Sparsity regularization and graph-based representation in medical imaging

Ph.D. Thesis Defense

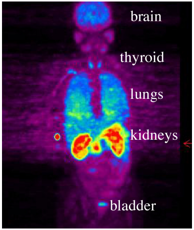
Katerina Gkirtzou

Supervisors : Nikos Paragios and Matthew Blaschko

17th December, 2013

Medical Imaging

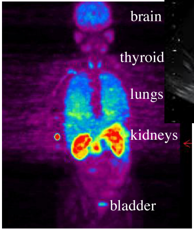
PET



Medical Imaging

Ultrasound

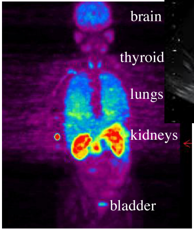
PET



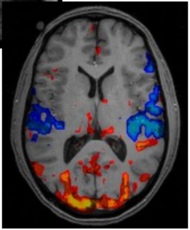
Medical Imaging

Ultrasound

PET



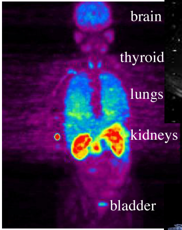
fMRI



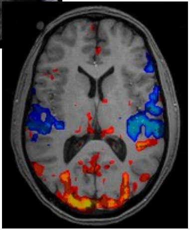
Medical Imaging

Ultrasound

PET



fMRI

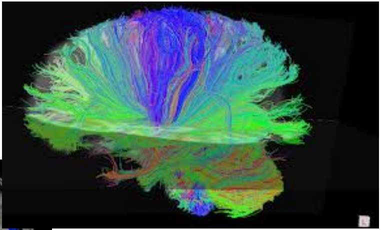


X-rays

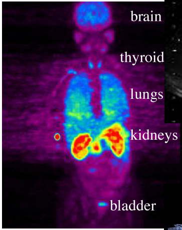
Medical Imaging

Ultrasound

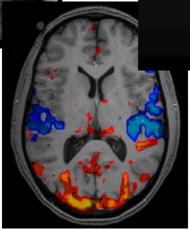
DTI



PET



fMRI

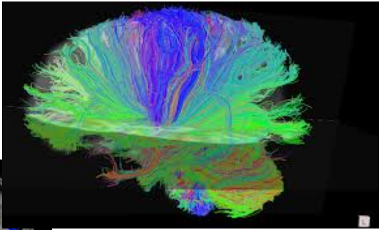


X-rays

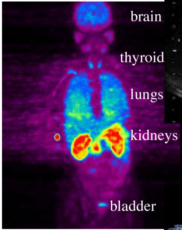
Medical Imaging

Ultrasound

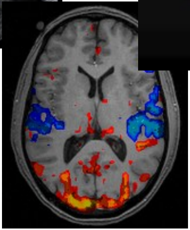
DTI



PET

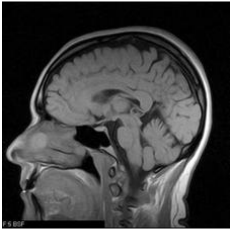


fMRI



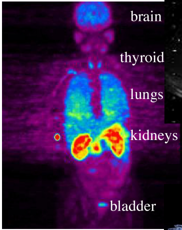
X-rays

MRI



Medical Imaging

PET

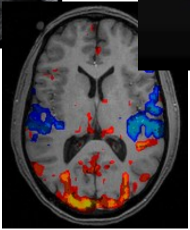


Ultrasound

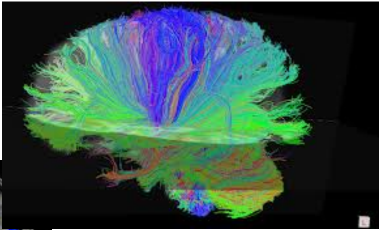
X-rays

Image Classification

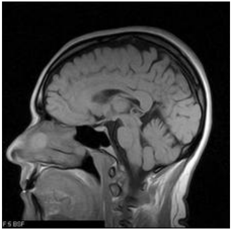
fMRI



DTI



MRI



Medical image classification

Given a set of n paired observations $\{(\mathbb{I}_i, y_i)\}_{1\leq i\leq n}$ where

- $\mathbb{I}_i$ is an medical image and
- $y_i \in \mathbb{R}$ is the classification label

the goal is to learn a *classification function* f.

Medical image classification

Given a set of n paired observations $\{(\mathbb{I}_i, y_i)\}_{1 \leq i \leq n}$ where

- $\mathbb{I}_i$ is an medical image and
- $y_i \in \mathbb{R}$ is the classification label

the goal is to learn a *classification function* f.

Problems

1. The representation of $\phi(\mathbb{I})$.
2. The learning process.

Medical image classification

Given a set of n paired observations $\{(\mathbb{I}_i, y_i)\}_{1 \leq i \leq n}$ where

- $\mathbb{I}_i$ is an medical image and
- $y_i \in \mathbb{R}$ is the classification label

the goal is to learn a *classification function* f.

Problems

1. The representation of $\phi(\mathbb{I})$.
   - Bag of Words approach
   - Graph representation.
2. The learning process.

Medical image classification

Given a set of n paired observations $\{(\mathbb{I}_i, y_i)\}_{1 \leq i \leq n}$ where

- $\mathbb{I}_i$ is an medical image and
- $y_i \in \mathbb{R}$ is the classification label

the goal is to learn a *classification function* f.

Problems

1. The representation of $\phi(\mathbb{I})$.
    - Bag of Words approach
    - Graph representation.
2. The learning process.

Medical image classification

Given a set of n paired observations $\{(\mathbb{I}_i, y_i)\}_{1 \leq i \leq n}$ where

- $\mathbb{I}_i$ is an medical image and
- $y_i \in \mathbb{R}$ is the classification label

the goal is to learn a *classification function* f.

Problems

1. The representation of $\phi(\mathbb{I})$.
   - Bag of Words approach
   - Graph representation.
2. Supervised statistical learning framework

$$\arg\min_{f \in \mathcal{F}} \lambda\Omega(f) + \overbrace{\frac{1}{n}\sum_{i=1}^{n} \mathbb{L}(f(\phi(\mathbb{I}_i)), y_i)}^{\text{Empirical Risk}}$$

where $\mathbb{L}$ is the *loss function* and $\lambda\Omega(f)$ is the regularization term.

Table of Content

Table of Content

What is a graph and why is it interesting?

Definition

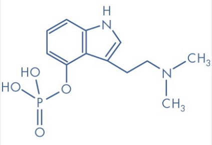
A *labeled graph* G is defined as a triplet $(V, E, \mathcal{L})$, where V is the *vertex set* and $E \subseteq V \times V$ is the *edge set* which represents a binary relation on V and $\mathcal{L} : X \mapsto \Sigma$ is a function assigning a label from an alphabet Σ to each element of the set X, which can be either V, E or $V \cup E$.

What is a graph and why is it interesting?

Definition

A *labeled graph* G is defined as a triplet $(V, E, \mathcal{L})$, where V is the *vertex set* and $E \subseteq V \times V$ is the *edge set* which represents a binary relation on V and $\mathcal{L} : X \mapsto \Sigma$ is a function assigning a label from an alphabet Σ to each element of the set X, which can be either V, E or $V \cup E$.

Areas of application



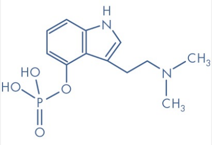
Chemoinformatics

What is a graph and why is it interesting?

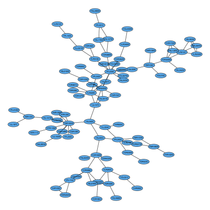
Definition

A *labeled graph* G is defined as a triplet $(V, E, \mathcal{L})$, where V is the *vertex set* and $E \subseteq V \times V$ is the *edge set* which represents a binary relation on V and $\mathcal{L} : X \mapsto \Sigma$ is a function assigning a label from an alphabet Σ to each element of the set X, which can be either V, E or $V \cup E$.

Areas of application



Chemoinformatics



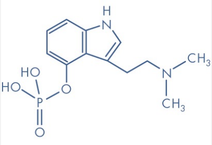
Bioinformatics

What is a graph and why is it interesting?

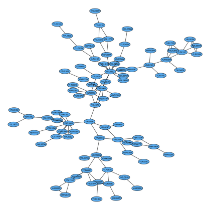
Definition

A *labeled graph* G is defined as a triplet $(V, E, \mathcal{L})$, where V is the *vertex set* and $E \subseteq V \times V$ is the *edge set* which represents a binary relation on V and $\mathcal{L} : X \mapsto \Sigma$ is a function assigning a label from an alphabet Σ to each element of the set X, which can be either V, E or $V \cup E$.

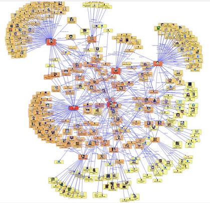
Areas of application



Chemoinformatics



Bioinformatics



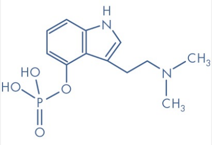
SNA

What is a graph and why is it interesting?

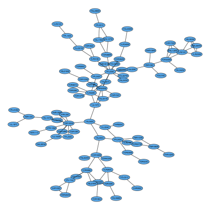
Definition

A *labeled graph* G is defined as a triplet $(V, E, \mathcal{L})$, where V is the *vertex set* and $E \subseteq V \times V$ is the *edge set* which represents a binary relation on V and $\mathcal{L} : X \mapsto \Sigma$ is a function assigning a label from an alphabet Σ to each element of the set X, which can be either V, E or $V \cup E$.

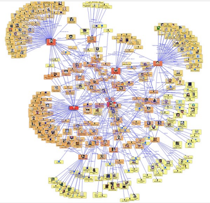
Areas of application



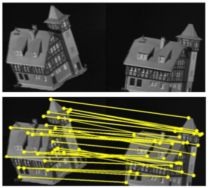
Chemoinformatics



Bioinformatics



SNA



Computer Vision

Graph Comparison Problem

Definition

Given a set $\mathcal{G}$ of graphs, the problem of graph comparison is defined as a function

$$k : \mathcal{G} \times \mathcal{G} \mapsto \mathbb{R}$$

such that $k(G, G')$ for $G, G' \in \mathcal{G}$ quantifies the similarity of G and G'.

Graph Comparison Problem

Definition

Given a set $\mathcal{G}$ of graphs, the problem of graph comparison is defined as a function

$$k : \mathcal{G} \times \mathcal{G} \mapsto \mathbb{R}$$

such that $k(G, G')$ for $G, G' \in \mathcal{G}$ quantifies the similarity of G and G'.

1st Approach

- Graph Isomorphism
- Subgraph Isomorphism
- Largest common subgraph

Graph Comparison Problem

Definition

Given a set $\mathcal{G}$ of graphs, the problem of graph comparison is defined as a function

$$k : \mathcal{G} \times \mathcal{G} \mapsto \mathbb{R}$$

such that $k(G, G')$ for $G, G' \in \mathcal{G}$ quantifies the similarity of G and G'.

1st Approach

- Graph Isomorphism - No efficient algorithm is known
- Subgraph Isomorphism - Proven to be NP-complete
- Largest common subgraph - Proven to be NP-hard

Graph Comparison Problem

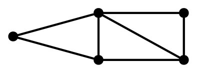
Definition

Given a set $\mathcal{G}$ of graphs, the problem of graph comparison is defined as a function

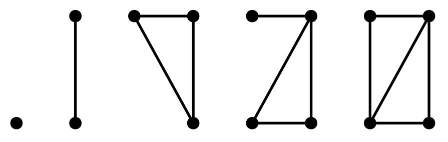
$$k : \mathcal{G} \times \mathcal{G} \mapsto \mathbb{R}$$

such that $k(G, G')$ for $G, G' \in \mathcal{G}$ quantifies the similarity of G and G'.

2nd Approach - R-convolution Kernels



$\Rightarrow$



$\cdots$

Graph Comparison Problem

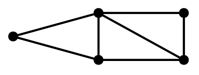
Definition

Given a set $\mathcal{G}$ of graphs, the problem of graph comparison is defined as a function

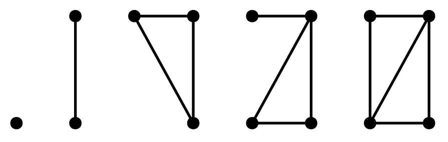
$$k : \mathcal{G} \times \mathcal{G} \mapsto \mathbb{R}$$

such that $k(G, G')$ for $G, G' \in \mathcal{G}$ quantifies the similarity of G and G'.

2nd Approach - R-convolution Kernels



$\Rightarrow$



...

Calculating all subgraphs is at least as hard as deciding whether two graphs are isomorphic [Gärtner 03]

Graph kernels

	Algorithm	Complexity[1]	Unlabeled	Discrete	Continuous	Vector
Walks	[Gärtner 03]	$\mathcal{O}(n^2 v^6)$	✓	✓	✓	✓
	[Mahé 04]		✓	✓		
	[Vishwanathan 10]	$\mathcal{O}(n^2 v^3)$	✓	✓	✓	✓
Paths	[Borgwardt 05]	$\mathcal{O}(n^2 v^4)$	✓	✓	✓	✓
	[Ralaivola 05]		✓	✓		
Graphlets	[Horváth 04]		✓	✓		
	[Shervashidze 09]	$\mathcal{O}(vd^{k-1})$	✓			
	[Costa 10]		✓	✓		
Subtree Patterns	[Ramon 03]	$\mathcal{O}(n^2 v^2 h 4^d)$	✓	✓		
	[Bach 08]		✓	✓		
	[Mahé 09]		✓	✓		
	[Shervashidze 11]	$\mathcal{O}(nhe + n^2 hv)$	✓	✓		

[1] where n is the number of graphs, v is the maximal number of nodes, e is the maximal number of edges, h is the height of subtree patterns, d is the maximum degree and k is the size of graphlets.

