Supervised Segmentation of Fiber Tracts

Emanuele Olivetti, Paolo Avesani

NeuroInformatics Laboratory (NILab)
Fondazione Bruno Kessler, Trento (FBK), Italy
Centro Interdipartimental Mente e Cervello (CIMeC), Università di Trento, Italy
http://nilab.fbk.eu , olivetti@fbk.eu

SIMBAD2011: 1st International Workshop on
Similarity-Based Pattern Analysis and Recognition
Venice (Italy), Sept. 28-30th 2011

Motivation: Brain Connectivity

- Trend in Neuroscience: unraveling the brain network.
- Neurological studies interested in anatomy of white matter (eg. Alzheimer Disease).
- Brain Connectivity & Machine Learning:
  - NIPS workshop: CINI2009.
  - International Conference on Data Mining, Contest 2009. (Pittsburgh Brain Connectivity Competition, 2009)

DISCLAIMER

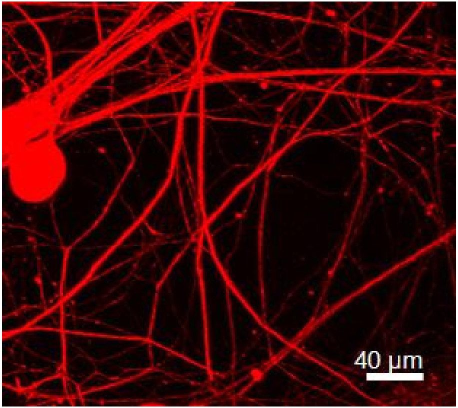
This talk is NOT about effective connectivity (i.e. causality) or functional connectivity (i.e. statistical dependence). We study *structural connectivity* (i.e. *anatomical links*).

Outline

Outline

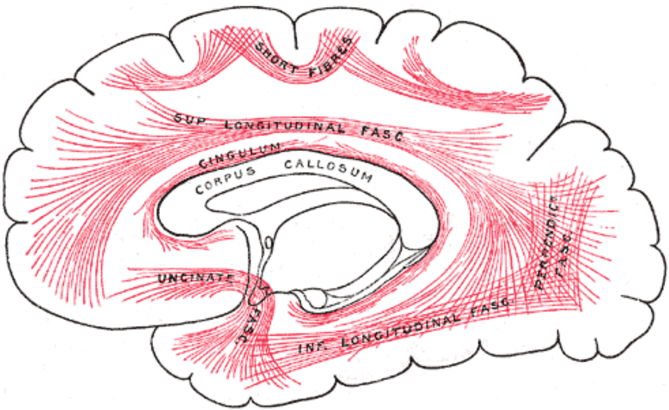
Fiber Tract Segmentation: Axons

Brain contains hundreds of millions of neuronal axons that constitute the white matter and act as wiring.



