Support Measure Data Description for Group Anomaly Detection

1 Jorge Guevara Díaz,
2 Stephane Canu,
1 Roberto Hirata Jr.

1 IME-USP, Brazil. 2 LITIS-INS, France.

November 03, 2015.
Outline

1. Who am I?
2. Introduction
3. Hilbert space embedding for distributions
4. SMDD models
5. Experiments
6. Conclusions and further research
Outline

1. Who am I?

2. Introduction

3. Hilbert space embedding for distributions

4. SMDD models

5. Experiments

6. Conclusions and further research
Who am I?

Bachelor graduate project
Title: A Speech Recognition system for vocals recognition using a feedforward neural net.
Supervisor: Professor Ronald Leon
Description: We performed experiments with a feed-forward neural net with back-propagation algorithm to classify vocal sounds.

Informatics engineering thesis
Title: Feature Extraction using Wavelets for Speech Recognition
Supervisor: Professor Ronald Leon
Description: We developed an algorithm for feature extraction from speech audio using wavelet theory, we applied such algorithm in isolated speech recognition. Text in spanish.

- Three papers in local peruvian conferences.
- Lorito. A isolated word recognition from scratch, in JAVA. https://github.com/jorjasso
Who am I?

Masters Thesis

<table>
<thead>
<tr>
<th>Title</th>
<th>Speech Recognition Framework for Information Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervisor</td>
<td>Professor Ronald Leon</td>
</tr>
<tr>
<td>Description</td>
<td>We explore speech recognition techniques for text retrieval from speech audios in spanish language. Text in spanish.</td>
</tr>
</tbody>
</table>

https://github.com/jorjasso

- Speech Miner. Speech recognition system using MFCC, HMM. using HTK
- SOM-TSP. SOM neural network to solve the TSP problem in JAVA.
- FNN. Neural network for digit classification in JAVA
Who am I?

Teaching Experience

2008–2009  **Speech Recognition**, *5rd year BSc*, National University of Trujillo, Peru.
2008–2009  **Image Processing**, *4rd year BSc*, National University of Trujillo, Peru.
2007–2008  **Artificial Intelligence**, *4rd year BSc*, National University of Trujillo, Peru.
Who am I?

Doctoral Thesis

Title  *Supervised Machine Learning using kernel methods, probability measures and fuzzy set theory*

Supervisor  Professor Roberto Hirata Junior, University of Sao Paulo, Brazil.

Internship  Professor Stephane Canu, INSA-ROUEN, France.

Description  This thesis explored the idea of learning on training sets of points, where each individual point is itself a set. We treat each point-set as a realization of a fuzzy random variable or, as a realization of a random probability measure. We develop kernel algorithms to deal with such data.

Similarity between fuzzy sets using kernels

- link between fuzzy systems and kernels
- theory of positive definite kernels on fuzzy sets
- kernels induced by fuzzy distance

A data description model for set of distributions
Doctoral Thesis

Title  Supervised Machine Learning using kernel methods, probability measures and fuzzy set theory

Supervisor  Professor Roberto Hirata Junior, University of Sao Paulo, Brazil.
Internship  Professor Stephane Canu, INSA-ROUEN, France.

Description  This thesis explored the idea of learning on training sets of points, where each individual point is itself a set. We treat each point-set as a realization of a fuzzy random variable or, as a realization of a random probability measure. We develop kernel algorithms to deal with such data.

Similarity between fuzzy sets using kernels

- link between fuzzy systems and kernels
- theory of positive definite kernels on fuzzy sets
- kernels induced by fuzzy distance

A data description model for set of distributions
https://github.com/jorjasso

- Fuzzy kernel hypothesis testing
- TSK kernels on fuzzy sets, classification of low quality datasets.
- Group anomaly detection using SMDD.
Who am I?
1. Who am I?

2. Introduction

3. Hilbert space embedding for distributions

4. SMDD models

5. Experiments

6. Conclusions and further research
Introduction

Definition (Anomaly)

An anomaly is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism.