Table of Content

Overview of the WLpyramid

Given a set $\mathcal{G} = \{G_i = (V_i, E_i, \mathcal{L}_i)\}_{1 \leq i \leq n}$ where $\mathcal{L}_i : V_i \mapsto \mathbb{R}^d$

1. A pyramid quantization of the label space.
2. Transformation of the initial graphs.
3. Produce subtree features with Weisfeiler-Lehman algorithm.
4. Learning the combination of the subtree features.

Overview of the WLpyramid

Given a set $\mathcal{G} = \{G_i = (V_i, E_i, \mathcal{L}_i)\}_{1 \leq i \leq n}$ where $\mathcal{L}_i : V_i \mapsto \mathbb{R}^d$

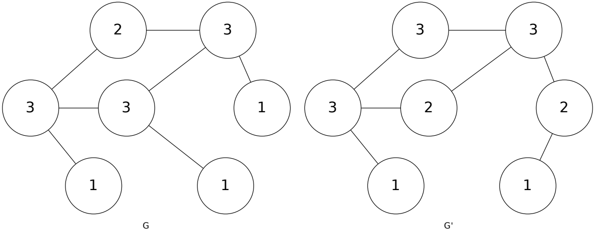
1. A pyramid quantization of the label space.
2. Transformation of the initial graphs.
3. Produce subtree features with Weisfeiler-Lehman algorithm.
4. Learning the combination of the subtree features.

Why Weisfeiler-Lehman?

1. Computational time $\mathcal{O}(nhe)$
   - n the number of graphs
   - e the maximal number of edges and and
   - h the height subtree features.
2. Competitive accuracy in several classification benchmark data sets [Shervashidze 11].

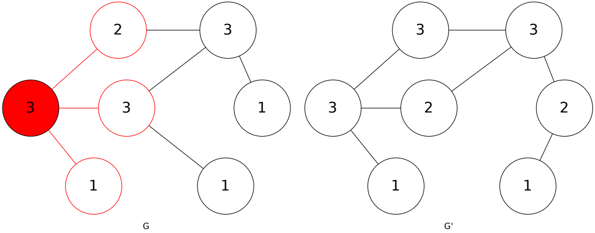
The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

Given labeled graphs G and G'



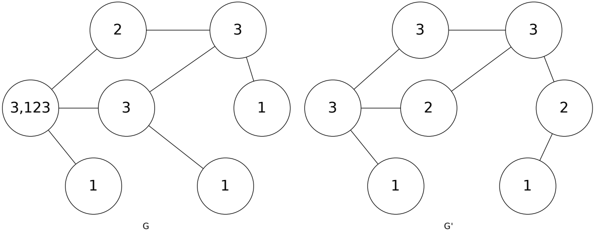
The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

Multi-label determination of G and G'



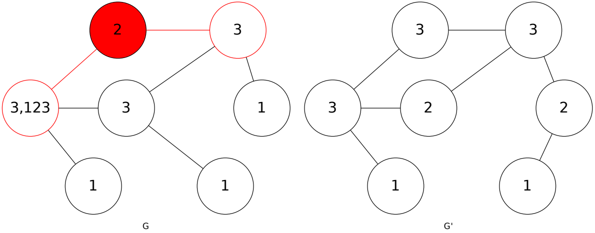
The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

Multi-label determination of G and G'



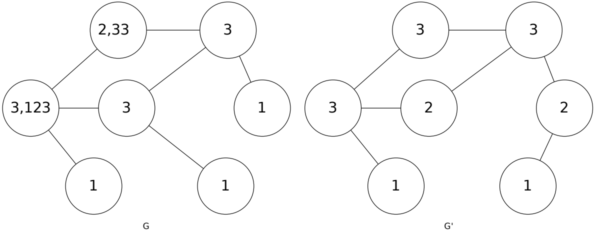
The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

Multi-label determination of G and G'



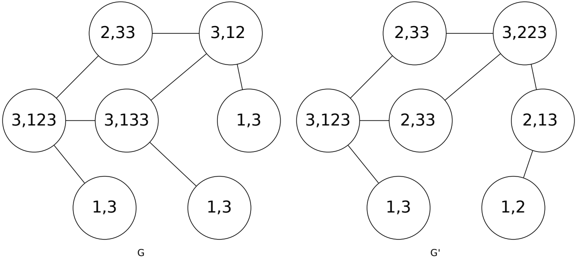
The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

Multi-label determination of G and G'



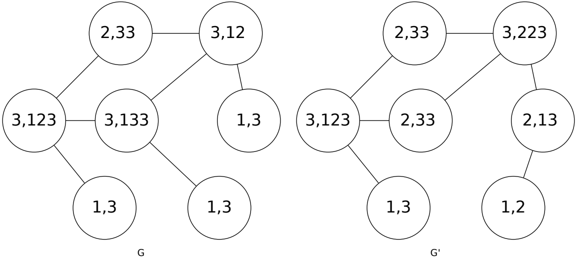
The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

Multi-label determination of G and G'



The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

Multi-label determination of G and G'



Label compression via hashing

1,2 ⟶ 4

1,3 ⟶ 5

2,13 ⟶ 6

2,33 ⟶ 7

3,12 ⟶ 8

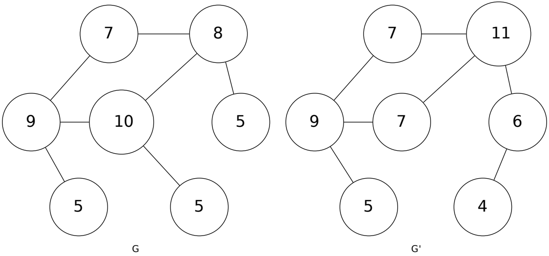
3,123 ⟶ 9

3,133 ⟶ 10

3,223 ⟶ 11

The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

Relabeling graphs G and G'

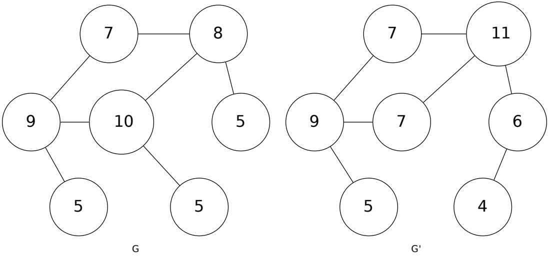


Label compression via hashing

1,2	→ 4	3,12	→ 8
1,3	→ 5	3,123	→ 9
2,13	→ 6	3,133	→ 10
2,33	→ 7	3,223	→ 11

The Weisfeiler-Lehman test of isomorphism [Weisfeiler 68]

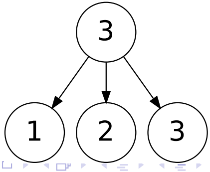
Relabeling graphs G and G'



Label compression via hashing

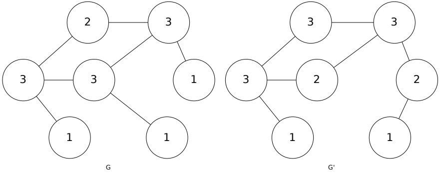
1,2	→ 4	3,12	→ 8
1,3	→ 5	3,123	→ 9
2,13	→ 6	3,133	→ 10
2,33	→ 7	3,223	→ 11

Subtree Pattern of Compressed label 9



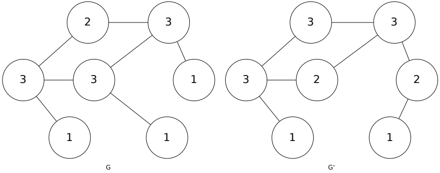
Weisfeiler-Lehman subtree features

Subtree patterns of depth 0.

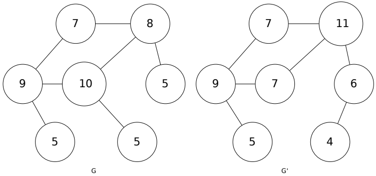


Weisfeiler-Lehman subtree features

Subtree patterns of depth 0.

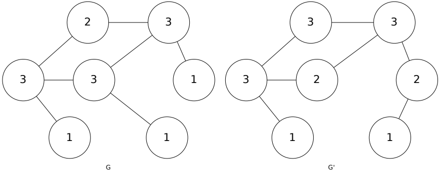


Subtree patterns of depth 1.

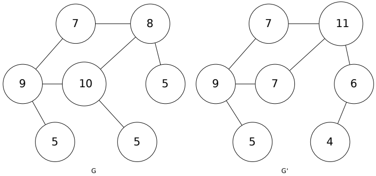


Weisfeiler-Lehman subtree features

Subtree patterns of depth 0.



Subtree patterns of depth 1.



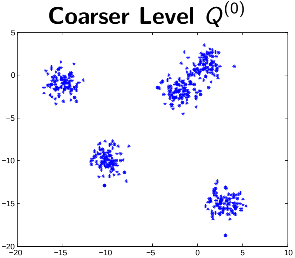
Original node labels Σ_0: 1, 2, 3 — Compressed node labels Σ_1: 4, …, 11

Labels $\{\Sigma_0, \Sigma_1\}$	=	{1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11}
$\phi_{(1)}(G)$	=	(3,	1,	3,	0,	3,	0,	1,	1,	1,	1,	0)
$\phi_{(1)}(G')$	=	(2,	2,	3,	1,	1,	1,	2,	0,	1,	0,	1)

$\phi_{(h)}(G)$ are histograms of occurences of the subtree patterns up to depth h in graph G.

The pyramid quantization strategy

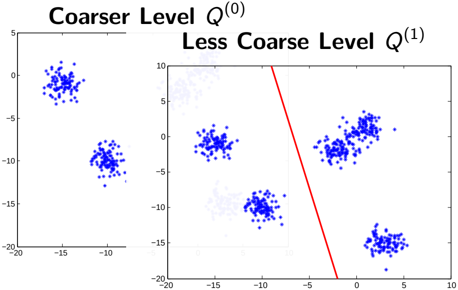
Given a set $\mathcal{G} = \{G_i = (V_i, E_i, \mathcal{L}_i)\}_{1 \leq i \leq n}$ where $\mathcal{L}_i : V_i \mapsto \mathbb{R}^d$



where $L = \lceil log_2 |V| \rceil$ and $|V| = \sum_i^n |V_i|$ [Grauman 07].

The pyramid quantization strategy

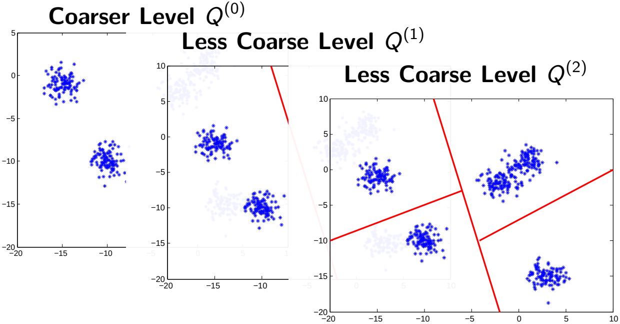
Given a set $\mathcal{G} = \{G_i = (V_i, E_i, \mathcal{L}_i)\}_{1 \leq i \leq n}$ where $\mathcal{L}_i : V_i \mapsto \mathbb{R}^d$



where $L = \lceil log_2 |V| \rceil$ and $|V| = \sum_i^n |V_i|$ [Grauman 07].

The pyramid quantization strategy

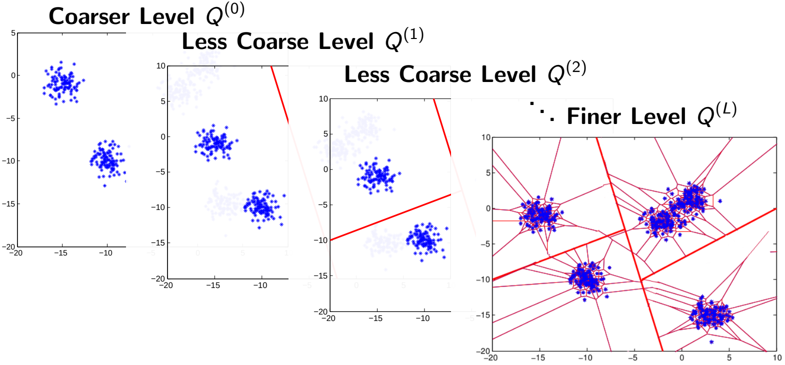
Given a set $\mathcal{G} = \{G_i = (V_i, E_i, \mathcal{L}_i)\}_{1 \leq i \leq n}$ where $\mathcal{L}_i : V_i \mapsto \mathbb{R}^d$



where $L = \lceil log_2 |V| \rceil$ and $|V| = \sum_i^n |V_i|$ [Grauman 07].

The pyramid quantization strategy

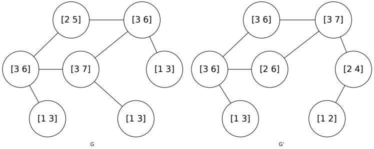
Given a set $\mathcal{G} = \{G_i = (V_i, E_i, \mathcal{L}_i)\}_{1 \leq i \leq n}$ where $\mathcal{L}_i : V_i \mapsto \mathbb{R}^d$



where $L = \lceil log_2 |V| \rceil$ and $|V| = \sum_i^n |V_i|$ [Grauman 07].

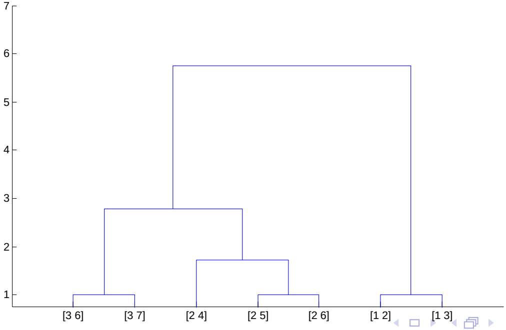
Data guided pyramid quantization scheme

Given labeled graphs G and G'



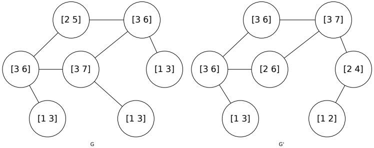
Notes

- Ward's minimum variance method over the image of V under $\mathcal{L}$.



