Ecole Doctorale Mathématiques et Informatique d'Aix-Marseille, ED184

Unités de recherche

LIF (UMR 6166), équipe BDAA, CMI, Technopôle de Chateau-Gombert, 39 rue Joliot-Curie 13453 Marseille cedex 13

LSIS (UMR 6168), équipe INCOD, Univ. Sud Toulon Var R229-BP20132-83957 La Garde CEDEX

Directeurs :

- Cécile Capponi, MCF (LIF), cecile.capponi@lif.univ-mrs.fr (assistée de F. Denis, Pr LIF)
- Hervé Glotin, MCF HDR (LSIS), glotin@univ-tln.fr

Titre de la thèse : Apprentissage semisupervisé multimodal, et application à l'indexation de données multimedia

Description :

La problématique centrale autour de laquelle se place cette thèse est le développement de méthodes et d'algorithmes inductifs pour l'apprentissage de modèles à partir de données multimodales, dans un cadre semi-supervisé. Avec ce type de données, de nombreuses problématiques, tant applicatives, algorithmiques, que théoriques sont à résoudre. Nous chercherons à optimiser nos résultats pour un traitement adapté de données multimédia en tenant compte de leur dimension multimodale, et du faible nombre de données annotées. Dans ce cadre, les approches actuelles sont principalement empiriques, basées sur la fusion : elles consistent à opérer des combinaisons de données et/ou de sorties de classifieurs, de façon séquentielle. On trouve dans [Kludas et al., 2007] un tout début de formalisation de la notion de fusion dans un cadre multimédia.

L'objectif principal de cette thèse est double, les deux axes étant complémentaires et devant être abordés simultanément :

1. étudier formellement les propriétés de plusieurs approches d'apprentissage statistique semi-supervisé [Balcan et al., 2005] dans un cadre multimodal, donc dans un cadre de fusion. Parmi ces approches, nous nous focaliserons sur le co-apprentissage [Blum et al., 1998] et l'apprentissage actif [Muslea 2006] ;

2. dériver un ou plusieurs algorithmes optimisant ces approches dans le cadre multimodal, en cherchant à maximiser les performances de la *coopération* entre les plusieurs vues (modalités) d'un même ensemble de données.

Les pistes à explorer, d'un point de vue informatique fondamentale, sont nombreuses car les approches d'apprentissage multimodal dans un cadre semi-supervisé restent empiriques donc difficilement généralisables. Notre orientation pourra dans un premier temps s'inspirer de quelques travaux en fusion multimodale, dans le cadre multimedia (annotation d'images web, de vidéos, de dialogue oraux, etc.), mais l'objectif majeur de la thèse reste une contribution en apprentissage automatique, fondamentalement basée sur la multimodalité des objets à annoter. Parmi les orientations que nous privilégierons, nous en avons identifié trois majeures ; toutes se basent sur l'algorithme de co-apprentissage dont nous cherchons à assouplir le cadre réel d'application.

1. Etude de l'impact du relâchement partiel de l'hypothèse d'indépendance des vues, centrale en co-apprentissage, sur les performances de l'algorithme. Cette étude nous permettra de concevoir des variantes opérationnelles du co-apprentissage, dans lesquelles l'hypothèse d'indépendance entre les vues est contrôlée. Ce type de recherche est fondamental dans la mesure où, en pratique, cette hypothèse est trop forte, et donc la convergence de l'algorithme n'est pas assurée. L'objectif ici est de garantir une convergence malgré une indépendance relative des vues, à caractériser.
2. Etude théorique du couplage co-apprentissage / apprentissage actif, dans la lignée des résultats de [Muslea et al., 2006]. En particulier, nous chercherons ici à optimiser dynamiquement le choix des vues de description lors de l'apprentissage coopératif, en exhibant des fonctionnelles caractérisant la qualité des classifieurs de chaque vue à un instant donné.
3. Etude de l'apprentissage semi-supervisé de type apprentissage actif (et coapprentissage, éventuellement), dans un cadre théorique non encore exploré, à savoir celui des exemples non i.i.d. (indépendamment et identiquement distribués). En effet, la grande majorité des algorithmes d'apprentissage semi-supervisé sont fondés sur l'hypothèse théorique, souvent fausse en pratique,

que les rares données étiquetées sont i.i.d. Nous pensons qu'étudier ces algorithmes sous un autre angle théorique permettrait d'élaborer d'autres approches plus réalistes face aux propriétés de distribution des données et échantillons.

Il est ici important de préciser que le travail de cette thèse devra être mené dans le soucis d'élaborer des approches passant à l'échelle, car le cadre multimédia et/ou web induit des volumétries gigantesques. A l'issue de ce travail, un prototype logiciel robuste devra avoir été réalisé pour l'expérimentation de protocoles d'annotation multimedia dans un cadre semi-supervisé.

Les résultats théoriques et algorithmiques devront être systématiquement éprouvés sur des données multimedia partagées par la communauté de fouille de données multimedia et web (par exemple, données Corel, TRECVID, CLEF, ESTER, etc.). Ici, l'extraction des caractéristiques accoustiques (type MCFF), visuelles (type SIFT, shistogramme de couleur, Gabor, etc.),ou encore textuelles après traitement classique de type Porter, pourra faire usage de sélection de traits ou d'approximation de sélection de traits dans ce cadre de données mal-étiquetées, comme le propose [Glotin et al., 2006]. Il est important de souligner l'apport de ce travail fondamental de thèse dans le cadre plus spécifique du multimedia. En effet, l'annotation de données multimedia dans un cadre semi-supervisé a été essentiellement étudiée empiriquement par application ad-hoc de techniques d'apprentissage ([Ricardi and Hakkani-Tur, 2004], [Gupta et al., 2008] par exemple). De ce fait, aucune étude spécifique des algorithmes utilisés n'a été menée dans ce cadre, et aucune variante de ces approches n'a donc été proposée. Nous pensons que, dans le cadre de cette thèse, des algorithmes plus performants dans le cadre multimedia peuvent être proposés en exploitant le caractère multimodal des données et en formaisant la notion de fusion.

Durant cette thèse, le doctorant intègrera le groupe WMM créé en septembre 2008 à Marseille auquel appartiennent les deux co-encadrants de cette thèse. Le doctorant participera à plusieurs challenges internationaux de type TRECVID ou CLEF, et sera amené(e) à publier ses résultats dans des revues et conférences spécifiques à l'apprentissage automatique, mais aussi dédiées au multimédia.

[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.

[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.

Connaissances et compétences : informatique fondamentale, éléments de statistiques, apprentissage automatique, éléments de traitement du signal, représentation des images/vidéos/sons. Compétences en algorithmique et programmation.

Thème prioritaire : (1.d) STIC__Recherches fondamentales en science informatique__multimédia