Figure: Rare starfish found. One in a million!


b) https://twitter.com/gotham3/status/421258659620855809
Some applications include:

- Fraud detection. i.e., abnormal buying patterns;
- Medicine.  i.e., Unusual symptoms, abnormal tests;
- Sports. i.e., Outstanding players;
- Measurement errors. i.e., abnormal values.
- Cyber-intrusion detection;
- Industrial damage detection;
- Image processing;
- Textual anomaly detection

(a) Unusual contraction. (b) Ceramic defects. (c) Traffic jam recognition.


b) https://www.nde-ed.org/

c) https://www.youtube.com/watch?v=DAXUzWnsiQk
Techniques:

- Generative models. i.e., HMM, GMM;
- Unsupervised methods. i.e., clustering, distance-based, density based;
- Discriminative models. i.e., SVM, neural networks;
- Information Theoretic Methods, Geometric methods

Figure: 2D-anomalies

Problem

Given a data set of the form

$$\mathcal{T} = \{s_i\}_{i=1}^N,$$

(1)

where $s_i = \{x_{1}^{(i)}, x_{2}^{(i)}, \ldots, x_{L_i}^{(i)}\} \sim \mathbb{P}_i$, and $\mathbb{P}_i$ defined on $(\mathbb{R}^D, \mathcal{B}(\mathbb{R}^D))$. Try to detect anomalies or group anomalies from $\mathcal{T}$.
Problem

Given a data set of the form

\[ \mathcal{T} = \{s_i\}_{i=1}^N, \]  

(2)

where \( s_i = \{x_1^{(i)}, x_2^{(i)}, \ldots, x_{L_i}^{(i)}\} \sim \mathbb{P}_i \), and \( \mathbb{P}_i \) defined on \( (\mathbb{R}^D, \mathcal{B}(\mathbb{R}^D)) \).

Try to detect anomalies or group anomalies from \( \mathcal{T} \)

A more complex scenario

![Diagrams showing data distribution and clustering](image-url)
Examples of datasets of this form

\[ \mathcal{T} = \left\{ s_i \right\}_{i=1}^{N} \quad (3) \]

(a) Cluster of galaxies  
(b) SIFT.  
(c) USPS.


Guevara 2015 (IME-USP)  IBM-Research  November 03, 2015.  15 / 49
Group of points with unexpected behavior wrt a dataset of group of points.

- **Point-based group anomalies** aggregation of anomalous points

- **Distributed-based group anomalies** anomalous aggregation of non-anomalous points.
Previous work

- Feature engineering approach
  - Feature extraction from each group\textsuperscript{a,b}
  - Clustering point anomalies

However, they ignore distributed-based group anomalies.

- Generative approach
  - Flexible genre models\textsuperscript{c}
  - Hierarchical probabilistic models\textsuperscript{d}

However, procedures rely on parametric assumptions.

- Discriminative approach
  - Support measure machines\textsuperscript{e}
  - Our work

nonparametric, performance depends on the kernel choice.

\textsuperscript{a} Chan et al. "Modeling multiple time series for anomaly detection," in Data Mining, IEEE.
\textsuperscript{b} Keogh et al. "HOT SAX: efficiently finding the most unusual time series subsequence," in Data Mining, IEEE.
\textsuperscript{c} L Xion et al. "Group Anomaly Detection using Flexible Genre Models", NIPS.
\textsuperscript{d} L Xion et al. "Hierarchical Probabilistic Models for Group Anomaly Detection", AISTATS.
\textsuperscript{e} Muandet et al. "One-Class Support Measure Machines for Group Anomaly Detection", UAI.
Outline

1. Who am I?
2. Introduction
3. Hilbert space embedding for distributions
4. SMDD models
5. Experiments
6. Conclusions and further research
Figure: Kernel mapping

Figure from Shawe-Taylor et al. "Kernel Methods for Pattern Analysis". Cambridge University Press.
RKHS, where the magic happens

Main ingredient

- A real-valued symmetric positive definite kernel $k$.