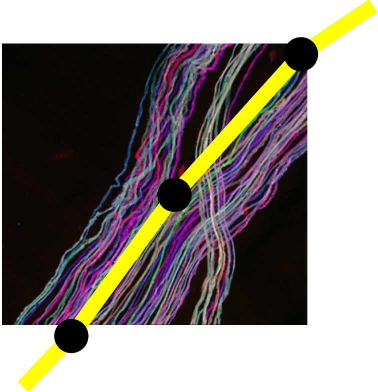
Fiber Tract Segmentation: Bundles

Axons are grouped in neuronal *pathways*/bundles/*tracts* sharing a common path

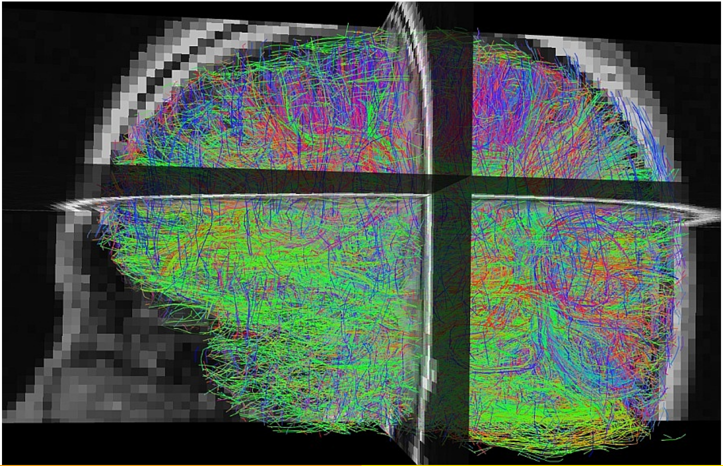


Fiber Tract Segmentation: Streamlines

- dMRI techniques allow the reconstruction of pathways in living subjects. Resolution $\approx$ 1mm.
- (deterministic) Tractography algorithms reconstruct streamlines/*fibers*.
- A streamline is a polyline representing thousands of axons.

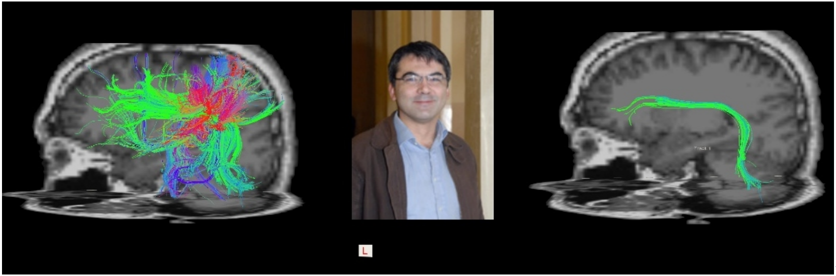


Fiber Tract Segmentation: Tractography ($\approx 10^5$ streaml). Here: 5%.



Fiber Tract Segmentation: Human Expert

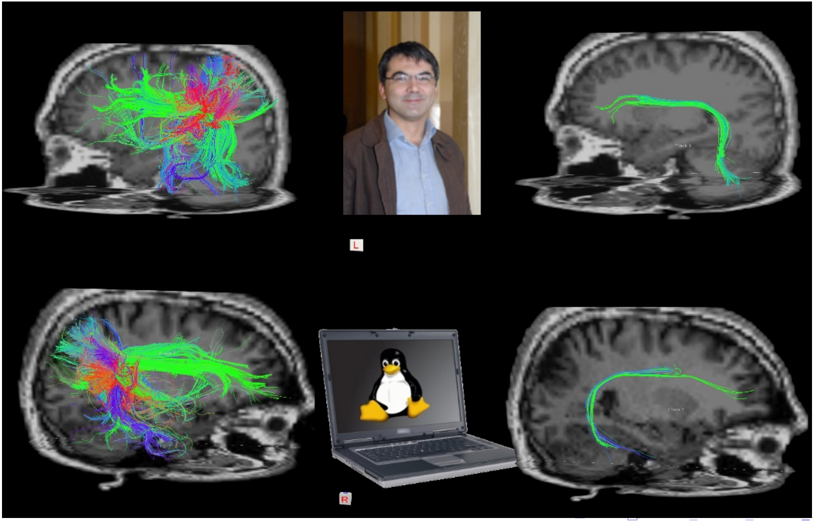
- Neuroanatomists are able to identify neural pathways/fiber bundles.
- Because of the large amount of streamlines and the anatomical variability among subjects, manual segmentation is difficult and lengthy.



Algorithmic Fiber Tract Segmentation

- High-level problem: *automatically segment a fiber tract on a given tractography given examples of that segmentation on different brains*.

Supervised Segmentation of Fiber Tracts



Previous Work

Unsupervised Tract Segmentation

- Unsupervised clustering of streamlines: agglomerative, k-means, . . . (updated short review: [Wang et al., 2011]).
- Several *ad-hoc* distance functions proposed in the literature, see [Zhang et al., 2008].

(partly) *Supervised* Tract Segmentation

- Affine reg. + *B*-spline based 1-NN to atlas [Maddah et al., 2005].
- Spectral Clustering [O'Donnell and Westin, 2007].
- Hierarchical Dirichelet Process [Wang et al., 2011].