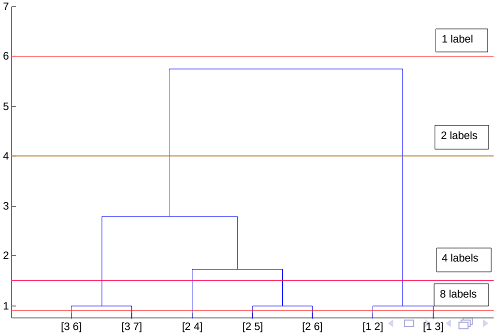
Data guided pyramid quantization scheme

Given labeled graphs G and G'



Notes

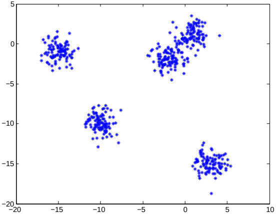
- Ward's minimum variance method over the image of V under $\mathcal{L}$.
- Selecting $L = \lceil log_2 D \rceil$, where $D \leq |V|$ the number of unique values in the image of V under $\mathcal{L}$
- Each level l has 2^l discrete labels.



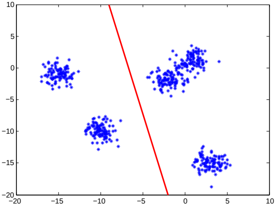
Transform the initial graphs as a sequence of graphs

The pyramid quantization of label space

Coarser Level $Q^{(0)}$

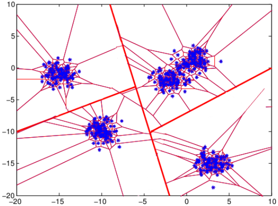


Less Coarse Level $Q^{(1)}$



. . .

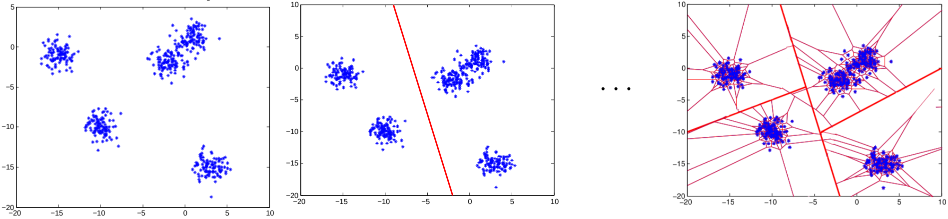
Finer Level $Q^{(L)}$



Transform the initial graphs as a sequence of graphs

The pyramid quantization of label space

Coarser Level $Q^{(0)}$ **Less Coarse Level $Q^{(1)}$** … **Finer Level $Q^{(L)}$**



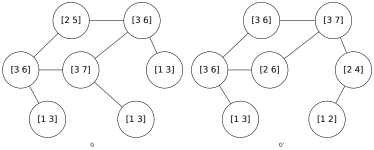
Sequence of discretely labeled graphs

$$G = (V, E, \mathcal{L}) \overset{Q^{(l)} \circ \mathcal{L}}{\underset{\forall l}{\approx}} \underbrace{\left(G^{(0)}, \dots, G^{(L)}\right)}_{\text{Increasing granularity} \longrightarrow} = \underbrace{\left((V, E, \mathcal{L}^{(0)}), \dots, (V, E, \mathcal{L}^{(L)})\right)}_{\text{Increasing granularity} \longrightarrow}$$

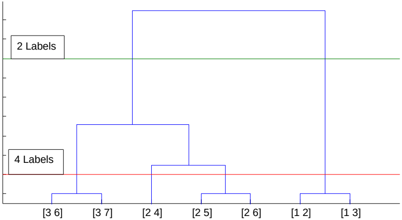
where $\mathcal{L}^{(l)} : V \rightarrow \Sigma_0^{(l)}$, $|\Sigma_0^{(l)}| = 2^l$ and $l \in \{0, \dots, L\}$.

A sequence of discretely labeled graphs

Given labeled graphs G and G'

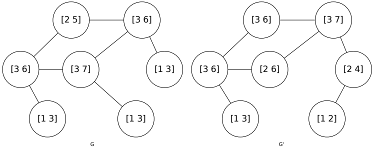


Data guided pyramid quantization.

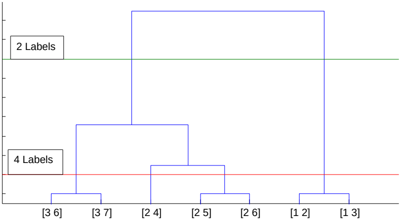


A sequence of discretely labeled graphs

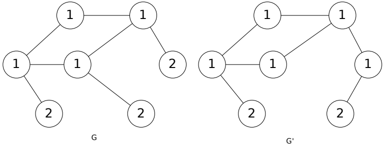
Given labeled graphs G and G'



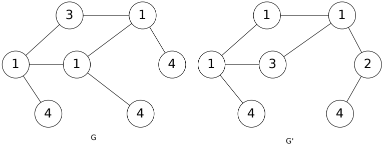
Data guided pyramid quantization.



Quantization level 1 with 2^1 number of labels.



Quantization level 2 with 2^2 number of labels



Creating and combining subtree features

Run Weisfeiler-Lehman on each quantization level

$$G = \left(G^{(0)}, \ldots, G^{(L)}\right) \xrightarrow[\text{Lehman}]{\text{Weisfeler}} \left(\phi^{(0)}_{(h)}(G^{(0)}), \ldots, \phi^{(L)}_{(h)}(G^{(L)})\right) = \widehat{\phi_{(h)}(G)}$$

where $\phi^{(l)}_{(h)}(G^{(l)})$ are histograms of occurences of the subtree patterns up to depth h at the quantization level l in graph G

Creating and combining subtree features

Run Weisfeiler-Lehman on each quantization level

$$G = \left(G^{(0)}, \ldots, G^{(L)}\right) \xrightarrow[Lehman]{Weisfeler} \left(\phi^{(0)}_{(h)}(G^{(0)}), \ldots, \phi^{(L)}_{(h)}(G^{(L)})\right) = \widehat{\phi_{(h)}(G)}$$

where $\phi^{(l)}_{(h)}(G^{(l)})$ are histograms of occurences of the subtree patterns up to depth h at the quantization level l in graph G

Learning to combine the quantization levels

1. Learn the selection of the subtree features $\widehat{\phi_{(h)}(G)}$.
2. Combine the subtree features $\phi^{(l)}_{(h)}(G^{(l)})$ per level l into a kernel and then learn the combination of kernels.

Learn the subtree patterns selection

- Labeled training data $\{(\widehat{\phi_{(h)}(G_i)}, y_i)\}_{1 \leq i \leq n} \in (\mathbb{N} \times \mathbb{R})^n$ where
  - $\widehat{\phi_{(h)}(G_i)}$ is the concatination of histograms of the occurences of subtree patterns up to depth h of graph G_i across all quantization levels,
  - y_i is the ground truth label and

Learn the subtree patterns selection

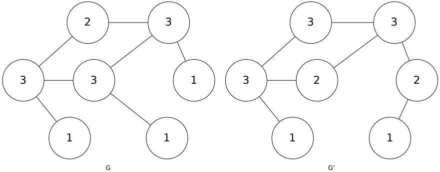
- Labeled training data $\{(\widehat{\phi_{(h)}(G_i)}, y_i)\}_{1\leq i\leq n} \in (\mathbb{N} \times \mathbb{R})^n$ where
  - $\widehat{\phi_{(h)}(G_i)}$ is the concatination of histograms of the occurences of subtree patterns up to depth h of graph G_i across all quantization levels,
  - y_i is the ground truth label and
- Elastic Net [Zou 05b]

$$\arg\min_{w\in\mathbb{R}^d} \underbrace{\lambda_1\|w\|_1}_{\ell_1 \text{ norm}} + \underbrace{\lambda_2\|w\|_2^2}_{\ell_2 \text{ norm}} + \underbrace{\frac{1}{n}\sum_{i=1}^n \left(\langle w, \widehat{\phi_{(h)}(G_i)}\rangle - y_i\right)^2}_{\text{Squared loss}}$$

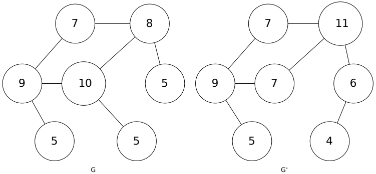
λ_1, λ_2 are scalar parameters controling the degree of regularization.

The intersection Weisfeiler-Lehman kernel

Subtree patterns of depth 0.



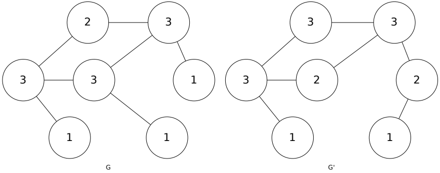
Subtree patterns of depth 1.



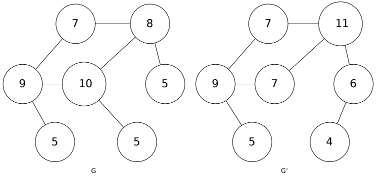
		Subtree patterns $h=0$			Subtree patterns $h=1$							
Labels $\{\Sigma_0, \Sigma_1\}$	$=$	$\{1,$	$2,$	$3,$	$4,$	$5,$	$6,$	$7,$	$8,$	$9,$	$10,$	$11\}$
$\phi^{(l)}_{(1)}(G^{(l)})$	$=$	$(3,$	$1,$	$3,$	$0,$	$3,$	$0,$	$1,$	$1,$	$1,$	$1,$	$0)$
$\phi^{(l)}_{(1)}(G'^{(l)})$	$=$	$(2,$	$2,$	$3,$	$1,$	$1,$	$1,$	$2,$	$0,$	$1,$	$0,$	$1)$

The intersection Weisfeiler-Lehman kernel

Subtree patterns of depth 0.



Subtree patterns of depth 1.



		Subtree patterns $h=0$			Subtree patterns $h=1$							
Labels $\{\Sigma_0, \Sigma_1\}$	=	{1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11}
$\phi^{(l)}_{(1)}(G^{(l)})$	=	(3,	1,	3,	0,	3,	0,	1,	1,	1,	1,	0)
$\phi^{(l)}_{(1)}(G'^{(l)})$	=	(2,	2,	3,	1,	1,	1,	2,	0,	1,	0,	1)
$\min\left(\phi^{(l)}_{(1)}(G^{(l)}), \phi^{(l)}_{(1)}(G'^{(l)})\right)$	=	(2,	1,	3,	0,	1,	0,	1,	0,	1,	0,	0)

The intersection Weisfeile-Lehman kernel is defined :

$$k^{(h)}_{i-WLsubtree}(G^{(l)}, G'^{(l)}) = \sum_{j}^{|\Sigma_0 \cup \Sigma_1|} \min\left(\phi^{(l)}_{(1)}(G^{(l)}), \phi^{(l)}_{(1)}(G'^{(l)})\right)_j$$

Multiple Kernel Learning

Problem

For each pair of graphs $G^{(l)}, G'^{(l)}$ for all the pyramid levels:

$$\left(K_{(h)}^{(0)}(G^{(0)}, G'^{(0)}), \ldots, K_{(h)}^{(L)}(G^{(l)}, G'^{(L)})\right)$$

we would like to learn a linear combination of them:

$$K_{(h)}(G, G') = \sum_{l=0}^{L} d_l K_{(h)}^{(l)}(G^{(l)}, G'^{(l)}), \text{ with } d_l \geq 0, \sum_{l=0}^{L} d_l = 1.$$

Multiple Kernel Learning

Problem

For each pair of graphs $G^{(l)}, G'^{(l)}$ for all the pyramid levels:

$$\left(K^{(0)}_{(h)}(G^{(0)}, G'^{(0)}), \ldots, K^{(L)}_{(h)}(G^{(l)}, G'^{(L)}) \right)$$

we would like to learn a linear combination of them:

$$K_{(h)}(G, G') = \sum_{l=0}^{L} d_l K^{(l)}_{(h)}(G^{(l)}, G'^{(l)}), \text{ with } d_l \geq 0, \sum_{l=0}^{L} d_l = 1.$$

Solutions

- Multiple kernel learning
- Average weighted kernel

Table of Content

fMRI Analysis

Key information

1. Inherent spatial structure brains
2. Voxel activation is a continuous value

fMRI Analysis

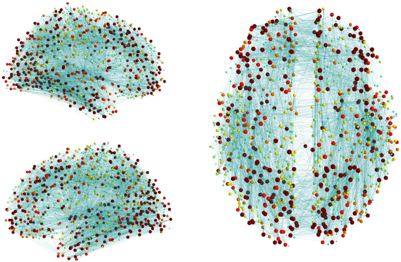
Key information

1. Inherent spatial structure brains
2. Voxel activation is a continuous value

⇓

Solution!

Represent fMRI as graphs with continuous labels.



Dataset

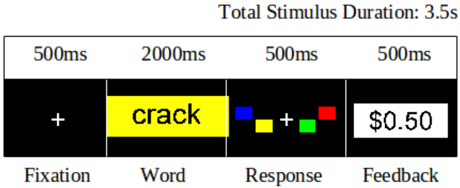
Cocaine Addiction Dataset

- 16 cocaine addicted vs 17 healthy subjects
- Drugstroop experiment with two varying conditions
  - the cue shown and
  - the monetary reward.

Input One contrast map per subject that is transformed into a graph.

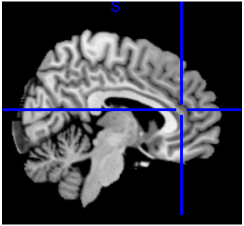
Objective The classification between cocaine abuser and control group.