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

- SVM
- kernel PCA
- Gaussian process
- SVDD
- MKL
- SMDD
- Kernel regression
- Kernel two-sample test
- kernel spectral clustering

\[
f^* = \arg\min_{f \in H} \text{Cost}(f(x_1, y_1, f(x_1)), \ldots, (x_N, y_N, f(x_N))) + \Omega(\|f\|)
\]

\[
f^*(x) = \sum \alpha_i k(x_i, x)
\]
Kernel methods

- SVM
- kernel PCA
- Gaussian process
- SVDD
- MKL
- SMDD
- Kernel regression
- Kernel two-sample test
- kernel spectral clustering

Representer Theorem

\[ f^* = \arg\min_{f \in \mathcal{H}} Cost((x_1, y_1, f(x_1)), \ldots, (x_N, y_N, f(x_N))) + \Omega(\|f\|) \]  \hspace{1cm} (4)

\[ f^*(.) = \sum_{i=1}^{N} \alpha_i k(., x_i) \]  \hspace{1cm} (5)
Remembering the problem

Detecting group anomalies from the set $\mathcal{T} = \{s_i\}_{i=1}^N$. 
Hilbert space embedding for distributions

Framework for embedding probability measures into a RKHS $\mathcal{H}$.

**Definition**

The embedding of probability measures $P \in \mathcal{P}$ into $\mathcal{H}$ is given by the mapping

$$
\mu : \mathcal{P} \rightarrow \mathcal{H}
$$

$$
P \mapsto \mu_P = \mathbb{E}_P[k(X, .)] = \int_{x \in \mathbb{R}^D} k(x, .) dP(x).
$$
Hilbert space embedding for distributions

Properties

- if $\mu_P(X) = \mathbb{E}_P[k(X, X)] < \infty$, with $k$ being measurable then $\mu_P \in \mathcal{H}$

$\mu_P$ is the representative function of $P$ if $k$ is characteristic then $\mu_P : P \rightarrow \mathcal{H}$ is injective

The term $\|\mu_P - \mu_{emp}\|$, is bounded, where $\mu_{emp}$ is a empirical estimator of $\mu_{emp}$
Properties

- if $\mu_P(X) = \mathbb{E}_P[k(X, X)] < \infty$, with $k$ being measurable then $\mu_P \in \mathcal{H}$
- Reproducing property $\langle f, \mu_P \rangle = \langle f, \mathbb{E}_P[k(X, .)] \rangle = \mathbb{E}_P[f(X)]$ holds for all $f \in \mathcal{H}$
Properties

- If \( \mu_P(X) = \mathbb{E}_P[k(X, X)] < \infty \), with \( k \) being measurable then \( \mu_P \in \mathcal{H} \)
- Reproducing property \( \langle f, \mu_P \rangle = \langle f, \mathbb{E}_P[k(X, .)] \rangle = \mathbb{E}_P[f(X)] \) holds for all \( f \in \mathcal{H} \)
- \( \mu_P \) is the representative function of \( P \)
Properties

- If $\mu_P(X) = \mathbb{E}_P[k(X, X)] < \infty$, with $k$ being measurable then $\mu_P \in \mathcal{H}$
- Reproducing property $\langle f, \mu_P \rangle = \langle f, \mathbb{E}_P[k(X, .)] \rangle = \mathbb{E}_P[f(X)]$ holds for all $f \in \mathcal{H}$
- $\mu_P$ is the representative function of $P$
- If $k$ is characteristic then $\mu : \mathcal{P} \rightarrow \mathcal{H}$ is injective
Hilbert space embedding for distributions

Properties

- if $\mu_P(X) = \mathbb{E}_P[k(X, X)] < \infty$, with $k$ being measurable then $\mu_P \in \mathcal{H}$

- Reproducing property $\langle f, \mu_P \rangle = \langle f, \mathbb{E}_P[k(X, .)] \rangle = \mathbb{E}_P[f(X)]$ holds for all $f \in \mathcal{H}$

- $\mu_P$ is the representative function of $P$

- if $k$ is characteristic then $\mu : \mathcal{P} \rightarrow \mathcal{H}$ is injective

- The term $\|\mu_P - \mu_{emp}\|$, is bounded, where $\mu_{emp}$ is a empirical estimator of $\mu_{emp}$
The mapping
\[
P \times P \rightarrow \mathbb{R}
\]
\[(P, Q) \mapsto \langle P, Q \rangle_P = \langle \mu_P, \mu_Q \rangle_H \]
defines an inner product on \( P \). The real-valued kernel on \( P \times P \), defined by
\[
\tilde{k}(P, Q) = \langle P, Q \rangle_P = \langle \mu_P, \mu_Q \rangle_H
\]
\[= \int_{x \in \mathbb{R}^D} \int_{x' \in \mathbb{R}^D} k(x, x') dP(x) dQ(x') \tag{6}\]
is positive definite.
Minimum volume sets (MV-set)

- a MV-set is a set satisfying some optimization criteria.

