Learning to Compare Examples

NIPS'06 Workshop

– Organizers –
David Grangier and Samy Bengio

IDIAP Research Institute, Switzerland
{grangier, bengio}@idiap.ch

Outline

Introduction

- Distances & kernels in Machine Learning
- Learning a distance/kernel from data
- Open issues in distance/kernel learning

Workshop Overview

Acknowledgments

Distances & Kernels in Machine Learning

Definition

- functions on example pairs, measure the proximity of examples.
- distance metric:

$$d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R},$$

non-negativity, identity, symmetry, triangle inequality

- kernel:

$$k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R},$$

symmetry, positive definiteness

Distances & Kernels in Machine Learning

Crucial for several approaches

- density estimator (e.g. Parzen windows, SV density estimation),
- clustering (e.g. k-means, spectral clustering),
- distance-based classifiers (e.g. RBF networks, k-NN classifiers),
- kernel-based classifiers (e.g. SVM)...

Selecting a Suitable Distance/Kernel

Standard procedure:

- a-priori selection
  (e.g. Euclidean distance, linear kernel)
- cross-validation within a small family of functions.
  (e.g. selecting the degree of the polynomial kernel)

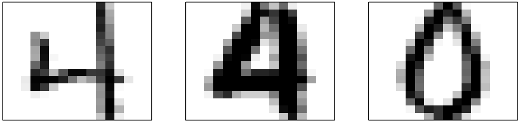
Selecting a Suitable Distance/Kernel

Standard procedure:

- a-priori selection
  (e.g. Euclidean distance, linear kernel)
- cross-validation within a small family of functions.
  (e.g. selecting the degree of the polynomial kernel)

Not always effective:

e.g. Euclidean distance or RBF kernel on USPS,



A B C

$$d(A, B) > d(B, C)$$

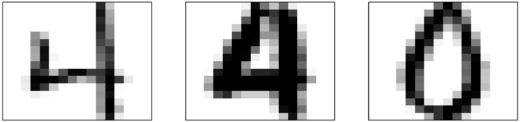
Selecting a Suitable Distance/Kernel

Standard procedure:

- a-priori selection
  (e.g. Euclidean distance, linear kernel)
- cross-validation within a small family of functions.
  (e.g. selecting the degree of the polynomial kernel)

Not always effective:

e.g. Euclidean distance or RBF kernel on USPS,



A B C

$$d(A, B) > d(B, C)$$

Better alternative ? Learn the distance/kernel function from data !

Learning the Distance/Kernel from Data

Different sources of information can be exploited:

Learning the Distance/Kernel from Data

Different sources of information can be exploited:

- labeled data

  e.g. find the distance metric which optimizes the performance of a k-NN classifier [Weinberger, Blitzer and Saul, NIPS'05]

Learning the Distance/Kernel from Data

Different sources of information can be exploited:

- labeled data
- invariance properties

  e.g. in face verification, 'picture of A with flash' should be close to 'picture of A without flash' [Chopra, Hadsell and LeCun, CVPR'05]

Learning the Distance/Kernel from Data

Different sources of information can be exploited:

- labeled data
- invariance properties
- proximity information

e.g. in text retrieval, this query Q is closer to the relevant document A than to an unrelated document B [Joachims, KDD'02]

Learning the Distance/Kernel from Data

Different sources of information can be exploited:

- labeled data
- invariance properties
- proximity information
- data labeled for another task

  e.g. in computer vision, learning that object A is far from object B can help to discriminate between objects C and D. [Fleuret and Blanchard, NIPS'05]

Learning the Distance/Kernel from Data

Different sources of information can be exploited:

- labeled data
- invariance properties
- proximity information
- data labeled for another task
- unlabeled data

  e.g. according to the cluster assumption, distance should be greater when crossing low density region. [Chapelle and Zien, AIStat'05]


Learning the Distance/Kernel from Data

Different sources of information can be exploited:

- labeled data
- invariance properties
- proximity information
- data labeled for another task
- unlabeled data
- etc.