Outline

Proposed Solution

1. Map all tractographies to a common space.
   - *Ask later if interested.*
2. Adopt the Statistical Learning Framework.
   - Definition and Notation
   - Use prior knowledge: streamline distances.
   - Classification strategies:
     - k-NN.
     - Indefinite streamline kernel.
     - (Dis)similarity space (+ Linear SVM).

Definitions and Notation

- Streamline: a polyline $s = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_s}\}$, where $\boldsymbol{x} \in \mathbb{R}^3$.
- Tractography: $T = \{s_1, \ldots, s_M\} \sim \mathfrak{T}$. Usually $|T| \simeq 3 \times 10^5$.
- Fiber Bundle / tract: $t \subset T$
- The neuroanatomist provides a segmentation: $Y = \{y_1, \ldots, y_M\}$, $y_i \in \{0, 1\}$.

Supervised Learning Problem

- Given a class-labeled sample $\{(s_1, y_1), \ldots, (s_N, y_N)\} \sim \mathfrak{P}$.
- Learn f^* from examples minimizing a given Loss L:

$$f^* = \operatorname*{argmin}_{f \in \mathcal{F}} E_{\mathfrak{P}}[L(f(s), y)]$$

Issues

Streamlines and Euclidean spaces

"*Most of the classification algorithms in the literature assume that objects live in a Euclidean feature space*", but:

1. Streamlines have different lengths across the brain.
2. The number of points of a streamline is not the same across the brain.

So streamlines cannot be *directly* embedded into a Euclidean space.

Distances between streamlines (from the literature)

non-metric modified Hausdorff distances [Zhang et al., 2008, Dubuisson and Jain, 1994]. Usually:

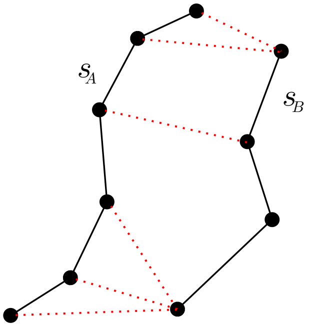
$$\boldsymbol{d}_1(s_A, s_B) = \frac{1}{K_A} \sum_{i=1}^{K_A} d(\boldsymbol{x}_i^A, s_B)$$

$$\boldsymbol{d}_2(s_A, s_B) = \min_{i=1,\dots,K_A} d(\boldsymbol{x}_i^A, s_B)$$

$$\boldsymbol{d}_3(s_A, s_B) = max_{i=1,\dots,K_A} d(\boldsymbol{x}_i^A, s_B)$$

where

$$d(\boldsymbol{x}_i^A, s_B) = \min_{j=1,\dots,K_B} ||\boldsymbol{x}_i^A - \boldsymbol{x}_j^B||_2$$



k-Nearest Neighbor

"*predict class-label as the most frequent one among the k nearest examples.*" (break ties at random)

Pros

- Simple and Effective.
- Universally Bayes-consistent [Stone, 1977] (for *metric* distances).

Cons

- *k* needs to be defined from data or prior knowledge.
- Sensitive to noise.

Kernel Methods: SVC, GPC, perceptron, etc.

"*Map data to a RKHS space to enhance linear separability. Make it easy with the kernel trick.*" $k(a, b) = \langle\phi(a), \phi(b)\rangle$

Pros

- Very Effective and Widely Adopted.
- Kernels are similarity functions, e.g. $k(a, b) = e^{-d(a,b)}$.
- Generalization bounds [Schölkopf and Smola, 2002].
- Convex optimization problem.

Cons

- Kernel k must be *positive semi-definite* (PSD).
- if d is not *metric*, then $k(a, b) = e^{-d(a,b)}$ is not PSD.

. . . and Indefinite Kernels

Issues

- Non convexity: local minima, saddle point.
- Increased amount of computation.
- No generalization error guarantees.