**Definition (MV-set for probability measures)**

Let \((\mathcal{P}, \mathcal{A}, \mathcal{E})\) be a probability space, where \(\mathcal{P}\) is the space of all probability measures \(\mathcal{P}\) on \((\mathbb{R}^D, \mathcal{B}(\mathbb{R}^D))\), \(\mathcal{A}\) is some suitable \(\sigma\)-algebra of \(\mathcal{P}\) and \(\mathcal{E}\) is a probability measure on \((\mathcal{P}, \mathcal{A})\). The MV-set is the set

\[
G^*_\alpha = \arg\min_{G \in \mathcal{A}} \{ \rho(G) | \mathcal{E}(G) \geq \alpha \}, \quad (7)
\]

where \(\rho\) is a reference measure on \(\mathcal{A}\) and \(\alpha \in [0, 1]\). The MV-set \(G^*_\alpha\) describes a fraction \(\alpha\) of the mass concentration of \(\mathcal{E}\).
Assuming that $\{P_i\}_{i=1}^{N}$ is an i.i.d. sample distributed according to $E$.

Assuming that each $P_i$ is unknown.
Minimum volume sets (MV-set)

- Assuming that $\{P_i\}_{i=1}^N$ is an i.i.d. sample distributed according to $\mathcal{E}$
- Assuming that each $P_i$ is unknown.

Examples of three different classes of volume-sets

$$\hat{G}_1(R, c) = \{P_i \in \mathcal{P} \mid ||\mu_{P_i} - c||_H^2 \leq R^2\}, \quad (8)$$
Assuming that $\{P_i\}_{i=1}^N$ is an i.i.d. sample distributed according to $\mathcal{E}$ and assuming that each $P_i$ is unknown.

Examples of three different classes of volume-sets

\[ \hat{G}_1(R, c) = \{ P_i \in \mathcal{P} \mid \| \mu_{P_i} - c \|_H^2 \leq R^2 \} , \tag{8} \]

\[ \hat{G}_2(R, c) = \{ P_i \in \mathcal{P} \mid \| \mu_{P_i} - c \|_H^2 \leq R^2 , \| \mu_{P_i} \|_H^2 = 1 \} . \tag{9} \]
Minimum volume sets (MV-set)

- Assuming that $\{\mathbb{P}_i\}_{i=1}^N$ is an i.i.d. sample distributed according to $\mathcal{E}$
- Assuming that each $\mathbb{P}_i$ is unknown.

Examples of three different classes of volume-sets

$$
\hat{\mathcal{G}}_1(R, c) = \{\mathbb{P}_i \in \mathcal{P} \mid \|\mu_{\mathbb{P}_i} - c\|_{\mathcal{H}}^2 \leq R^2\}, \quad (8)
$$

$$
\hat{\mathcal{G}}_2(R, c) = \{\mathbb{P}_i \in \mathcal{P} \mid \|\mu_{\mathbb{P}_i} - c\|_{\mathcal{H}}^2 \leq R^2, \|\mu_{\mathbb{P}}\|_{\mathcal{H}}^2 = 1\}. \quad (9)
$$

$$
\hat{\mathcal{G}}_3(\mathcal{K}) = \{\mathbb{P}_i \in \mathcal{P} \mid \mathbb{P}_i(\|k(X_i, \cdot) - c\|_{\mathcal{H}}^2 \leq R^2) \geq 1 - \kappa_i\}. \quad (10)
$$
First SMDD model

$$\hat{G}_1(R, c) = \{P_i \in \mathcal{P} \mid \|\mu_{P_i} - c\|_H^2 \leq R^2\}$$