Drugstroop Experiment



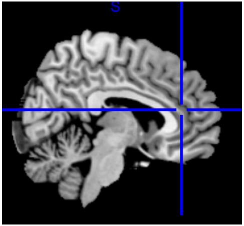
Graph Transformation

Contrast map



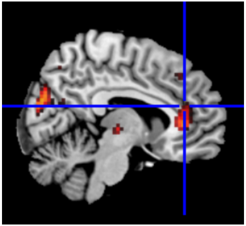
Graph Transformation

Contrast map



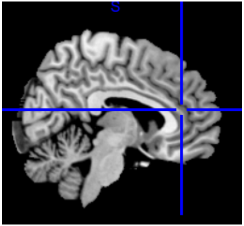
Elastic Net $\longrightarrow$

Selected voxels



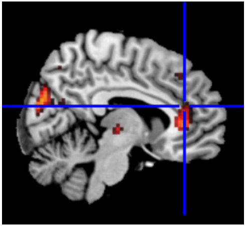
Graph Transformation

Contrast map

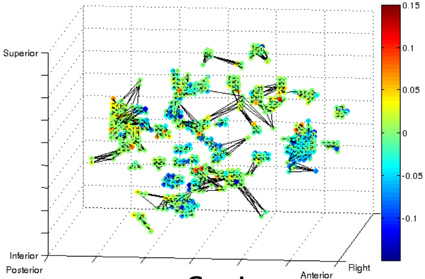


Elastic Net →

Selected voxels

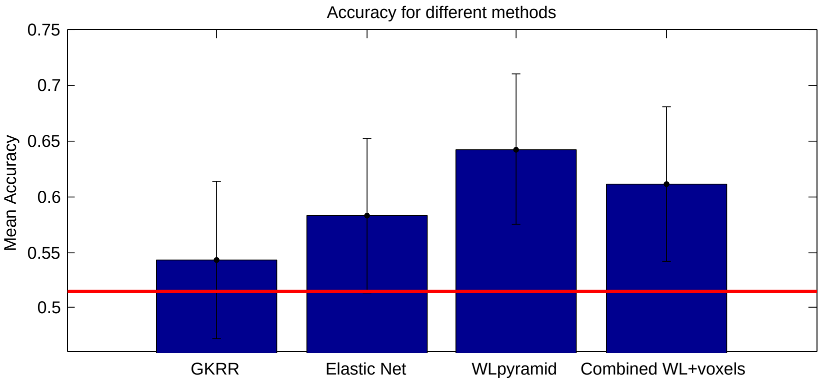


knn ↙

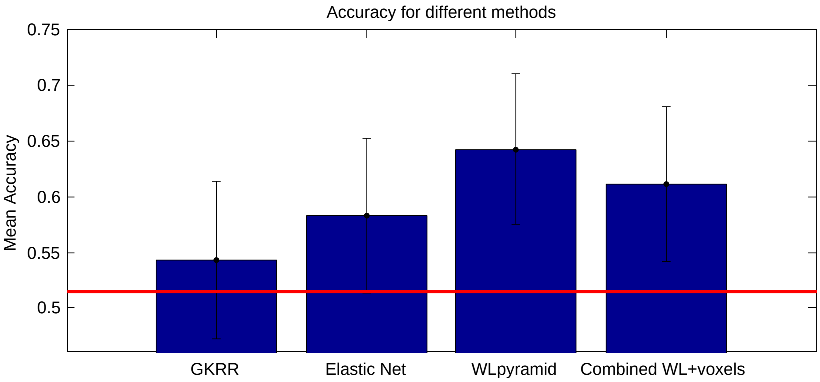


Graph

Performance



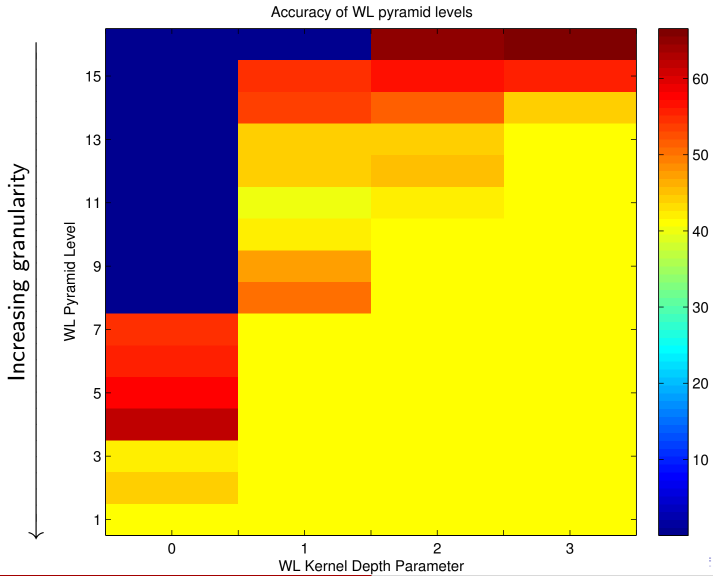
Performance



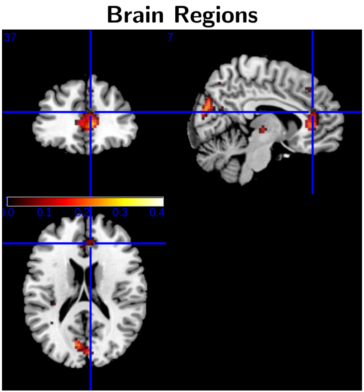
WLpyramid vs Elastic Net on raw voxels

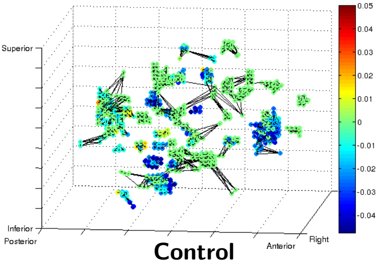
Wilcoxon signed rank with $p = 0.02$ show statistical significance.

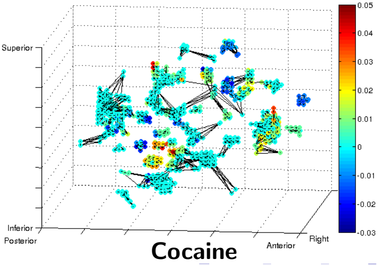
Performance per pyramid quantization level



Visualization of learned function



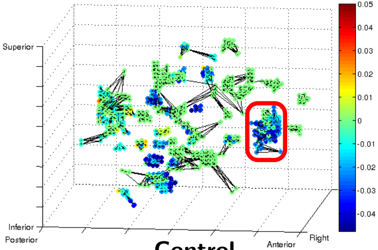




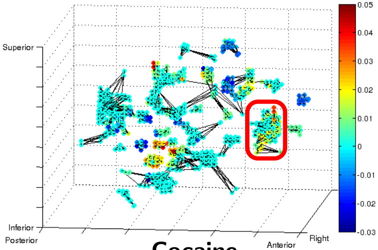
Visualization of learned function

Rostral Anterior Cingulate Cortex

- In cocaine addicted subjects deactivates during the drug Stroop experiment as compared to baseline.
- Its activity is normalized by oral methylphenidate where the dopamine transporters increase the extracellular dopamine, an increase which is associated with lower task-related impulsivity.



Control

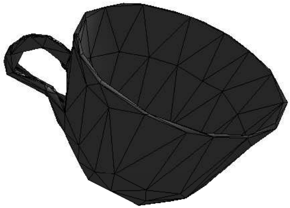


Cocaine

3D shape classification

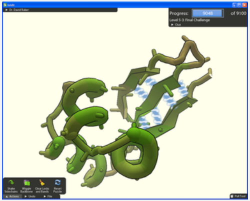
3D shape problems

- Storage
- Classification
- Retrieval

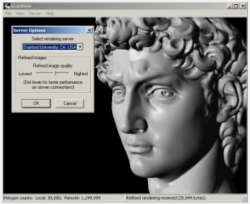


Areas of applications

3D Game

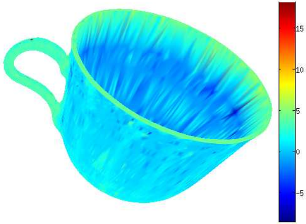


Chemoinformatics



Cultural heritage

3D shape classification

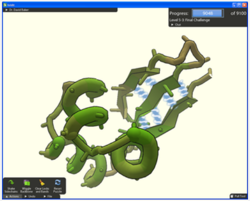


3D shape problems

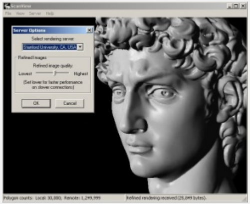
- Storage
- Classification
- Retrieval

Areas of applications

3D Game



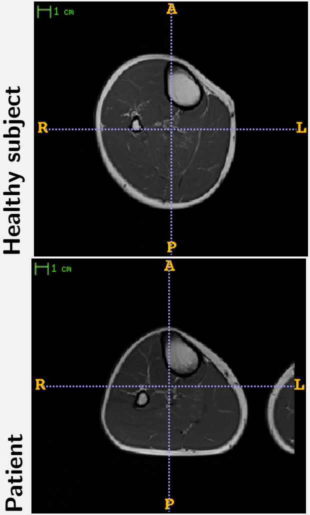
Chemoinformatics



Cultural heritage

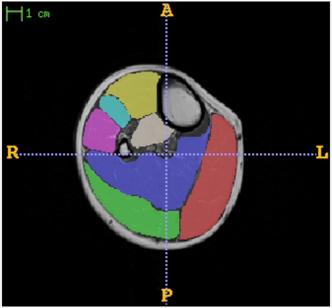
3D shape datasets

Muscle Dataset



3D shape datasets

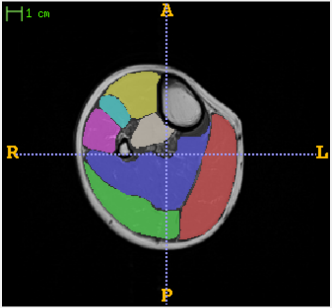
Muscle Dataset



- 27 patients vs 14 healthy subjects
- MRI images of the calf muscles
- Segmented into 7 muscles

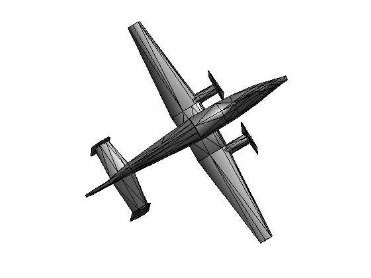
3D shape datasets

Muscle Dataset



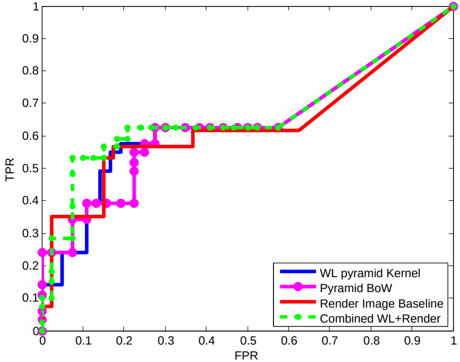
- 27 patients vs 14 healthy subjects
- MRI images of the calf muscles
- Segmented into 7 muscles

SHREC 2013 Dataset



- 20 classes of generic objects, such as bed, biplane, mug, etc.
- Each class contains 18 models.
- In total 360 3D objects.

Performance on the muscle dataset



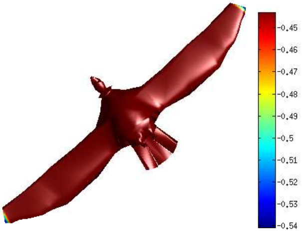
	WLpyramid Our Work	pyramid BoW	Rendering	Combined
Accuracy	78.00%	73.00%	75.50%	82.93%
AUC	0.6410	0.6361	0.6300	0.6648

Performance on SHREC 2013

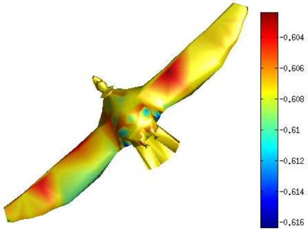
Class	WLpyramid Our Work	pyramid BoW	Rendering	Combined
Bird	0.85	0.83	0.85	**0.86**
Bicycle	0.84	0.87	0.90	0.90
Biped	0.89	0.88	0.99	0.99
Biplane	0.60	0.63	0.68	**0.69**
Bird	0.73	0.73	0.80	0.80
Bottle	0.76	0.76	0.79	**0.80**
Car	0.78	0.79	0.80	0.80
CellPhone	0.74	0.80	0.88	**0.89**
Chair	0.69	0.68	0.70	**0.72**
Cup	0.85	0.84	0.88	0.88
Desklamp	0.80	0.80	0.88	**0.89**
Fish	1.00	1.00	1.00	1.00
Floorlamp	0.80	0.77	0.89	0.89
Insect	0.64	0.60	0.62	**0.66**
Monoplane	0.84	0.82	0.88	**0.90**
Mug	0.82	0.82	0.85	**0.87**
Phone	0.83	0.74	0.72	**0.83**
Quadruped	0.89	0.86	0.97	**0.98**
Sofa	0.76	0.75	0.74	**0.75**
Wheelchair	0.81	0.79	0.88	**0.90**
Average	0.80	0.79	0.84	0.85

SHREC 2013 - Visualization of the learned weights

Subtree patterns of depth 0



Subtree patterns up to depth 2



Subtree patterns up to depth 1

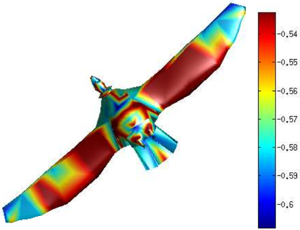


Table of Content

fMRI analysis and regularization methods

Regularizer	$\lambda\Omega(w)$
LASSO [Tibshirani 96]	$\lambda_1\|w\|_1$

where

- λ is a scalar controlling the degree of regularization,
- $|w|_i^{\downarrow}$ is the ith largest element of the vector $|w|$,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that correlates with the cardinality of w and
- r is the unique integer in $\{0, \ldots, k-1\}$ automatically selected by the algorithm.