Available Solutions [Chen et al., 2009]

- "*Issues are not so big, do not worry.*"
- *Massage* the kernel matrix $K_{ij} = k(s_i, s_j)$. Since $K = U^T \Lambda U$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots)$:
  - Clip: $\lambda_i = \max(0, \lambda_i)$
  - Flip: $\lambda_i = |\lambda_i|$
  - Shift: $\lambda_i = \lambda_i + |\min(\lambda_{min}, 0)|$

(Dis)similarity representation [Pekalska et al., 2002]

"*Given a set of prototypes/landmarks* $R = \{\tilde{s}_1, \ldots, \tilde{s}_D\}$ *map streamlines to* $\mathbb{R}^D$ *via* $\psi_R(s) = [d(s, \tilde{s}_1), \ldots, d(s, \tilde{s}_D)]^T$*.*"

Pros

- Every classification algorithm can be used.
- d has no constraints.

Cons

- Computationally more expensive: construction of feat.space.
- How to select prototypes $\{\tilde{s}_i\}_i$?
- Generalization?

(Dis)similarity representation CONT.

General results [Balcan et al., 2008a, Balcan et al., 2008b]

If d is *good* (= expected intraclass similarity):

- We achieve good generalization.
- ...even when $\{\tilde{s}_i\}_i$ is a random subset of data.
- ...and we have an upper bound on D.

Issues

- do assumptions hold?

Proposed Method

Proposed Method

- Use **dis**similarity representation.
- Use d_1:

$$\boldsymbol{d}_1(s_A, s_B) = \frac{1}{K_A} \sum_{i=1}^{K_A} d(\boldsymbol{x}_i^A, s_B)$$

where

$$d(\boldsymbol{x}_i^A, s_B) = \min_{j=1,\dots,K_B} ||\boldsymbol{x}_i^A - \boldsymbol{x}_j^B||_2$$

- Use random prototypes from $\{\text{Train} \cup \text{Test}\}$.
- Train a linear SVM.

Outline

Dataset: *enhanced* PBCC 2009

Pittsburgh Brain Connectivity Competition (PBCC) 2009, Spring
www.braincompetition.org 3 subjects $\times$ 8 fiber tracts.

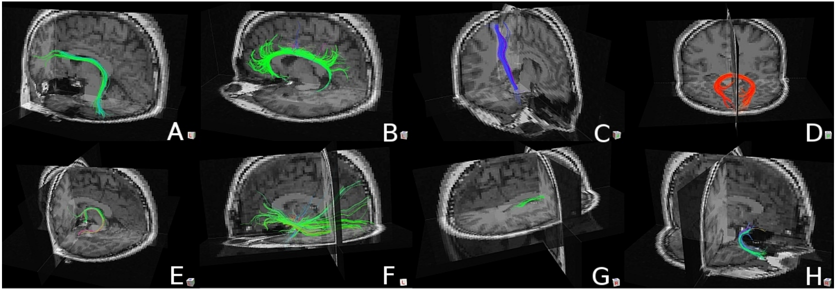


Figure: Arcuate (A), cingulum (B), corticospinal (C), forceps major (D), fornix (E), i.o.f.f. (F), subcallosal (G), uncinate (H).

Evaluation Criterion

	true 0	true 1
pred 0	TN	FN
pred 1	FP	TP

PBCC2009 score:

$$r = \frac{TP - FP}{TP + FN}$$

Note

- $r \in [-\frac{|S_\tau|}{|t|} + 1, 1]$ ($r < 0$: bad, $r > 0$:good)
- r focuses on sensitivity and penalize predicting a large tract.
- $r = \frac{TP}{TP+FN} - \frac{FP}{TP+FN} = \text{sensitivity} - \frac{FP}{TP+FN}$

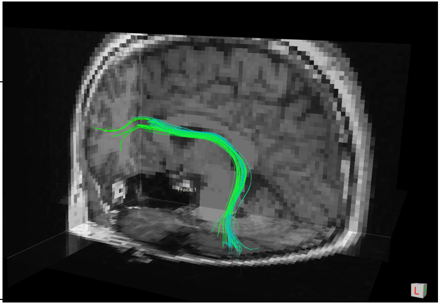
Results: Same Subject, Proposed Solution

tract	Subj0	Subj1	Subj2
arcuate	0.94	0.96	0.93
cingulum	0.85	0.89	0.92
corticosp.	0.94	0.95	0.92
forceps	0.98	0.94	0.92
fornix	0.81	0.86	0.72
ioff	0.70	0.72	0.90
subcall.	0.92	0.83	0.87
uncinate	0.84	0.75	0.63

PBCC2009 score averaged over 4 draws of 100 random prototypes, SVM linear kernel, 10-fold CV. **Std-mean** $\approx$ **0.02**, Testset $=$ 5000.