Given the mean functions \(\{\mu_{P_i}\}_{i=1}^N\) of \(\{P_i\}_{i=1}^N\), the SMDD is:

**Problem**

\[
\begin{align*}
\min_{c \in \mathcal{H}, R \in \mathbb{R}^+, \xi \in \mathbb{R}^N} & \quad R^2 + \lambda \sum_{i=1}^N \xi_i \\
\text{subject to} & \quad \|\mu_{P_i} - c\|_H^2 \leq R^2 + \xi_i, \ i = 1, \ldots, N \\
& \quad \xi_i \geq 0, \ i = 1, \ldots, N.
\end{align*}
\]
Proposition (Dual form)

The dual form of the previously problem is given by:

Problem

$$\max_{\alpha \in \mathbb{R}^N} \sum_{i=1}^{N} \alpha_i \tilde{k}(P_i, P_i) - \sum_{i,j=1}^{N} \alpha_i \alpha_j \tilde{k}(P_i, P_j)$$

subject to

$$0 \leq \alpha_i \leq \lambda, \quad i = 1, \ldots, N$$

$$\sum_{i=1}^{N} \alpha_i = 1$$

where $$\tilde{k}(P_i, P_j) = \langle \mu_{P_i}, \mu_{P_j} \rangle_H$$ and $$\alpha$$ is a Lagrange multiplier vector with non negative components $$\alpha_i$$. 
Proposition (Representer theorem)

\[ c(.) = \sum_{i} \alpha_i \mu_{P_i}, \quad i \in \{i \in \mathcal{I} \mid 0 < \alpha_i \leq \lambda \}, \]

where \( \mathcal{I} = \{1, 2, \ldots, N\} \). Furthermore,

- all \( P_i, \ i \in \{i \in \mathcal{I} \mid \alpha_i = 0\} \) are inside the MV-set \( \hat{G}_\alpha^* \).
- All \( P_i, \ i \in \{i \in \mathcal{I} \mid \alpha_i = \lambda\} \) are the training errors.
- All \( P_i, \ i \in \{i \in \mathcal{I} \mid 0 < \alpha_i < \lambda\} \) are the support measures.
Proposition (Representer theorem)

\[ c(.) = \sum_{i} \alpha_i \mu_{P_i}, \quad i \in \{i \in \mathcal{I} \mid 0 < \alpha_i \leq \lambda\}, \]

where \( \mathcal{I} = \{1, 2, \ldots, N\} \). Furthermore,

- all \( P_i, \ i \in \{i \in \mathcal{I} \mid \alpha_i = 0\} \) are inside the MV-set \( \hat{G}^*_{\alpha} \).
- All \( P_i, \ i \in \{i \in \mathcal{I} \mid \alpha_i = \lambda\} \) are the training errors.
- All \( P_i, \ i \in \{i \in \mathcal{I} \mid 0 < \alpha_i < \lambda\} \) are the support measures.

Theorem

Let \( \eta \) be the Lagrange multiplier of the constraint \( \sum_{i=1}^{N} \alpha_i = 1 \), then

\[ R^2 = -\eta + \|c\|^2_{\mathcal{H}}. \]
Proposition (Representer theorem)

\[ c(.) = \sum_i \alpha_i \mu_{P_i}, \quad i \in \{ i \in \mathcal{I} \mid 0 < \alpha_i \leq \lambda \}, \]

where \( \mathcal{I} = \{1, 2, \ldots, N\} \). Furthermore,

- All \( P_i, \ i \in \{ i \in \mathcal{I} \mid \alpha_i = 0 \} \) are inside the MV-set \( \hat{G}_\alpha^* \).
- All \( P_i, \ i \in \{ i \in \mathcal{I} \mid \alpha_i = \lambda \} \) are the training errors.
- All \( P_i, \ i \in \{ i \in \mathcal{I} \mid 0 < \alpha_i < \lambda \} \) are the support measures.