fMRI analysis and regularization methods

Regularizer	$\lambda\Omega(w)$
LASSO [Tibshirani 96]	$\lambda_1\|w\|_1$
Elastic Net [Zou 05a]	$\lambda_1\|w\|_1 + \lambda_2\|w\|_2^2$

where

- λ is a scalar controlling the degree of regularization,
- $|w|_i^{\downarrow}$ is the ith largest element of the vector $|w|$,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that correlates with the cardinality of w and
- r is the unique integer in $\{0, \ldots, k-1\}$ automatically selected by the algorithm.

fMRI analysis and regularization methods

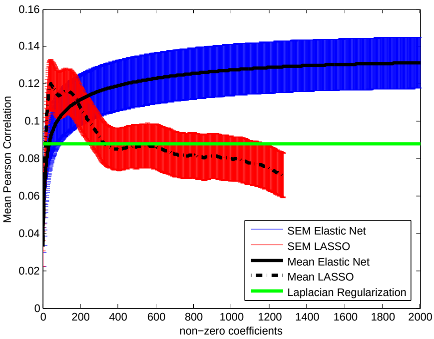
Regularizer	$\lambda\Omega(w)$
LASSO [Tibshirani 96]	$\lambda_1\|w\|_1$
Elastic Net [Zou 05a]	$\lambda_1\|w\|_1 + \lambda_2\|w\|_2^2$
k-support norm [Argyriou 12]	$\lambda\left(\overbrace{\sum_{i=1}^{k-r-1}(\lvert w\rvert_i^{\downarrow})^2}^{\ell_2 \text{ norm}} + \frac{1}{r+1}\overbrace{\left(\sum_{i=k-r}^{d}\lvert w\rvert_i^{\downarrow}\right)^2}^{\ell_1 \text{ norm}}\right)^{\frac{1}{2}}$

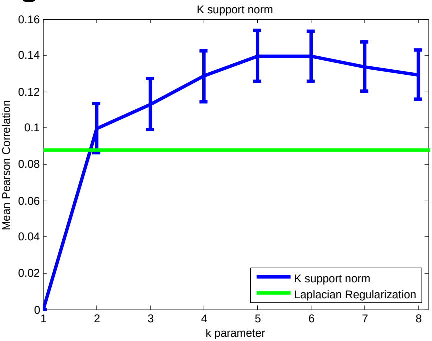
where

- λ is a scalar controlling the degree of regularization,
- $|w|_i^{\downarrow}$ is the ith largest element of the vector $|w|$,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that correlates with the cardinality of w and
- r is the unique integer in $\{0, \ldots, k-1\}$ automatically selected by the algorithm.

Quantitative Results

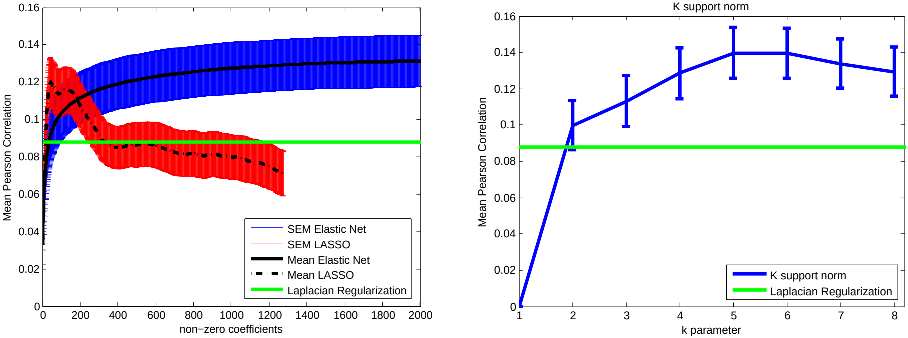
Picture viewing dataset





Quantitative Results

Picture viewing dataset



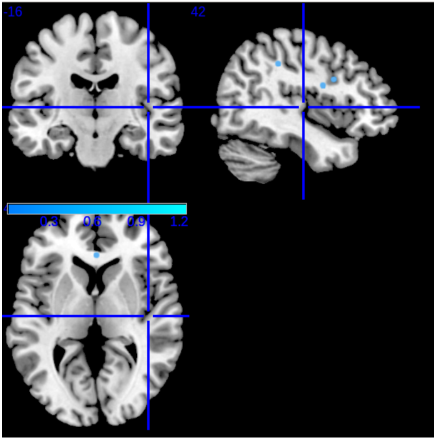
k-support norm vs LASSO and Elastic Net

Wilcoxon signed rank test with $p = 0.05$ show statistical significance.

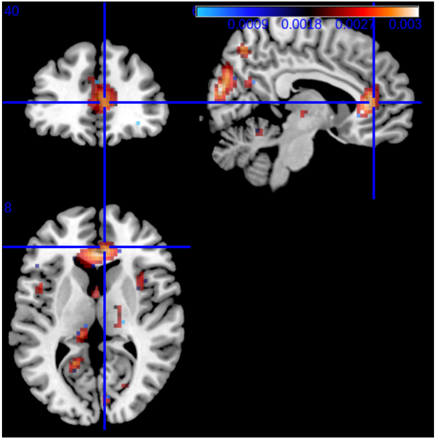
Qualitative Results

Cocaine addiction dataset

LASSO

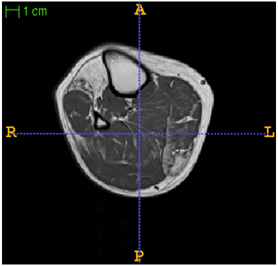


***k*-support norm**

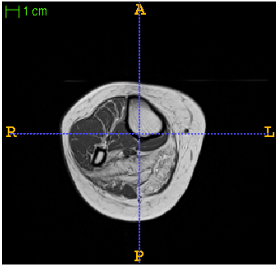


Neuromuscular disease classification

FSH

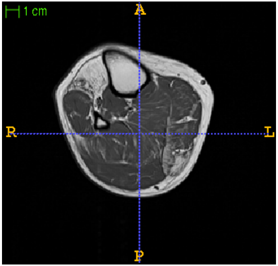


DM1

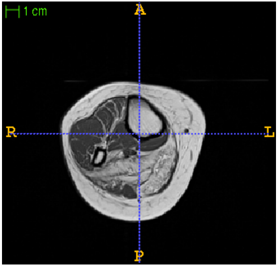


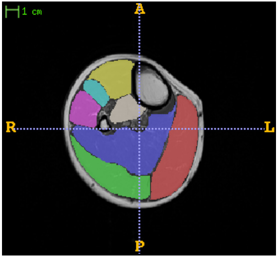
Neuromuscular disease classification

FSH



DM1





- the mean T1/T2 signal,
- the Signal to Noise Ration,
- the Fractional Anisotropy,
- the trace of the diffusion tensor,
- the volume of the tensor, etc.

k-support regularized SVM

The Support Vector Machine is defined as the following optimization problem:

$$\begin{aligned} \min_{w\in\mathbb{R}^d, b\in\mathbb{R}, \xi\in\mathbb{R}^n} \quad & \lambda\|w\|_2^2 + \sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & y_i\left(\langle w, x_i\rangle + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i. \end{aligned}$$

where

- λ is a scalar, user supplied parameter controling the degree of regularization,

k-support regularized SVM

The k-support norm regularized SVM ($k$sup-SVM) is defined as the following optimization problem:

$$\begin{aligned} \min_{w\in\mathbb{R}^d, b\in\mathbb{R}, \xi\in\mathbb{R}^n} \quad & \lambda \|w\|_k^{sp} + \sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & y_i\left(\langle w, x_i\rangle + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i. \end{aligned}$$

where

- λ is a scalar, user supplied parameter controling the degree of regularization,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that negative correlates with the cardinality of w and
- $\|w\|_k^{sp}$ is the k-support penalty.

k-support regularized SVM

The k-support norm regularized SVM ($k$sup-SVM) is defined as the following optimization problem:

$$\min_{w\in\mathbb{R}^d, b\in\mathbb{R}, \xi\in\mathbb{R}^n} \quad \lambda \|w\|_k^{sp} + \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad y_i\left(\langle w, x_i\rangle + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i.$$

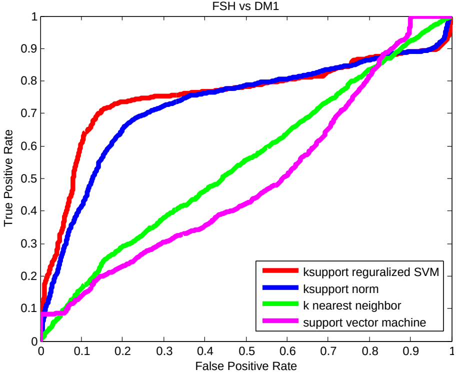
where

- λ is a scalar, user supplied parameter controling the degree of regularization,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that negative correlates with the cardinality of w and
- $\|w\|_k^{sp}$ is the k-support penalty.

Advantage

Solution is sparse but correlated subset of discriminative variables.

Performance

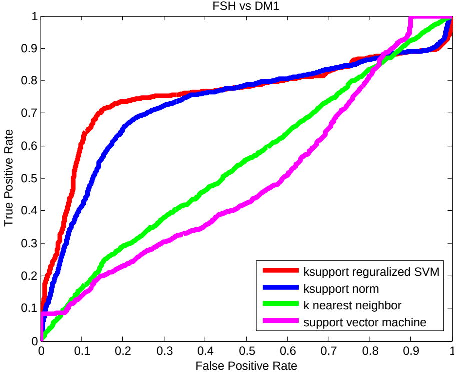


Method	AUC
*k*sup-SVM - Our work	0.756
k-support with SL	0.726
*k*nn	0.537
SVM	0.494

Method	Accuracy
*k*sup-SVM - Our work	77 ± 0.013
k-support with SL	74 ± 0.006
*k*nn	61 ± 0.015
SVM	59 ± 0.015

Chance is 60%.

Performance



Method	AUC
*k*sup-SVM - Our work	0.756
k-support with SL	0.726
*k*nn	0.537
SVM	0.494

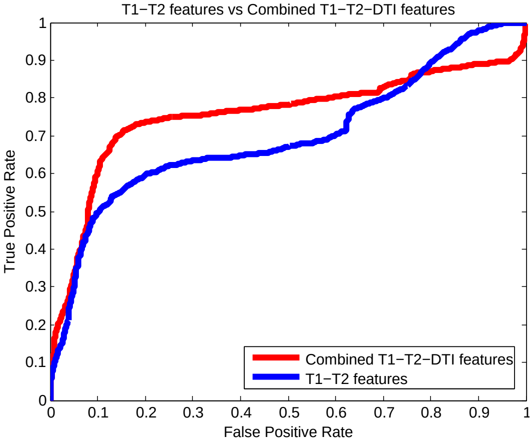
Method	Accuracy
*k*sup-SVM - Our work	77 ± 0.013
k-support with SL	74 ± 0.006
*k*nn	61 ± 0.015
SVM	59 ± 0.015

Chance is 60%.

*k*sup-SVM vs the rest methods

Wilcoxon signed rank test with $p \ll 10^{-9}$ show statistical significance.

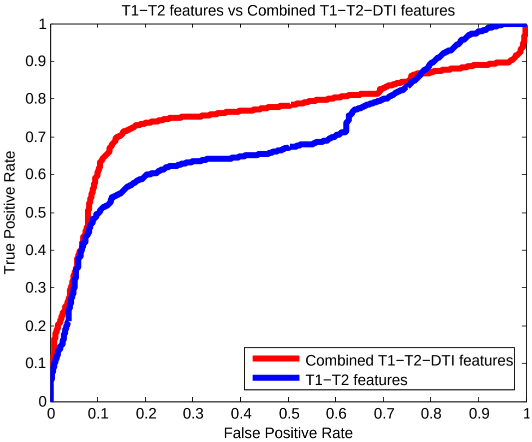
Structured and DTI features vs Structured features only



Features Used	AUC
MRI and DTI	0.756
MRI only	0.697

Features Used	Accuracy
MRI and DTI	77 $\pm$ 0.013
MRI only	73 $\pm$ 0.006

Structured and DTI features vs Structured features only



Features Used	AUC
MRI and DTI	0.756
MRI only	0.697

Features Used	Accuracy
MRI and DTI	77 ± 0.013
MRI only	73 ± 0.006

Structured and DTI features vs Structured features only

Wilcoxon signed rank test with $p \ll 0.05$ show statistical significance.

Table of Content

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.

k-support regularized SVM

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.

k-support regularized SVM

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.
- Linear computation time in the number of graphs, in the number of edges in the graphs and in the depth of subtree patterns.

k-support regularized SVM

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.
- Linear computation time in the number of graphs, in the number of edges in the graphs and in the depth of subtree patterns.
- Evaluation on two domains

k-support regularized SVM

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.
- Linear computation time in the number of graphs, in the number of edges in the graphs and in the depth of subtree patterns.
- Evaluation on two domains
  - fMRI analysis and

k-support regularized SVM

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.
- Linear computation time in the number of graphs, in the number of edges in the graphs and in the depth of subtree patterns.
- Evaluation on two domains
  - fMRI analysis and
  - 3D shape classification.

k-support regularized SVM

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.
- Linear computation time in the number of graphs, in the number of edges in the graphs and in the depth of subtree patterns.
- Evaluation on two domains
  - fMRI analysis and
  - 3D shape classification.
- Visualizations of the learned functions provide interpretability.

k-support regularized SVM

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.
- Linear computation time in the number of graphs, in the number of edges in the graphs and in the depth of subtree patterns.
- Evaluation on two domains
  - fMRI analysis and
  - 3D shape classification.
- Visualizations of the learned functions provide interpretability.

k-support regularized SVM

- A novel regularized SVM algorithm.

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.
- Linear computation time in the number of graphs, in the number of edges in the graphs and in the depth of subtree patterns.
- Evaluation on two domains
  - fMRI analysis and
  - 3D shape classification.
- Visualizations of the learned functions provide interpretability.

k-support regularized SVM

- A novel regularized SVM algorithm.
- Correlated sparse solution under the SVM framework.