Results Cross-Subject. Arcuate Fasciculus

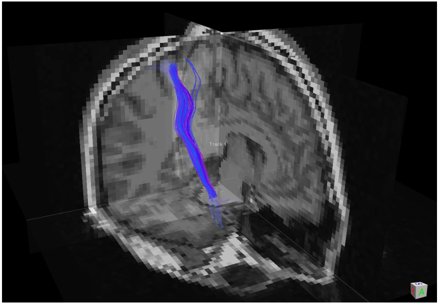
train $\mapsto$ test	1-NN	d_1^{100}+ℓSVM
$1_L \mapsto 2_R$	0.224	**0.328**
$1_L \mapsto 3_R$	0.338	**0.711**
$2_R \mapsto 1_L$	-0.021	**0.333**
$2_R \mapsto 3_R$	0.697	**0.860**
$3_R \mapsto 1_L$	0.260	**0.792**
$3_R \mapsto 2_R$	**0.229**	0.187



Cross-Subject Segmentation of the *arcuate fasciculus*. Sizes - Subj1 : 96/4027, Subj2 : 406/5050, Subj3 : 228/5142.

Results Cross-Subject. Corticospinal tract

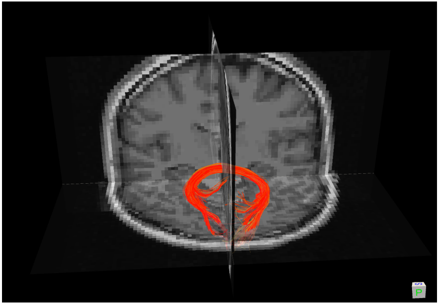
train $\mapsto$ test	1-NN	d_1^{100}+ℓSVM
$1_R \mapsto 2_L$	0.402	**0.767**
$1_R \mapsto 3_L$	0.091	**0.387**
$2_L \mapsto 1_R$	0.446	**0.749**
$2_L \mapsto 3_L$	**0.852**	0.588
$3_L \mapsto 1_R$	0.417	**0.869**
$3_L \mapsto 2_L$	0.459	**0.698**



Cross-Subject Segmentation of the *corticospinal tract*. Sizes - Subj1 : $175/6615$, Subj2 : $331/4877$, Subj3 : $243/5211$.

Results Cross-Subject. Forceps Major

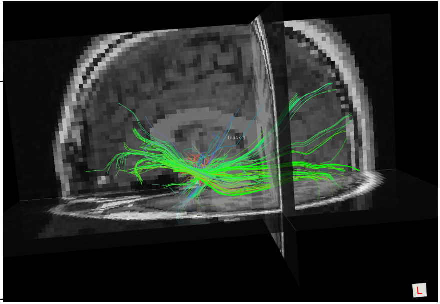
train $\mapsto$ test	1-NN	d_1^{100}+ℓSVM
$1 \mapsto 2$	**0.732**	0.506
$1 \mapsto 3$	**0.323**	0.194
$2 \mapsto 1$	0.158	**0.544**
$2 \mapsto 3$	0.658	**0.726**
$3 \mapsto 1$	0.014	**0.347**
$3 \mapsto 2$	0.366	**0.743**



Cross-Subject Segmentation of the *forceps major*. Sizes - Subj1 : $366/8333$, Subj2 : $385/4586$, Subj3 : $263/4504$.

