Learning in Non-(geo)metric Spaces - Workshop @ ICML 2010
The SIMBAD FP7 Project
Beyond Features:
Similarity-Based Pattern Analysis and Recognition

1. Università Ca' Foscari di Venezia (IT), \textit{coordinator}
2. University of York (UK)
3. Technische Universität Delft (NL)
4. Insituto Superior Técnico, Lisbon (PL)
5. Università degli Studi di Verona (IT)
6. ETH Zürich (CH)
Pattern Recognition and Hume’s Similarity Principle

« I have found that such an object has always been attended with such an effect, and I foresee, that other objects, which are, in appearance, similar, will be attended with similar effects. »

David Hume

An Enquiry Concerning Human Understanding
(1748)
The Classical “Feature-based” Approach and Its Limitations

Traditional pattern recognition techniques are centered on the notion of feature, i.e. they derive similarities from vector representations.

But, there are various application domains where either it is not possible to find satisfactory features or they are inefficient for learning purposes.

This is typically the case, e.g.,

- when experts cannot define features in a straightforward way
- when data are high dimensional
- when features consist of both numerical and categorical variables,
- in the presence of missing or inhomogeneous data
- when objects are described in terms of structural properties, such as parts and relations between parts, as is the case in shape recognition
By departing from vector-space representations one is confronted with the challenging problem of dealing with (dis)similarities that do not necessarily possess the Euclidean behavior or not even obey the requirements of a metric.

The lack of the Euclidean and/or metric properties undermines the very foundations of traditional pattern recognition theories and algorithms!
SIMBAD aims at bringing to full maturation a paradigm shift that is currently just emerging within the pattern recognition and machine learning domains, where researchers are becoming increasingly aware of the importance of similarity information *per se*, as opposed to the classical feature-based approach.

The whole project will revolve around two main themes, which basically correspond to the two fundamental questions that arise when abandoning the realm of vectorial representations, namely:

- How can one *obtain* suitable similarity information from object representations that are more powerful than, or simply different from, the vectorial?

- How can one *use* similarity information in order to perform learning and classification tasks?
The structure of SIMBAD

1. Deriving similarities for non-vectorial data
   -- Structural (generative/compression) kernels
   -- Learning and combining similarities

2. Learning and classification with non-(geo)metric similarities
   -- Foundations of non (geo)metric similarities
   -- Imposing geometricity on non-geometric similarities (embedding)
   -- Learning with non-(geo)metric similarities (game theory)

3. Biomedical applications
   -- Analysis of tissue micro-array (TMA) images of renal cell carcinoma
   -- Analysis of brain magnetic resonance (MR) scans for the diagnosis of mental illness
For more information:

http://simbad-fp7.eu
Tentative Schedule:

Proposal submission: Next few days…
Call for Papers issued: End of Summer 2010
Submission deadline: October 2010
1st reviews: March 2011
Revised papers: July 2011
2nd reviews/decisions: October 2011
Publication: Early 2012
The SIMBAD Workshop Series

The first edition of the workshop (**SIMBAD 2011**) will take place in Italy, in the spring of 2011, at the end of the project (in conjunction with the final SIMBAD meeting).

**Format:**
- a few invited talks
- contributed oral/poster presentation
- panel discussion

Call for Papers issued early in summer 2010.

Within the workshop we also plan to run a **SIMBAD challenge** to ensure that the methods are compared on as level a playing field as possible.
We’ll ask PASCAL to support with infrastructures.
09:10 - 09:40  
**Shai Ben David** (University of Waterloo)  
*Clustering without any subjective similarity information*

09:40 - 10:10  
**Maria-Florina Balcan** (Georgia Tech)  
*Learning with general similarity functions*

10:10 - 10:40  
**Amnon Shashua** (The Hebrew University of Jerusalem)  
*On probabilistic hypergraph matching*

10:40 - 11:00  
Coffee break

11:00 - 11:30  
**Alex Smola** (Yahoo! Research)  
*From collaborative filtering to multitask learning*

11:30 - 12:00  
**Robert Krauthgamer** (The Weizmann Institute of Science)  
*A metric notion of dimension and its applications to learning*

12:00 - 14:00  
Lunch
14:00 - 14:30  
Nicolò Cesa-Bianchi (University of Milan)  
*Scalable algorithms for learning on graphs*

14:30 - 15:00  
Ulrike von Luxburg (Max Planck Institute for Biological Cybernetics)  
*A note of caution regarding distances on graphs*

15:00 - 16:00  
*Poster session* (includes refreshment)

16:00 - 17:00  
*Panel discussion:*
*Is non-(geo)metricity an issue for machine learning?*

Panelists:  
Shai Ben-David (University of Waterloo)  
Joachim Buhmann (ETH Zurich)  
Edwin Hancock (University of York, UK)  
Alex Smola (Yahoo! Research)