Contributions - Methodological

The *pyramid quantized Weisfeiler-Lehman graph representation*

- A novel algorithm for comparing graphs with vector labels.
- Based on subtree patterns.
- Linear computation time in the number of graphs, in the number of edges in the graphs and in the depth of subtree patterns.
- Evaluation on two domains
  - fMRI analysis and
  - 3D shape classification.
- Visualizations of the learned functions provide interpretability.

k-support regularized SVM

- A novel regularized SVM algorithm.
- Correlated sparse solution under the SVM framework.
- Evaluation on a neuromuscular disease task.

Contributions

Methodological

Code from both algorithms is available online under GNU-GPL at
http://cvc.centrale-ponts.fr/personnel/gkirtzou/code

Contributions

Methodological

Code from both algorithms is available online under GNU-GPL at
`http://cvc.centrale-ponts.fr/personnel/gkirtzou/code`

Clinical and Applications

- Investigate the applicability of sparsity regularizers in fMRI analysis.

Contributions

Methodological

Code from both algorithms is available online under GNU-GPL at
`http://cvc.centrale-ponts.fr/personnel/gkirtzou/code`

Clinical and Applications

- Investigate the applicability of sparsity regularizers in fMRI analysis.
- In the fMRI analysis, we saw that the interconnections between voxels can contain additional information about brain structure.

Contributions

Methodological

Code from both algorithms is available online under GNU-GPL at
`http://cvc.centrale-ponts.fr/personnel/gkirtzou/code`

Clinical and Applications

- Investigate the applicability of sparsity regularizers in fMRI analysis.
- In the fMRI analysis, we saw that the interconnections between voxels can contain additional information about brain structure.
- In the neuromuscular dystrophy classification task, we saw that features extracted from DTI images provide significant information.

Contributions

Methodological

Code from both algorithms is available online under GNU-GPL at
`http://cvc.centrale-ponts.fr/personnel/gkirtzou/code`

Clinical and Applications

- Investigate the applicability of sparsity regularizers in fMRI analysis.
- In the fMRI analysis, we saw that the interconnections between voxels can contain additional information about brain structure.
- In the neuromuscular dystrophy classification task, we saw that features extracted from DTI images provide significant information.
- Interpretation of 3D shape meshes as annotated graphs.

Future Work

Medical image analysis

- Evaluation of k-support norm regularization on fMRI analysis problem in larger scale.
- Evaluation of k-support regularized SVM on neuromuscular disease discrimination in larger scale.
- Exploration of different constructions of the graphs from fMRI.

Future Work

Medical image analysis

- Evaluation of k-support norm regularization on fMRI analysis problem in larger scale.
- Evaluation of k-support regularized SVM on neuromuscular disease discrimination in larger scale.
- Exploration of different constructions of the graphs from fMRI.

Graph kernels

- Comparison on partially matching subtree patterns.
- Comparison on partially labeled graphs.

Publications

- K. Gkirtzou, and M. Blaschko
  The pyramid quantization Weisfeiler-Lehman graph representation
  Submitted to Pattern Recognition

- K. Gkirtzou, J. Honorio, D. Samaras, R. Goldstein and M. Blaschko
  FMRI analysis of cocaine addiction using k-support sparsity
  In International Symposium on Biomedical Imaging 2013
  Oral Presentation - 19% acceptance rate

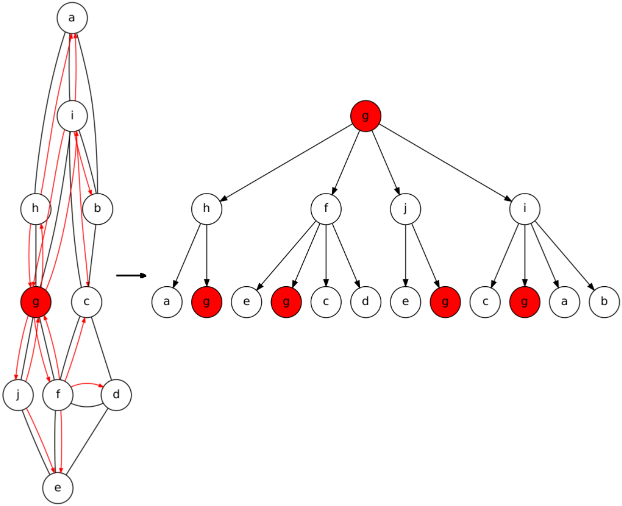
- K. Gkirtzou, DJ. Francois, G. Bassez, A. Sotiras, A.Rahmouni, T. Varacca, N. Paragios and M. Blaschko
  Sparse classification with MRI based markers for neuromuscular disease classification.
  In LNCS series of Machine Learning in Medical Imaging 2013
  Oral Presentation - 26% acceptance rate

- K. Gkirtzou, J. Honorio, D. Samaras, R. Goldstein and M. Blaschko
  fMRI Analysis with Sparse Weisfeiler-Lehman Graph Statistics
  In LNCS series of Machine Learning in Medical Imaging 2013
  Poster Presentation - 56% acceptance rate

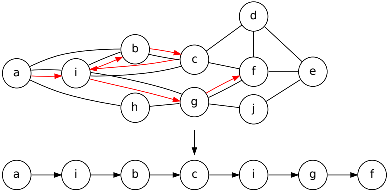
Table of Contents - Appendix

Table of Content

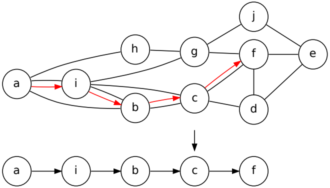
Subtree patterns



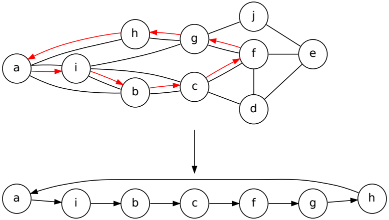
Walk, path and cycle on graph



(a) Graph Walk



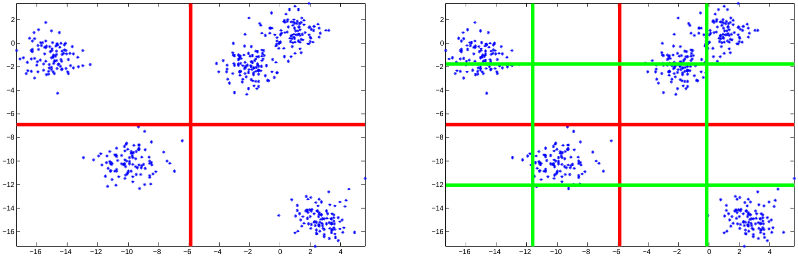
(b) Graph Path



(c) Graph Cycle

Binning strategies

Fixed binning



Data guided binning

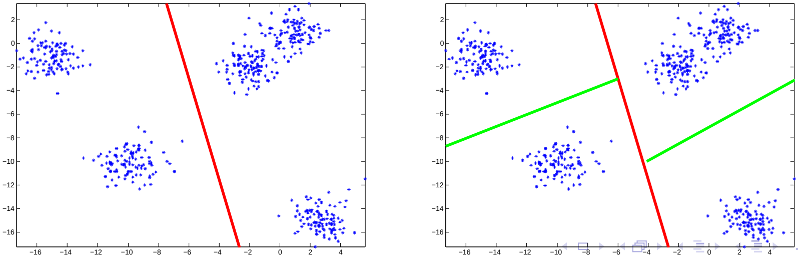


Table of Content

fMRI Analysis of Cocaine Addiction using k-support sparsity

Exploring regularization techniques for fMRI analysis

K. Gkirtzou[1,2], Jean Honorio[3], Dimitris Samaras[1,3], Rita Goldstein[4], Matthew B. Blaschko[1,2]

[1] CENTRALE PARIS [2] Inria INVENTORS FOR THE DIGITAL WORLD [3] STONY BROOK UNIVERSITY [4] Mount Sinai

10th April, 2013

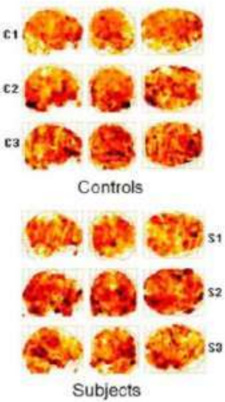
fMRI Analysis

Goal of fMRI analysis

The goal of fMRI data analysis is to detect correlations between brain activation and a task the subject performs during the scan.

Common problems in fMRI analysis

- high-dimensional space
- small number of samples
- high levels of noise



Related Work

Previous Related Work

- Generalized Linear Model
- Support vector machines [Song 11]
- Kernel canonical correlation analysis [Hardoon 07, Blaschko 09, Blaschko 11].
- Independent component analysis [Bartels 04, Bartels 05]
- Regression models (OLS, Ridge Regression, LASSO, Elastic Net) [Carroll 09, Ng 12].

Related Work

Previous Related Work

- Generalized Linear Model
- Support vector machines [Song 11]
- Kernel canonical correlation analysis [Hardoon 07, Blaschko 09, Blaschko 11].
- Independent component analysis [Bartels 04, Bartels 05]
- Regression models (OLS, Ridge Regression, LASSO, Elastic Net) [Carroll 09, Ng 12].

Regularized methods explored in this work

- LASSO [Tibshirani 96]
- Elastic Net [Zou 05a]
- k-support norm [Argyriou 12]

Sparsity regularization - Mathematical framework

- Labeled training data $\{(x_1, y_1), \ldots, (x_n, y_n)\} \in \left(\mathbb{R}^d \times \mathbb{R}\right)^n$
  - x_i is the output of a fMRI scan
  - y_i is the ground truth label
- Loss function

$$\arg\min_{w \in \mathbb{R}^d} \lambda\Omega(w) + \frac{1}{n}\sum_{i=1}^{n}(\langle w, x_i \rangle - y_i)^2$$

  - λ is a scalar parameter controling the degree of regularization
  - Ω is a scalar valued function monotonic in a norm of w.

Sparsity regularization - Mathematical framework

- Labeled training data $\{(x_1, y_1), \ldots, (x_n, y_n)\} \in \left(\mathbb{R}^d \times \mathbb{R}\right)^n$
  - x_i is the output of a fMRI scan
  - y_i is the ground truth label
- Loss function

$$\arg\min_{w \in \mathbb{R}^d} \lambda\Omega(w) + \frac{1}{n}\sum_{i=1}^{n}(\langle w, x_i \rangle - y_i)^2$$

Penalty function

Regularizer	$\lambda\Omega(w)$
LASSO [Tibshirani 96]	$\lambda_1\|w\|_1$
Elastic net [Zou 05a]	$\lambda_1\|w\|_1 + \lambda_2\|w\|_2^2$

k support norm - Penalty function

The k-support norm [Argyriou 12] can be computed as

$$\lambda\Omega(w) = \lambda\|w\|_k^{sp} = \lambda \left(\overbrace{\sum_{i=1}^{k-r-1} (|w|_i^{\downarrow})^2}^{\ell_2 \text{ norm}} + \frac{1}{r+1} \overbrace{\left(\sum_{i=k-r}^{d} |w|_i^{\downarrow} \right)^2}^{\ell_1 \text{ norm}} \right)^{\frac{1}{2}}$$

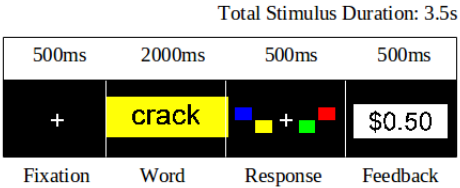
where

- λ is a scalar, user supplied parameter controling the degree of regularization,
- $|w|_i^{\downarrow}$ is the ith largest element of the vector $|w|$,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that correlates with the cardinality of w and
- r is the unique integer in $\{0, \ldots, k-1\}$ automatically selected by the algorithm.

