

## **Mathematics and Computer Science Doctoral School – Marseille/Toulon, France.**

### **Research Labs**

LIF (UMR 6166), BDAA research team, CMI, Technopôle de Chateau-Gombert, 39 rue Joliot-Curie 13453 Marseille cedex 13, France. <http://www.lif.univ-mrs.fr>

LSIS (UMR 6168), Information Dynamics project, Univ. Sud Toulon Var R229-BP20132-83957 La Garde CEDEX, France. <http://www.lsis.org>

### **Supervisors**

- François Denis, PR (LIF), [francois.denis@lif.univ-mrs.fr](mailto:francois.denis@lif.univ-mrs.fr)
- Hervé Glotin, MCF HDR (LSIS), [glotin@univ-tln.fr](mailto:glotin@univ-tln.fr)

**Title** : multimodal semi-supervised machine learning applied to the indexation of multimedia data

### **Description of the thesis:**

The main problem addressed in this research is the design and experimentation of semi-supervised machine learning methods and algorithms for inferring models from multimedia data. Such data lead up to many theoretical and algorithmical problems, because of the curse of dimension of multimodal partly annotated data. Most encountered approaches to deal with such data are empirical fusion based ([Ricardi and Hakkani-Tur, 2004], [Gupta et al., 2008]), : usually, data descriptors or classifiers are combined following a sequential or, less often a hierarchical, schema. [Kludas et al., 2007] presents a first formalism of data in the multimedia framework. Moreover as denoted in [Chapelle et al., 2006], the semi-supervised machine learning approaches are of interest each time that only few data are annotated among a huge amount of available data.

This thesis intends to deal with two, complementary research tracks that must be tackled simultaneously :

- To formally study several semi-supervised, statistical, machine-learning approaches in order to fit them to a multimodal framework and fusion-based schemes. Among these approaches to be studied, we plan to mainly focus on co-training [Blum et al., 1998] and active learning [Muslea 2006].
- To optimize above methods in a multimedia environment, through the performances maximization of multi-views (modalities) cooperation onto a same data set.

From the theoretical point of view, numerous directions have to be studied, for semi-supervised machine learning methods applied to the indexing of multimodal data are currently empirical hence hardly generally applicable. Past and current works on multimedia fusion should guide us to design first, generic solutions in a multimodal semi-supervised framework. However, the main objective of this thesis is to contribute to machine learning within a multimodal environment, independently from any application. Among the directions we are intended to follow, one of them is the core of the thesis. It concerns the relaxing of the

major assumption for co-training to be theoretically efficient (convergence). This hypothesis is that the two modalities that cooperate to draw a classifier must be conditionally independent the one from the other. Most of the time, such an hypothesis is not realistic in practice. Our aim is thus to theoretically study and control the relaxing of this hypothesis, then to design a two-views co-training like algorithm that achieves such a control. We plan to carefully experiment such an algorithm on benchmarked multimedia data. Such an algorithm might then be extended to a multi-views co-training-like algorithm.

Some other secondary directions might also be followed during the thesis in collaboration with members of the research group. For example, the above study of co-training together with active learning [Muslea et al., 2006] should help us to

exhibit relevant functions for dynamically optimizing the choice of the most efficient view at each step of co-training. Another – more prospective – direction is

to study co-training in a new theoretical framework, namely where annotated data are not independently and identically distributed (not i.i.d) in order to design algorithms more adapted to realistic multimedia data, where samples and examples are not as homogeneous as expected by usual approaches in statistical semi-supervised machine learning.

The theoretical and empirical study achieved during this thesis will be evaluated on scaled benchmarks. At the end of the thesis, a software shall be designed, that permits the experimentations of semi-supervised algorithms onto multimodal data,

and web data. In order to compare them to most recent ones, they will be applied on data of selected multimedia indexing international challenges, namely TRECVID. For such a purpose, the extraction of acoustic features (MFCC type), visual features (color, shape, texture...), and textual data, might be achieved efficiently and processed through approximative features selection (i.e. where labels are noisy as in [Glotin et al., 2006]).

The PhD student will integrate the working group « Web Multimedia Mining », which features researchers from two complementary labs in computer science (LIF: <http://www.lif.univ-mrs.fr> and LSIS: <http://www.lsis.org>), specialised in machine learning and multimedia data. The student must participate to some challenges, in particular TRECVID, with the research group, and should publish his (her) work in both machine learning and multimedia-oriented journals and conferences.

The thesis must start in fall 2009, for a 3-year duration.

**For more information:** please contact the supervisors.

### **Short references list**

[Balcan et al., 2005] M.F. Balcan and A. Blum. A PACstyle model for learning from labeled and unlabeled data. In Proceedings of Computational Learning Theory (COLT), pp111126, 2005.

[Blum et al., 1998] A. Blum and T. Mitchell. Combining labeled and unlabeled data with cotraining. Proceedings of Computational Learning Theory (COLT), pp92100, 1998.

[Chapelle et al., 2006]: Chapelle, O., B. Schölkopf and A. Zien. Semi-Supervised Learning. MIT Press, Cambridge, MA (2006)

[Glotin et al., 2006] H. Glotin, S. Tollari, Pascale Giraudet, "Shape reasoning on mis-segmented and mis-labeled objects using approximated Fisher criterion" in: Computers & Graphics, Elsevier, 30(2):177-184, April 2006.

[Gupta et al., 2008] S. Gupta, J. Kim, K. Grauman and R. Mooney. Watch, Listen & Learn: Co training on Captioned Images and Videos. Proceedings of European Conference on Machine Learning (ECML), pp 457472, 2008.

[Kludas et al., 2007] J. Kludas, E. Bruno, and S. MarchandMaillet. Information Fusion in Multimedia Information Retrieval. Proceedings of 5th international Workshop on Adaptive Multimedia Retrieval (AMR), Paris, France, 2007.

[Muslea et. al, 2006] Ion Muslea, Steven Minton, Craig A. Knoblock: Active Learning with Multiple Views. J. Artif. Intell. Res. (JAIR) 27: 203233, 2006.

[Riccardi and HakkaniTur, 2004] G. Riccardi and D. HakkaniTür. Active Learning: Theory and Applications to Automatic Speech Recognition. IEEE Transactions on Speech and Audio Processing, 13(4), 2005.

**Required skills** : fundamental computer science, statistics, machine learning, signal processing, multimedia data representation. Algorithmics and programming. The candidate must have a master degree in computer science. Matlab and C programming are welcome.