} \sum_{j=1}^{p} \frac{(m_j - m_j^t)^2}{\sigma_j^2 + (\sigma_j^t)^2} \right\}
\]

**Gaussian MF with parameter \( \gamma \in \mathbb{R} \)**

\[
k_{tks, \gamma}(X, Y) = \exp \left\{ -\frac{1}{2} \sum_{j=1}^{p} \frac{(m_j - m_j^t)^2}{\sigma_j^2 + (\sigma_j^t)^2 + \gamma} \right\}
\]
More kernels on fuzzy sets

If $k_1(.,.)$ and $k_2(.,.)$ are PD kernels on fuzzy sets, by closure properties of PD kernels, also are PD kernels on fuzzy sets:

1. $k_1(X, Y) + k_2(X, Y)$;
2. $\alpha k_1(X, Y)$, $\alpha \in \mathbb{R}^+$;
3. $k_1(X, Y)k_2(X, Y)$;
4. $f(X)f(Y)$, $f : \mathcal{F}(S \subset \Omega) \rightarrow \mathbb{R}$;
5. $k_1(f(X), f(Y))$, $f : \mathcal{F}(S \subset \Omega) \rightarrow \mathcal{F}(S \subset \Omega)$;
6. $\exp(k_1(X, Y))$;
7. $p(k_1(X, Y))$, $p$ is a polynomial with positive coefficients.
More kernels on fuzzy sets could be obtained using the nonlinear mapping

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

$$X \mapsto \phi(X),$$

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

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

$$= k(X, Y) - 2k(X, Y) + k(Y, Y),$$
The following kernels are PD kernels on fuzzy sets:

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

\[
k_{pol}(X, Y) = (\langle \phi(X), \phi(Y) \rangle_H + \alpha)^\beta = (k(X, Y) + \alpha)^\beta.
\]

- **Fuzzy Gaussian kernel** \( \gamma > 0 \)

\[
k_{gauss}(X, Y) = \exp(-\gamma \|\phi(X) - \phi(Y)\|_H^2) = \exp(-\gamma D(X, Y)).
\]

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

\[
k_{ratio}(X, Y) = (1 + \frac{\|\phi(X) - \phi(Y)\|_H^2}{\alpha \beta^2})^{-\alpha} = (1 + \frac{D(X, Y)}{\alpha \beta^2})^{-\alpha}.
\]
**Conditionally Positive Definite Kernels on Fuzzy Sets**

**Lema**

*CPD kernels are symmetric kernels satisfying:*

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0, \quad \sum_{i=1}^{N} c_i = 0.
\]  

(19)

\(\forall N \in \mathbb{N}, \forall c_i, c_j \in \mathbb{R}.

- **Fuzzy Multiquadric kernel**

\[
k_{\text{multi}}(X, Y) = -\sqrt{\|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2 + \alpha^2} = -\sqrt{D(X, Y) + \alpha^2}.
\]

- **Fuzzy Inverse Multiquadric kernel**

\[
k_{\text{invmult}}(X, Y) = (\sqrt{\|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2 + \alpha^2})^{-1}
\]
Machine Learning Applications

- ML Applications in fuzzy data, i.e., \( \{X_i\}_{i=1}^N \), where each \( X_i \) is FS.

Kernels on FS can be used modularly in SVM, SVDD, kernel PCA, Gaussian process.

ML on big data, by data squashing.

Promising results on low quality data.

Reference

Machine Learning Applications

- ML Applications in fuzzy data, i.e., \( \{X_i\}_{i=1}^{N} \), where each \( X_i \) is FS.
- Kernels on FS can be used modularly in SVM, SVDD, kernel PCA, Gaussian process.

Reference

Machine Learning Applications

- ML Applications in fuzzy data, i.e., $\{X_i\}_{i=1}^N$, where each $X_i$ is FS.
- Kernels on FS can be used modularly in SVM, SVDD, kernel PCA, Gaussian process.
- ML on big data, by data squashing.
Machine Learning Applications

- ML Applications in fuzzy data, i.e., \( \{X_i\}_{i=1}^N \), where each \( X_i \) is FS.
- Kernels on FS can be used modularly in SVM, SVDD, kernel PCA, Gaussian process.
- ML on big data, by data squashing.
- Promising results on low quality data.

Reference

Table of Contents

1 Motivation
2 Preliminares
3 Kernel on Fuzzy sets
4 Practical Applications
5 Conclusions
Theory of kernels allow us to represent fuzzy sets as functions in a RKHS.
Tool to use: semiring of sets and measures
Important applications on fuzzy data
Summary

- kernels on fuzzy sets
Summary

- kernels on fuzzy sets
- PD fuzzy intersection kernels

Acknowledgement

The authors are thankful with FAPESP grant # 2011/50761-2, CNPq, CAPES, NAP eScience - PRP - USP for their financial support.
kernels on fuzzy sets
PD fuzzy intersection kernels
CPD fuzzy kernels

Acknowledgements

The authors are thankful with FAPESP grant # 2011/50761-2, CNPq, CAPES, NAP eScience - PRP - USP for their financial support.