Results: Inferior Occipito-Frontal Fasciculus (IOFF)

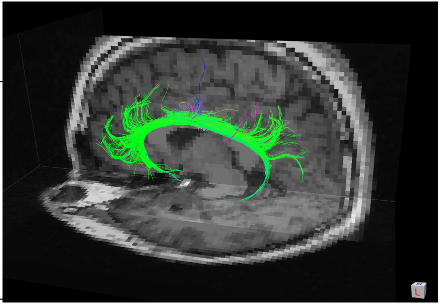
train $\mapsto$ test	1-NN	d_1^{100}+ℓSVM
$1_L \mapsto 2_L$	-0.853	**0.323**
$1_L \mapsto 3_L$	-1.170	**0.567**
$2_L \mapsto 1_L$	-0.095	**0.189**
$2_L \mapsto 3_L$	-0.025	**0.415**
$3_L \mapsto 1_L$	0.090	**0.229**
$3_L \mapsto 2_L$	-0.049	**0.203**



Cross-Subject Segmentation of the *inferior occipito-frontal fasciculus* (ioff). Sizes - Subj1 : 433/3152, Subj2 : 266/3234, Subj3 : 282/4858.

Results Cross-Subject. Cingulum

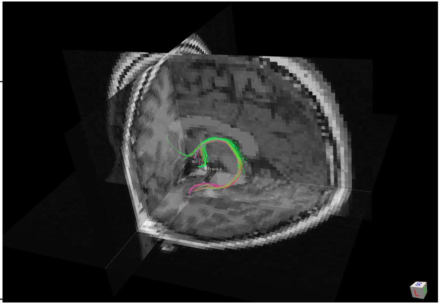
train $\mapsto$ test	1-NN	d_1^{100}+ℓSVM
$0_L \mapsto 1_R$	-1.243	-0.778
$0_L \mapsto 2_L$	-0.201	**0.211**
$1_R \mapsto 0_L$	-0.343	0.000
$1_R \mapsto 2_L$	**0.608**	**0.624**
$2_L \mapsto 0_L$	0.360	**0.558**
$2_L \mapsto 1_R$	0.351	**0.465**



Sizes - Subj1 : $539/4211$, Subj2 : $185/4117$, Subj3 : $194/5375$.

Results Cross-Subject. Fornix

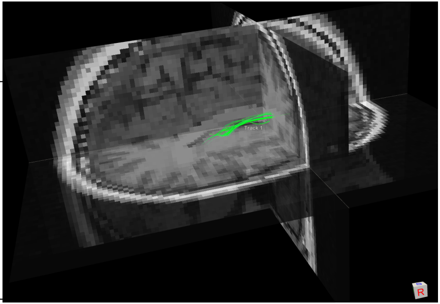
train $\mapsto$ test	1-NN	$\boldsymbol{d}_1^{100}+\ell$SVM
$0_L \mapsto 1_L$	0.156	**0.321**
$0_L \mapsto 2_L$	-0.319	**0.553**
$1_L \mapsto 0_L$	-0.407	**0.111**
$1_L \mapsto 2_L$	-0.404	**0.362**
$2_L \mapsto 0_L$	0.296	**0.370**
$2_L \mapsto 1_L$	-0.431	-0.505



Sizes - Subj1 : $54/3999$, Subj2 : $109/4908$, Subj3 : $47/4659$.

Results Cross-Subject. Subcallosal

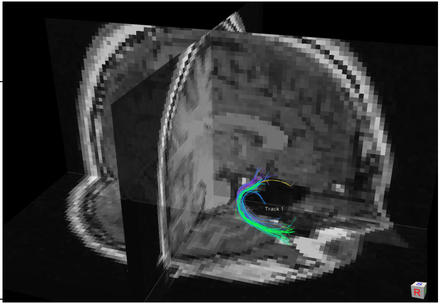
train $\mapsto$ test	1-NN	$\boldsymbol{d}_1^{100}+\ell$SVM
$0_R \mapsto 1_R$	-1.333	0.000
$0_R \mapsto 2_L$	-0.794	0.000
$1_R \mapsto 0_R$	-0.667	0.000
$1_R \mapsto 2_L$	**0.441**	0.000
$2_L \mapsto 0_R$	-1.259	0.000
$2_L \mapsto 1_R$	-1.278	0.000



Sizes - Subj1 : $27/3419$, Subj2 : $18/3464$, Subj3 : $34/4434$.