Theorem

Let \( \eta \) be the Lagrange multiplier of the constraint \( \sum_{i=1}^{N} \alpha_i = 1 \), then

\[ R^2 = -\eta + \| c \|^2_\mathcal{H}. \]

If \( P_t \) is described by this SMDD model, then this must be true:

\[ \| \mu_{P_t} - c \|^2_\mathcal{H} = \tilde{k}(P_t, P_t) - 2 \sum_i \alpha_i \tilde{k}(P_i, P_t) + \sum_{i,j} \alpha_i \alpha_j \tilde{k}(P_i, P_j) \leq R, \quad (11) \]
Second SMDD model

- Mean maps with stationary kernels do not have constant norm
  \[ \| \mu_P \|_\mathcal{H} = \| \mathbb{E}_P[k_I(X, .)] \|_\mathcal{H} \leq \mathbb{E}_P[\| k_I(X, .) \|_\mathcal{H}] = \sqrt{|\epsilon|} \]
- Normalize mean maps to lie on a surface of some hypersphere
  \[ \tilde{k}(P_i, P_j) = \frac{\langle \mu_P, \mu_Q \rangle_\mathcal{H}}{\sqrt{\langle \mu_P, \mu_P \rangle_\mathcal{H} \langle \mu_Q, \mu_Q \rangle_\mathcal{H}}} \]
- The injectivity of \( \mu : \mathcal{P} \rightarrow \mathcal{H} \) is preserved.

Figure: Figure from Muandet et al "One-class support measure machines for group anomaly detection."
This could be modeled by the class of volume sets

\[
\hat{G}_2(R, c) = \{ \mathbb{P}_i \in \mathcal{P} \mid \|\mu_{\mathbb{P}_i} - c\|_H^2 \leq R^2, \|\mu_{\mathbb{P}}\|_H^2 = 1 \}. \quad (13)
\]

and formulated by the following optimization problem:

**Problem (M2)**

\[
\begin{align*}
\max_{\alpha \in \mathbb{R}^N} & - \sum_{i,j=1}^{N} \alpha_i \alpha_j \tilde{k}(\mathbb{P}_i, \mathbb{P}_j) \\
\text{subject to} & \quad 0 \leq \alpha_i \leq \lambda, \quad i = 1, \ldots, N \\
& \quad \sum_{i=1}^{N} \alpha_i = 1.
\end{align*}
\]
Estimate the MV-set over the class of the volume-sets given by

\[ \hat{G}_3(\mathcal{K}) = \left\{ \mathbb{P}_i \in \mathcal{P} \mid \mathbb{P}_i(\|k(X_i,.) - c\|_{\mathcal{H}}^2 \leq R^2) \geq 1 - \kappa_i \right\}. \quad (14) \]

Given the mean functions \( \{\mu_{\mathbb{P}_i}\}_{i=1}^{N} \) of \( \{\mathbb{P}_i\}_{i=1}^{N} \), and \( \{\kappa_i\}_{i=1}^{N} \), \( \kappa_i \in [0, 1] \), the SMDD model is the following chance constrained problem:

\[
\begin{align*}
\min & \quad c \in \mathcal{H}, R \in \mathbb{R}, \xi \in \mathbb{R}^N \\
\text{subject to} & \quad \mathbb{P}_i(\|k(X_i,.) - c(.)\|_{\mathcal{H}}^2 \leq R^2 + \xi_i) \geq 1 - \kappa_i, \\
& \quad \xi_i \geq 0, \\
& \quad \text{for all } i = 1, \ldots, N.
\end{align*}
\]
- **Markov's inequality.**

\[ \mathbb{P}_i(\| k(X_i, .) - c(.) \|_H^2 \geq R^2 + \xi_i) \leq \frac{\mathbb{E}_\mathbb{P}[\| k(X_i, .) - c(.) \|_H^2]}{R^2 + \xi_i}, \quad (15) \]

holds, for all \( i = 1, 2, \ldots, N \).