Learning the Distance/Kernel from Data

Different formalizations of the problem,

Learning the Distance/Kernel from Data

Different formalizations of the problem,

- margin maximization for SVM or k-NN

Training set	labeled examples.
Learning Objective	minimize a lower bound of generalization error of the final classifier.
e.g.	[Lanckriet et al., JMLR'04]
	[Weinberger, Blitzer and Saul, NIPS'05]

Learning the Distance/Kernel from Data

Different formalizations of the problem,

- margin maximization for SVM or k-NN
- classification of pairs

Training set	similar and dissimilar pairs.
Learning objective	for any new pair (x, x'),
	$d(x, x') = \begin{cases} 0 & \text{if the pair is similar} \\ +\infty & \text{if the pair is dissimilar.} \end{cases}$
e.g.	[Xing et al., NIPS'03]
	[Chopra, Hadsell and LeCun, CVPR'05]

Learning the Distance/Kernel from Data

Different formalizations of the problem,

- margin maximization for SVM or k-NN
- classification of pairs
- proximity constraints

Training set	proximity constraints 'a is closer to b than c'.
Learning objective	for any new constraint (a, b, c), $d(a,b) < d(b,c)$
e.g.	[Joachims, KDD'02] [Burges et al., ICML'05]

Learning the Distance/Kernel from Data

Different formalizations of the problem,

- margin maximization for SVM or k-NN
- classification of pairs
- proximity constraints
- density-based approaches

Training set	unlabeled examples
Learning objective	shorter distance across densely populated areas
e.g.	[Lebanon, UAI'03]
	[Chapelle and Zien, AIStat'05]



Learning the Distance/Kernel from Data

Different formalizations of the problem,

- margin maximization for SVM or k-NN
- classification of pairs
- proximity constraints
- density-based approaches
- etc.

Open issues in Learning to Compare Examples

- Parameterization
  Mainly, Mahalanobis distance, $d(x, y)^2 = (x - y)^T \ A^T A \ (x - y)$, or linear combination of kernels, $k(x, y) = \sum_i \lambda_i k_i(x, y)$.
- Regularization:
  What should be the regularizer for a kernel, a metric ?
- Efficiency:
  Most approaches working on example pairs are expensive to train.
- Multi-objective learning:
  How to jointly learn
  - a kernel relying on proximity data or data labeled for another task,
  - a kernel-based classifier relying on labeled data ?
- etc.

Workshop Overview

An application-oriented morning session:

- invited talk by Yann LeCun,

  Learning Similarity Metrics with Invariance Properties,
- 3 contributed talks, mainly on computer vision,
- followed by a discussion on

  Applications of Learning to Compare Examples.

Workshop Overview

A more theoretic afternoon session:

- invited talk by Sam Roweis,

  Neighborhood Components Analysis & Metric Learning,
- 4 contributed talks, mainly on distance metric learning,
- followed by a discussion on

  Kernel & Distance Learning.

Workshop Overview

Note on Contributed Talks:

The program committee

- Samy Bengio, IDIAP Research Institute
- Gilles Blanchard, Fraunhofer FIRST
- Chris Burges, Microsoft Research
- Francois Fleuret, EPFL
- David Grangier, IDIAP Research Institute
- Thomas Hofmann, Google Switzerland
- Guy Lebanon, Purdue University
- Thorsten Joachims, Cornell University
- Yoram Singer, The Hebrew University
- Alex Smola, National ICT Australia

reviewed 14 papers out of which 7 were accepted.

Acknowledgments

This workshop would not have been possible without,

- the program committee,
- the invited speakers,
- the contributing authors,

and, of course,

- the attendees !

Thanks also to the PASCAL European Network for its financial support.

Learning to Compare Examples – Morning Session

7:30am Introduction by D. Grangier
Learning to Compare Examples

8:00am Invited talk by Y. LeCun
Learning Similarity Metrics with Invariance Properties

8:45am E. Nowak and F. Jurie
Learning Visual Distance Function for Object Identification from one Example

9:10am *Coffee break*

9:30am A. Maurer
Learning to Compare using Operator-Valued Large-Margin Classifiers

9:55am M. B. Blaschko and T. Hofmann
Conformal Multi-Instance Kernels

10:20am Discussion
Suggested Topic: Applications of Learning to Compare Examples

Learning to Compare Examples – Afternoon Session

3:30pm	Invited talk by S. Roweis Neighborhood Components Analysis & Metric Learning
4:15pm	J. Peltonen, J. Goldberger and S. Kaski Fast Discriminative Component Analysis for Comparing Examples
4:40pm	J. Davis, B. Kulis, S. Sra and I. Dhillon Information-Theoretic Metric Learning
5:05pm	*Coffee break*
5:25pm	J. Dillon, Y. Mao, G. Lebanon and J. Zhang Statistical Translation, Heat Kernels, and Expected Distances
5:50pm	S. Andrews and T. Jebara Structured Network Learning
6:15pm	Discussion Suggested Topic: Kernel and Distance Learning