Results Cross-Subject. Uncinate

train $\mapsto$ test	1-NN	d_1^{100}+ℓSVM
$0_R \mapsto 1_R$	**0.263**	0.075
$0_R \mapsto 2_R$	**0.328**	0.197
$1_R \mapsto 0_R$	**0.427**	0.280
$1_R \mapsto 2_R$	-0.090	0.016
$2_R \mapsto 0_R$	-0.134	**0.402**
$2_R \mapsto 1_R$	-0.500	-0.375



Sizes - Subj1 : $82/4148$, Subj2 : $80/4829$, Subj3 : $122/3711$.

Outline

Conclusions

Take-home message

- Same-Subject
  - Experiments confirm that dissimilarity representation works well.
- Cross-Subject
  - (Dis)similarity + Lin.SVM works *consistently* better than 1-NN.
  - There is large anatomical variability across-subject and proposed registration seems the weak link.

Future work

- Test different number of random prototypes.
- Optimal prototypes? [Snelson and Ghahramani, 2005]
- Experiment with *indefinite* kernels.
- Tackle cross-subject anatomical differences by *Domain Adaptation*/*Transfer Learning*.
    - (Semi-)Supervised (e.g. [Daumé et al., 2010]).
    - Unsupervised (e.g. Transductive SVM. DOES NOT WORK!)

Thank You!

Balcan, M.-F., Blum, A., and Srebro, N. (2008a).
A theory of learning with similarity functions.
Machine Learning, 72(1):89–112.

Balcan, M.-F., Blum, A., and Srebro, N. (2008b).
Improved Guarantees for Learning via Similarity Functions.
In Servedio, R. A. and Zhang, T., editors, *COLT*, pages 287–298. Omnipress.

Chen, Y., Garcia, E. K., Gupta, M. R., Rahimi, A., and Cazzanti, L. (2009).
Similarity-based Classification: Concepts and Algorithms.
J. Mach. Learn. Res., 10:747–776.

Daumé, H., Kumar, A., and Saha, A. (2010).
A Co-regularization Based Semi-supervised Domain Adaptation.

In *Proceedings of the Conference on Neural Information Processing Systems (NIPS)*, Vancouver, Canada.

Dubuisson, M. P. and Jain, A. K. (1994).
A modified Hausdorff distance for object matching.
In *Proceedings of 12th International Conference on Pattern Recognition*, volume 1, pages 566–568. IEEE Comput. Soc. Press.

Maddah, M., Mewes, A. U. J., Haker, S., Grimson, and Warfield, S. K. (2005).
Automated Atlas-Based Clustering of White Matter Fiber Tracts from DTMRI.
In Duncan, J. S. and Gerig, G., editors, *Medical Image Computing and Computer-Assisted Intervention âĂŞ MICCAI 2005*, volume 3749 of *Lecture Notes in Computer*

Science, chapter 24, pages 188–195. Springer Berlin / Heidelberg, Berlin, Heidelberg.

O'Donnell, L. J. and Westin, C.-F. F. (2007).
Automatic tractography segmentation using a high-dimensional white matter atlas.
IEEE transactions on medical imaging, 26(11):1562–1575.

Pekalska, E., Paclik, P., and Duin, R. P. W. (2002).
A generalized kernel approach to dissimilarity-based classification.
J. Mach. Learn. Res., 2:175–211.

Schölkopf, B. and Smola, A. J. (2002).
Learning with kernels : support vector machines, regularization, optimization, and beyond.
MIT Press, 1st edition.

Snelson, E. and Ghahramani, Z. (2005).
Sparse Gaussian Processes using Pseudo-inputs.

Stone, C. J. (1977).
Consistent Nonparametric Regression.
The Annals of Statistics, 5(4).

Wang, X., Grimson, and Westin, C.-F. (2011).
Tractography segmentation using a hierarchical Dirichlet processes mixture model.
NeuroImage, 54(1):290–302.

Zhang, S., Correia, S., and Laidlaw, D. H. (2008).
Identifying White-Matter Fiber Bundles in DTI Data Using an Automated Proximity-Based Fiber-Clustering Method.
Visualization and Computer Graphics, IEEE Transactions on, 14(5):1044–1053.