- **Trace of the covariance operator:** \( \Sigma^\mathcal{H} : \mathcal{H} \rightarrow \mathcal{H} \),

\[ tr(\Sigma^\mathcal{H}) = \mathbb{E}_\mathbb{P}[k(X, X)] - \tilde{k}(\mathbb{P}, \mathbb{P}), \quad (16) \]

- This allows to use the following result

**Lemma**

\[ \mathbb{E}_\mathbb{P}[\| k(X, .) - c(.) \|_H^2] = tr(\Sigma^\mathcal{H}) + \| \mu_\mathbb{P} - c(.) \|_H^2. \]
Given the mean functions \( \{\mu_{P_i}\}_{i=1}^N \) of \( \{P_i\}_{i=1}^N \) and \( \{\kappa_i\}_{i=1}^N \), \( \kappa_i \in (0, 1] \), the SMDD model is:

**Problem**

\[
\begin{align*}
\min_{c \in \mathcal{H}, R \in \mathbb{R}, \xi \in \mathbb{R}^N} & \quad R^2 + \lambda \sum_{i=1}^N \xi_i \\
\text{subject to} & \quad \|\mu_{P_i} - c(.)\|_\mathcal{H}^2 \leq (R^2 + \xi_i)\kappa_i - \text{tr}(\Sigma_i^\mathcal{H}), \\
& \quad \xi_i \geq 0,
\end{align*}
\]

for all \( i = 1, \ldots, N \)
Proposition (Dual form)

The dual form is given by the following fractional programming problem

Problem (M1)

\[
\max_{\alpha \in \mathbb{R}^N} \sum_{i=1}^{N} \alpha_i \langle \mu_{P_i}, \mu_{P_i} \rangle_H - \frac{\sum_{i,j=1}^{N} \alpha_i \alpha_j \langle \mu_{P_i}, \mu_{P_j} \rangle_H}{\sum_{i=1}^{N} \alpha_i} \\
+ \sum_{i=1}^{N} \alpha_i \text{tr}(\Sigma_i^H) \\
\text{subject to } 0 \leq \alpha_i \kappa_i \leq \lambda, \ i = 1, \ldots, N \\
\sum_{i=1}^{N} \alpha_i \kappa_i = 1,
\]

where \( \langle \mu_{P_i}, \mu_{P_j} \rangle_H \) is computed by \( \tilde{k}(P_i, P_j) \)
Proposition (Representer theorem)

\[ c(.) = \frac{\sum_i \alpha_i \mu P_i}{\sum_i \alpha_i}, \quad i \in \{i \in \mathcal{I} \mid 0 < \alpha_i \kappa_i \leq \lambda\}, \tag{17} \]

where \( \mathcal{I} = \{1, 2, \ldots, N\} \). Furthermore,

- all \( P_i, \ i \in \{i \in \mathcal{I} \mid \alpha_i = 0\} \) are inside the MV-set \( \hat{G}^*_\alpha \).
- All \( P_i, \ i \in \{i \in \mathcal{I} \mid \alpha_i \kappa_i = \lambda\} \) are the training errors.
- All \( P_i, \ i \in \{i \in \mathcal{I} \mid 0 < \alpha_i \kappa_i < \lambda\} \) are the support measures.

Theorem

Let \( \eta \) be the Lagrange multiplier of the constraint \( \sum_{i=1}^{N} \alpha_i \kappa_i = 1 \) then \( R^2 = -\eta \).
Group anomalies could be detected if the term $\|\mu_{P_t} - c\|_H^2 + tr(\Sigma_H^t) \geq R$ is true, where $\|\mu_{P_t} - c\|$ is given by

$$\tilde{k}(P_t, P_t) - 2 \sum_i \alpha_i \tilde{k}(P_i, P_t) + \sum_{i,j} \alpha_i \alpha_j \tilde{k}(P_i, P_j) + tr(\Sigma_H^t)$$

(18)
1. Who am I?

2. Introduction

3. Hilbert space embedding for distributions

4. SMDD models

5. Experiments

6. Conclusions and further research
The Most Mysterious Star in Our Galaxy

Astronomers have spotted a strange mess of objects whirling around a distant star. Scientists who search for extraterrestrial civilizations are scrambling to get a closer look.

In the Northern hemisphere’s sky, hovering above the Milky Way, there are two constellations—Cygnus the swan, her wings outstretched in full flight, and Lyra, the harp that accompanied poetry in ancient Greece, from which we take our word “lyric.”