Dataset 1

Cocaine Addiction Dataset

- 16 cocaine addicted vs 17 control subjects
- Drugstroop experiment with two varying conditions
  - the cue shown
  - the monetary reward
- Using one contrast map per subject
- The discriminative task is the classification between cocaine abuser and control group



Drugstroop Experiment

Dataset 2

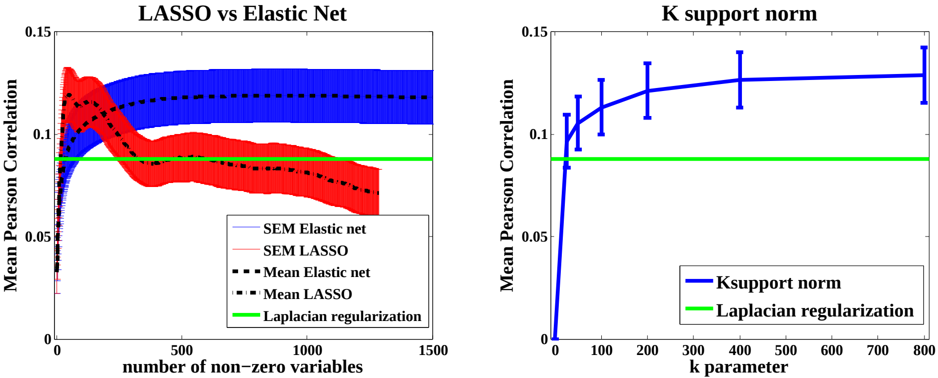
Natural viewing dataset

- A healthy subject in a free-viewing setting
- Using the whole time series
- The discriminative task is the prediction of a "Temporal Contrast" variable

Experimental setup

- 100 random permutations trials
- Training on the 80% of data
- Testing on the rest 20% of data

Quantitative results

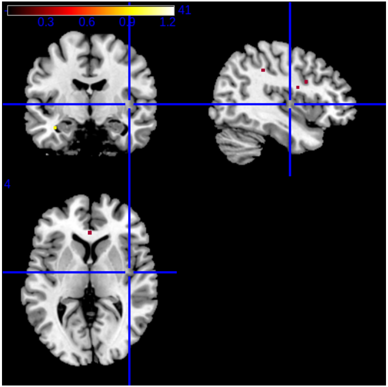


Natural viewing dataset

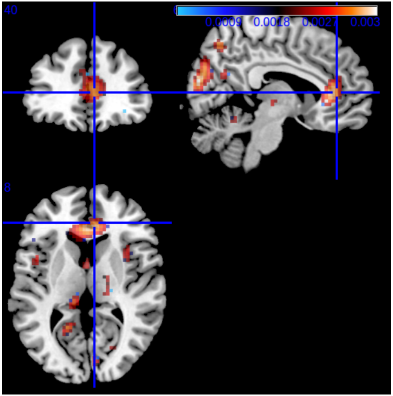
Elastic Net vs *k*-support norm

Wilcoxon signed rank test with p-value $\ll 0.05$

Qualitative results



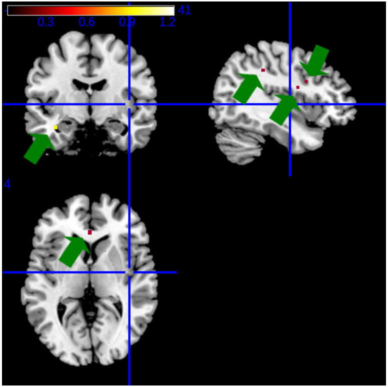
LASSO



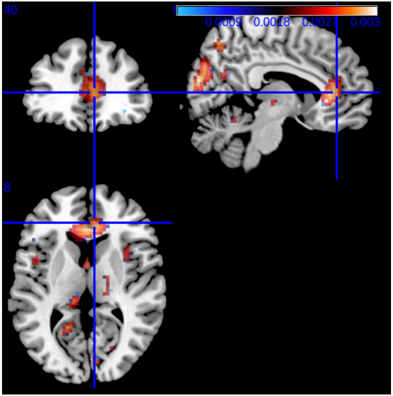
k-support norm

Cocaine addiction dataset

Qualitative results



LASSO



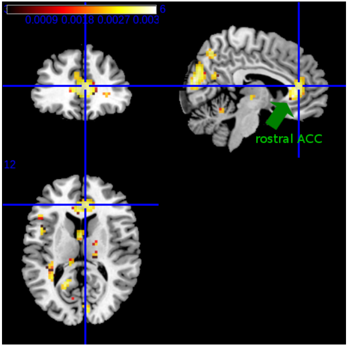
k-support norm

Cocaine addiction dataset

Rostral Anterior Cingulate Cortex

Rostral Anterior Cingulate Cortex

- In cocaine addicted subjects deactivates during the drug Stroop experiment as compared to baseline.
- Its activity is normalized by oral methylphenidate where the dopamine transporters increase the extracellular dopamine, an increase which is associated with lower task-related impulsivity.
- In cigarette smokers was responsive to pharmacotherapeutic interventions.
- In depression may be a marker of treatment response.



k-support norm

Conclusions - Future Work

Conclusions

- The k-support norm can boost the predictive performance of the LASSO and elastic net.
- The LASSO does not show a meaningful sparsity pattern.
- The brain regions implicated in addiction by the k-support norm coincide with previous results on addiction.
- Code available online :: `http://www.centrale-ponts.fr/personnel/gkirtzou/code/`

Conclusions - Future Work

Conclusions

- The k-support norm can boost the predictive performance of the LASSO and elastic net.
- The LASSO does not show a meaningful sparsity pattern.
- The brain regions implicated in addiction by the k-support norm coincide with previous results on addiction.
- Code available online :: `http://www.centrale-ponts.fr/personnel/gkirtzou/code/`

Future Work

- Exploring the structural information of the brain

Questions

Acknowledgements

- This work is partially funded by
    - the European Research Council under the Seventh Framework Programme (FP7/2007-2013)/ERC Grant 259112.
    - NIDA R21 DA034954 under the SUBSample project from the DIGITEO Institute, France.
    - "Machine learning discovery of patterns of self-regulation in drug addiction and intermittent explosive disorder" NIA 1R21DA034954-01
- We thank A. Bartels for providing data.

Table of Content

Sparse classification with MRI based markers for neuromuscular disease categorization

Exploring regularization techniques for fMRI analysis

Katerina Gkirtzou[1,2], Jean-François Deux[3], Guillaume Bassez[3], Aristeidis Sotiras[4], Alain Rahmouni[3], Thibault Varacca[3], Nikos Paragios[1,2], Matthew B. Blaschko[1,2]

1 CENTRALE PARIS 2 Inria INVENTORS FOR THE DIGITAL WORLD 3 hm GROUPE HOSPITALIER HENRI MONDOR ALBERT CHENEVIER - GEORGES CLEMENCEAU JOFFRE-DUPUYTREN - EMILE ROUX 4 Penn UNIVERSITY of PENNSYLVANIA

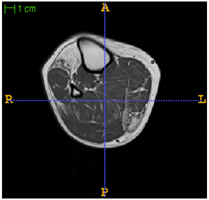
22th September, 2013

Neuromuscular Diseases

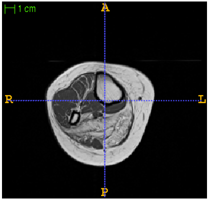
T1-weighted MR images of the calf.

Problem

- Myopathies are neuromuscular diseases that result functional anomalies including
  - fat infiltration
  - atrophy
  - weakness of the muscle and
  - paralysis.
- In this study, we focus on categorization patients between Facioscapulohumeral muscular dystrophy (FSH) and myotonic muscular dystrophy type 1 (DM1) using MRI based markers.



(d) FSH



(e) DM1

Our approach and Related Work

Our approach

1. Features extracted from both structured MR Imaging (T1 and T2 weighted) and Diffusion Tensor Imaging
2. a novel structured sparsity algorithm, the *k*-support regularized SVM (*k*sup-SVM)

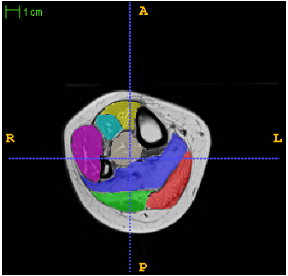
Related Work

- DTI on Neuroimaging studies
  - Alzheimer's disease [Klöppel 08]
  - male-female or older-younger classification [Lao 04]
  - temporal classification of block design fMRI data [LaConte 05]
  - the study of autism spectrum disorder [Ingalhalikar 11]
- DTI on different clinical scenarios
  - the human tongue [Gilbert 05]
  - the heart muscle [Gilbert 05]
  - the human calf muscle [Galban 04]

Data description

Dataset

- 25 subjects, 10 affected by FSH and 15 affected by DM1.
- T1-weighted, T2-weighted and Diffusion Tensor Images of the calf muscle.
- Obtained volumes $64 \times 64 \times 20$ voxels with voxel resolution $3.125\text{mm} \times 3.125\text{mm} \times 7\text{mm}$



Color	Muscle
Yellow	the anterior tibialis
Cyan	the extensor digitorum longus
Magenta	the peroneous longus
White	the posterior tibialis
Blue	the soleus
Green	the lateral gastrocnemius
Red	the medial gastrocnemius

Structural and DTI features

We extract for every muscle the following features from the structural data:

1. the absolute volume,
2. the mean T1 signal,
3. the mean T2 signal, and
4. the Signal to Noise Ration (SNR).

and from the DTI data:

1. the Fractional Anisotropy (FA),
2. the trace of the diffusion tensor,
3. the volume of the tensor,
4. the eigenvalues (L1, L2, L3),
5. the planar coefficient (Cp), and
6. the linear coefficient (Cl).

Structural and DTI features

We extract for every muscle the following features from the structural data:

1. the absolute volume,
2. the mean T1 signal,
3. the mean T2 signal, and
4. the Signal to Noise Ration (SNR).

and from the DTI data:

1. the Fractional Anisotropy (FA),
2. the trace of the diffusion tensor,
3. the volume of the tensor,
4. the eigenvalues (L1, L2, L3),
5. the planar coefficient (Cp), and
6. the linear coefficient (Cl).

Number of Features

4 Structured features $\times$ 8 DTI features $\times$ 7 muscles $=$ 84 features

k support norm - Regularization term

The k-support norm [Andreas Argyriou 12] can be computed as

$$\lambda\Omega(w) = \lambda\|w\|_k^{sp} = \lambda \left(\overbrace{\sum_{i=1}^{k-r-1} (|w|_i^{\downarrow})^2}^{\ell_2 \text{ norm}} + \frac{1}{r+1} \overbrace{\left(\sum_{i=k-r}^{d} |w|_i^{\downarrow} \right)^2}^{\ell_1 \text{ norm}} \right)^{\frac{1}{2}}$$

where

- λ is a scalar, user supplied parameter controling the degree of regularization,
- $|w|_i^{\downarrow}$ is the ith largest element of the vector $|w|$,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that negative correlates with the cardinality of w and
- r is the unique integer in $\{0, \ldots, k-1\}$ automatically selected by the algorithm.

k-support regularized SVM

The Support Vector Machine is defined as the following optimization problem:

$$\begin{aligned} \min_{w\in\mathbb{R}^d, b\in\mathbb{R}, \xi\in\mathbb{R}^n} \quad & \lambda\|w\|_2^2 + \sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & y_i\left(\langle w, x_i\rangle + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i. \end{aligned}$$

where

- λ is a scalar, user supplied parameter controling the degree of regularization,

k-support regularized SVM

The k-support norm regularized SVM ($k$sup-SVM) is defined as the following optimization problem:

$$\min_{w \in \mathbb{R}^d, b \in \mathbb{R}, \xi \in \mathbb{R}^n} \quad \lambda \|w\|_k^{sp} + \sum_{i=1}^{n} \xi_i$$
$$\text{s.t.} \quad y_i \left(\langle w, x_i \rangle + b \right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i.$$

where

- λ is a scalar, user supplied parameter controling the degree of regularization,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that negative correlates with the cardinality of w and
- $\|w\|_k^{sp}$ is the k-support penalty.

k-support regularized SVM

The k-support norm regularized SVM ($k$sup-SVM) is defined as the following optimization problem:

$$\min_{w\in\mathbb{R}^d, b\in\mathbb{R}, \xi\in\mathbb{R}^n} \quad \lambda\|w\|_k^{sp} + \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad y_i\left(\langle w, x_i\rangle + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i.$$

where

- λ is a scalar, user supplied parameter controling the degree of regularization,
- $k \in \{1, \ldots, d\}$ is a scalar, user supplied parameter that negative correlates with the cardinality of w and
- $\|w\|_k^{sp}$ is the k-support penalty.

Advantage

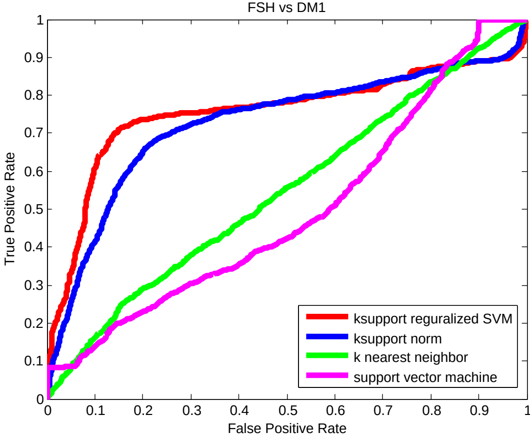
Solution is sparse but correlated subset of discriminative variables.

Experimental Setting

- Methods under comparison:
  - $k$sup-SVM with $k \in \{1, 10, 20, 40, 80\}$ and $\lambda \in \{1, 10, 1000\}$.
  - knn with $k \in \{1, 3, 5, 7, 10\}$
  - SVM with kernel functions
    - linear,
    - polynomial of third degree, and
    - radial basis function (RBF)

    with a soft-margin parameter $C \in \{10^{-3}, 10^{0}, 10^{3}\}$.
  - k-support norm with $k \in \{1, 10, 20, 40, 80\}$ and $\lambda \in \{1, 10, 1000\}$.
- 1000 trials of random split, with 80% of the data used for training and the rest 20% for testing.

Performance

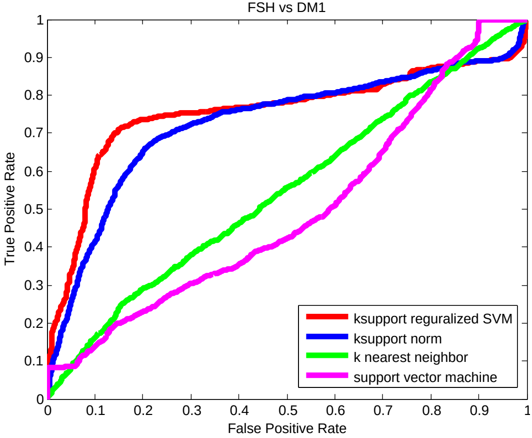


Method	AUC
*k*sup-SVM	0.756
k-support norm	0.726
*k*nn	0.537
SVM	0.494

Method	Accuracy
*k*sup-SVM	77 ± 0.013
k-support norm	74 ± 0.006
*k*nn	61 ± 0.015
SVM	59 ± 0.015

Chance is 60%.

Performance



Method	AUC
$k$sup-SVM	0.756
k-support norm	0.726
$k$nn	0.537
SVM	0.494

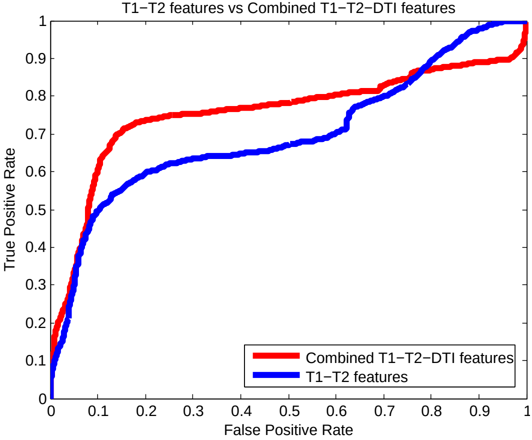
Method	Accuracy
$k$sup-SVM	77 ± 0.013
k-support norm	74 ± 0.006
$k$nn	61 ± 0.015
SVM	59 ± 0.015

Chance is 60%.