Between these constellations sits an unusual star, invisible to the naked eye, but visible to the Kepler Space Telescope, which stared at it for more than four years, beginning in 2009.
The Star KIC 8462852 phenomena

In the Northern hemisphere’s sky, hovering above the Milky Way, there are two constellations—Cygnus the swan, her wings outstretched in full flight, and Lyra, the harp that accompanied poetry in ancient Greece, from which we take our word “lyric.”

Between these constellations sits an unusual star, invisible to the naked eye, but visible to the Kepler Space Telescope, which stared at it for more than four years, beginning in 2009.
Group anomaly detection in astronomical data

The Star KIC 8462852 phenomena

In the Northern hemisphere’s sky, hovering above the Milky Way, there are two constellations—Cygnus the swan, her wings outstretched in full flight, and Lyra, the harp that accompanied poetry in ancient Greece, from which we take our word “lyric.”

Between these constellations sits an unusual star, invisible to the naked eye, but visible to the Kepler Space Telescope, which stared at it for more than four years, beginning in 2009.

Has the Kepler Space Telescope Discovered an Alien Megastructure?

By Ian O’Neill, Discovery News | October 15, 2015 12:13pm ET
Group anomaly detection in astronomical data

(a) Cluster of galaxies

(b) A galaxy.

(c) Spectrum of a galaxy.

Outline

1. Who am I?

2. Introduction

3. Hilbert space embedding for distributions

4. SMDD models

5. Experiments

6. Conclusions and further research
Conclusions and further research

- SMDD is a non parametric kernel method, with promising results in group anomaly detection task.
Conclusions and further research

- SMDD is a non-parametric kernel method, with promising results in group anomaly detection task.
- Performance depends on the kernel choice, and regularization parameter.

Further research:
- Create models for classification, regression and other ML tasks that exploit structural information of groups.

References:
- Guevara, Jorge and Canu, Stephane and Hirata Jr, Roberto. Support Measure Data Description for Group Anomaly Detection. 2015, SIGACM KDD, ODDx3'15.
Conclusions and further research

- SMDD is a non parametric kernel method, with promising results in group anomaly detection task.
- Performance depends on the kernel choice, and regularization parameter.
- More models can be easily induced based on the idea of MV-sets.

References

Guevara, Jorge and Canu, Stephane and Hirata Jr, Roberto Support Measure Data Description for Group Anomaly Detection 2015, SIGACM KDD, ODDx3'15.

Conclusions and further research

- SMDD is a non parametric kernel method, with promising results in group anomaly detection task
- Performance depends on the kernel choice, and regularization parameter
- More models can be easily induced based on the idea of MV-sets

Further research

- Create models for classification, regression and other ML tasks
Conclusions and further research

- SMDD is a non parametric kernel method, with promising results in group anomaly detection task.
- Performance depends on the kernel choice, and regularization parameter.
- More models can be easily induced based on the idea of MV-sets.

Further research

- Create models for classification, regression and other ML tasks.
- Exploit structural information of groups.
Conclusions and further research

- SMDD is a non parametric kernel method, with promising results in group anomaly detection task
- Performance depends on the kernel choice, and regularization parameter
- More models can be easily induced based on the idea of MV-sets

Further research

- Create models for classification, regression and other ML tasks
- exploit structural information of groups

References

- Guevara, Jorge and Canu, Stephane and Hirata Jr, Roberto Support Measure Data Description for Group Anomaly Detection 2015, SIGACM KDD, ODDx3’15.
- Guevara, Jorge and Canu, Stephane and Hirata Jr, Roberto Support Measure Data Description, 2014, Techical Report IME-USP.
Any questions?
Point-based group anomaly detection over a Gaussian mixture distribution dataset

- M4: OCSMM
- M5: SVDD
- Kernel on probability measures: gaussian kernel, \( \gamma \) given by the median heuristic.
- 200 runs, training set = 50 groups, test set = 30 groups (20 group anomalies).
- métricas AUC e ACC

(a) AUC.  
(b) ACC Non A.  
(c) ACC Anom.  
(d) Médias grup.
Distribution-based group anomaly detection over a Gaussian mixture distribution dataset

(a) AUC.  (b) ACC Non A.  (c) ACC Anom.  (d) Médias grup.

(e) AUC.  (f) ACC Non A.  (g) ACC Anom.  (h) Médias grup.