$k$sup-SVM vs rest methods

Wilcoxon signed rank test with p-value $<< 10^{-9}$

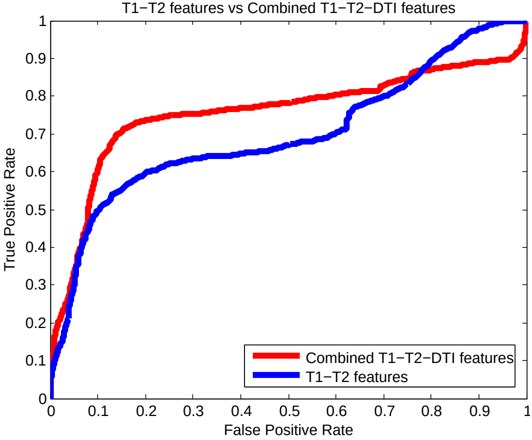
Structured and DTI features vs Structured features only



Features Used	AUC
T1, T2 and DTI	0.756
T1 and T2	0.697

Features Used	Accuracy
T1, T2 and DTI	77 ± 0.013
T1 and T2	73 ± 0.006

Structured and DTI features vs Structured features only



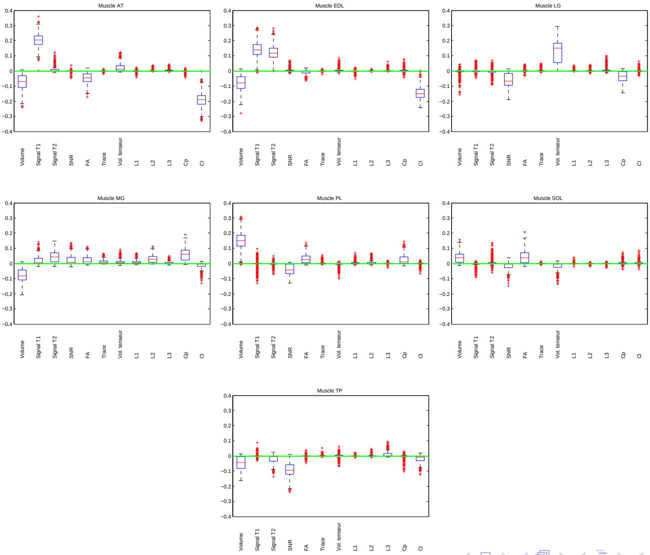
Features Used	AUC
T1, T2 and DTI	0.756
T1 and T2	0.697

Features Used	Accuracy
T1, T2 and DTI	77 ± 0.013
T1 and T2	73 ± 0.006

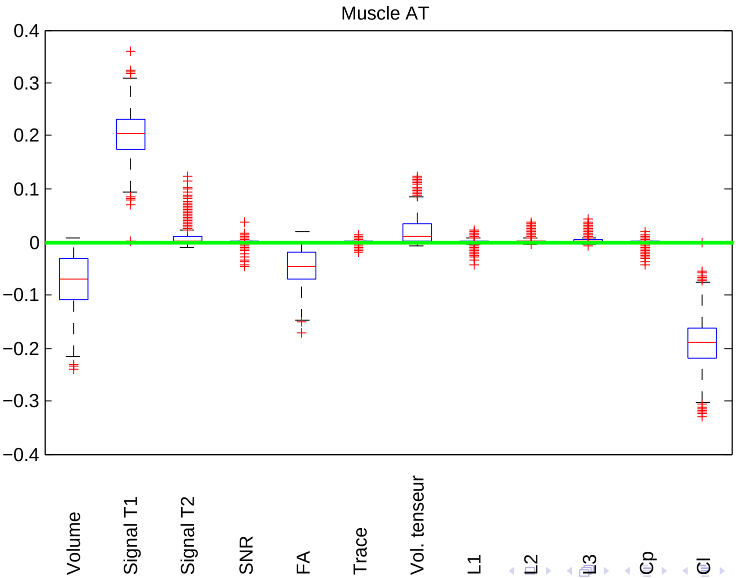
Structured and DTI features vs Structured features only

Wilcoxon signed rank test with p-value $<< 0.05$

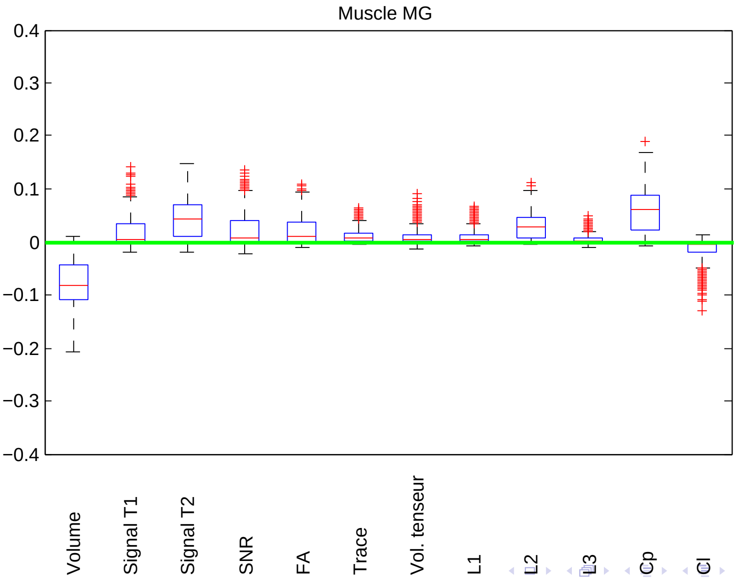
Feature Evaluation



Feature Evaluation on anterior tibialis



Feature Evaluation on medial gastrocnemius



Conclusion

- We studied the more difficult and clinically relevant task of discriminating between two myopathies (FSH vs DM1).
- MRI markers, and DTI tensor features in particular, can discriminate between disease conditions.
- Sparsity regularization appears to be a more important property of the learning algorithm than non-linearity.
- We introduced a novel machine learning algorithm, the *k*sup-SVM.
- The *k*sup-SVM achieved a mean accuracy of 77%.

Source code is available

`https://gitorious.org/ksup-svm`

Questions

Acknowledgements

This work is partially funded by

- the European Research Council under the Seventh Framework Programme (FP7/2007-2013)/ERC Grant 259112.
- the AFM-Telethon foundation.

Rina Foygel Andreas Argyriou & Nathan Srebro.
Sparse Prediction with the k-Support Norm.
In Advances in Neural Information Processing Systems (NIPS), 2012.

A. Argyriou, R. Foygel & N. Srebro.
Sparse Prediction with the k-Support Norm.
In NIPS. 2012.

Francis R. Bach.
Graph kernels between point clouds.
In Proceedings of the 25th international conference on Machine learning, International Conference on Machine Learning '08, pages 25–32, 2008.

A. Bartels & S. Zeki.
The chronoarchitecture of the human brain–natural viewing conditions reveal a time-based anatomy of the brain.
NeuroImage, vol. 22, no. 1, pages 419 – 433, 2004.

A. Bartels & S. Zeki.
Brain dynamics during natural viewing conditions–A new guide for mapping connectivity in vivo.
NeuroImage, vol. 24, no. 2, pages 339–349, 2005.

M.B. Blaschko, J.A. Shelton & A. Bartels.
Augmenting Feature-driven fMRI Analyses: Semi-supervised learning and resting state activity.
In NIPS. 2009.

M.B. Blaschko, J.A. Shelton, A. Bartels, C.H. Lampert & A. Gretton.
Semi-supervised kernel canonical correlation analysis with application to human fMRI.
Pattern Recognition Letters, vol. 32, no. 11, pages 1572 – 1583, 2011.

Karsten M. Borgwardt & Hans-Peter Kriegel.
Shortest-Path Kernels on Graphs.
In Proceedings of the Fifth IEEE International Conference on Data Mining, ICDM '05, pages 74–81, Washington, DC, USA, 2005. IEEE Computer Society.

M.K. Carroll, G.A. Cecchi, I. Rish, R. Garg & A.R. Rao.
Prediction and interpretation of distributed neural activity with sparse models.
NeuroImage, vol. 44, no. 1, pages 112 – 122, 2009.

Fabrizio Costa & Kurt De Grave.
Fast neighborhood subgraph pairwise distance kernel.
In Proceedings of the 26th International Conference on Machine Learning, pages 255–262, 2010.

Craig J Galban, Stefan Maderwald, Kai Uffmann, Armin de Greiff & Mark E Ladd.
Diffusive sensitivity to muscle architecture: a magnetic resonance diffusion tensor imaging study of the human calf.
European journal of applied physiology, vol. 93, no. 3, pages 253–262, 2004.

Thomas Gärtner, Peter Flach & Stefan Wrobel.
On Graph Kernels: Hardness Results and Efficient Alternatives.
In Bernhard Schölkopf & Manfred K. Warmuth, editeurs, Learning Theory and Kernel Machines, volume 2777 of *Lecture Notes in Computer Science*, pages 129–143. Springer Berlin Heidelberg, 2003.

Richard J Gilbert & Vitaly J Napadow.
Three-dimensional muscular architecture of the human tongue determined in vivo with diffusion tensor magnetic resonance imaging.
Dysphagia, vol. 20, no. 1, pages 1–7, 2005.

Kristen Grauman & Trevor Darrell.
The Pyramid Match Kernel: Efficient Learning with Sets of Features.
Journal of Machine Learning Research, vol. 8, pages 725–760, May 2007.

D.R. Hardoon, J. Mourão-Miranda, M. Brammer & J. Shawe-Taylor.
Unsupervised analysis of fMRI data using kernel canonical correlation.
NeuroImage, vol. 37, no. 4, pages 1250 – 1259, 2007.

Tamás Horváth, Thomas Gärtner & Stefan Wrobel.
Cyclic pattern kernels for predictive graph mining.
In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '04, pages 158–167, 2004.

Madhura Ingalhalikar, Drew Parker, Luke Bloy, Timothy PL Roberts & Ragini Verma.
Diffusion based abnormality markers of pathology: Toward learned diagnostic prediction of ASD.
Neuroimage, vol. 57, no. 3, pages 918–927, 2011.

Stefan Klöppel, Cynthia M Stonnington, Carlton Chu, Bogdan Draganski, Rachael I Scahill, Jonathan D Rohrer, Nick C Fox, Clifford R Jack, John Ashburner & Richard SJ Frackowiak.
Automatic classification of MR scans in Alzheimer's disease.
Brain, vol. 131, no. 3, pages 681–689, 2008.

Stephen LaConte, Stephen Strother, Vladimir Cherkasskyet *et al.*
Support vector machines for temporal classification of block design fMRI data.
NeuroImage, vol. 26, no. 2, page 317, 2005.

Zhiqiang Lao, Dinggang Shen, Zhong Xue, Bilge Karacali, Susan M Resnick & Christos Davatzikos.
Morphological classification of brains via high-dimensional shape transformations and machine learning methods.
Neuroimage, vol. 21, no. 1, pages 46–57, 2004.

P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret & J.-P. Vert.
Extensions of marginalized graph kernels.
In Proceedings of the Twenty-First International Conference on Machine Learning (ICML 2004), pages 552–559, 2004.

Pierre Mahé & Jean-Philippe Vert.
Graph kernels based on tree patterns for molecules.
Machine Learning, vol. 75, no. 1, pages 3–35, 2009.

B. Ng, V. Siless, G. Varoquaux, J.-B. Poline, B. Thirion & R. Abugharbieh.
Connectivity-informed Sparse Classifiers for fMRI Brain Decoding.
In Pattern Recognition in Neuroimaging, 2012.

Liva Ralaivola, Sanjay Joshua Swamidass, Hiroto Saigo & Pierre Baldi.
Graph kernels for chemical informatics.
Neural Networks, vol. 18, no. 8, pages 1093–1110, 2005.

Jan Ramon & Thomas Gaertner.
Expressivity versus efficiency of graph kernels.
In Proceedings of the First International Workshop on Mining Graphs, Trees and Sequences, pages 65–74, 2003.

N. Shervashidze, S. V. N. Vishwanathan, T. Petri, K. Mehlhorn & K. Borgwardt.
Efficient graphlet kernels for large graph comparison.
In Proceedings of the International Workshop on Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics, 2009.

Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn & Karsten M. Borgwardt.
Weisfeiler-Lehman Graph Kernels.
Journal of Machine Learning Research, vol. 12, pages 2539–2561, November 2011.

S. Song, Z. Zhan, Z. Long, J. Zhang & L. Yao.
Comparative Study of SVM Methods Combined with Voxel Selection for Object Category Classification on fMRI Data.
PLoS One, vol. 6, no. 2, page e17191, 2011.

R. Tibshirani.
Regression Shrinkage and Selection via the Lasso.
Journal of the Royal Statistical Society Series B, vol. 58, pages 267–288, 1996.

S. Vichy N. Vishwanathan, Nicol N. Schraudolph, Risi Imre Kondor & Karsten M. Borgwardt.
Graph Kernels.
Journal of Machine Learning Research, vol. 11, pages 1201–1242, 2010.

Boris Weisfeiler & A.A. Lehman.
A reduction of a graph to a canonical form and an algebra arising during this reduction.
Nauchno-Technicheskaya Informatsia, vol. 2, no. 9, pages 12–16, 1968.

H. Zou & T. Hastie.
Regularization and variable selection via the elastic net.
Journal of the Royal Statistical Society Series B, vol. 67, no. 2, pages 301–320, 2005.

Hui Zou & Trevor Hastie.
Regularization and variable selection via the Elastic Net.
Journal of the Royal Statistical Society, Series B, vol. 67, pages 301–320, 2005.