

Colloque du
Département d'Informatique et de Recherche Opérationnelle

Simon Lacoste-Julien

Laboratoire d'Informatique de l'École Normale Supérieure, Paris, France

Exploiter la structure des données pour l'apprentissage statistique discriminatif

Résumé: Le domaine de l'apprentissage statistique s'est grandement développé durant les deux dernières décennies, avec un impact dans des domaines divers tels que la vision par ordinateur, le traitement automatique du langage, la biologie computationnelle et les sciences sociales. Par contre, ces nouvelles applications ont mis en évidence que les données issues de tâches réelles possèdent souvent une structure plus riche que celle normalement modélisée par les paradigmes classiques, tels que la classification binaire ou la régression. En traduction automatique par ordinateur, par exemple, l'algorithme doit choisir parmi un nombre exponentiel de séquences de mots possibles en tant que traduction, et non un simple petit nombre de choix tel que pour la reconnaissance automatique de chiffres écrits. Un défi clé pour le domaine de l'apprentissage statistique moderne est de trouver des moyens de modéliser les structures complexes de manière évolutive ('scalable') tout en restant robuste à la mauvaise spécification du modèle. Dans cette présentation, je vais exposer comment exploiter deux types de structure sur les données : une structure combinatoire représentée par des graphes pour la tâche d'alignement de mots pour le traitement automatique du langage, et une structure probabiliste avec variable latente utilisée dans la modélisation de texte pour la classification automatique de documents et la recherche d'information. Une contribution clé de l'approche proposée est de rendre les modèles robustes en les apprenant avec un critère d'apprentissage *discriminatif*, ce qui permet d'obtenir des gains substantiels de performance. Un autre thème clé sera la 'scalabilité' – je vais expliquer comment utiliser des outils modernes d'optimisation pour maintenir cette propriété. Ces exemples mettront en évidence comment l'applicabilité des algorithmes d'apprentissage statistique peut être accrue en exploitant de manière plus appropriée la structure des données complexes.

Biosketch: Simon Lacoste-Julien est présentement un Research in Paris fellow à l'École Normale Supérieure / INRIA. Il a complété un B.Sc. Triple Honours en mathématiques, physique et informatique à l'Université McGill en 2003, un Ph.D. en informatique et statistiques à UC Berkeley en 2009 sous la supervision de Michael I. Jordan, et il a ensuite travaillé en tant que post-doc avec Zoubin Ghahramani à l'Université Cambridge. Il a obtenu la McGill Moyses Travelling Scholarship ainsi que le UC Berkeley College of Engineering Graduate Student Prize. Ses problématiques de recherche comprennent l'apprentissage statistique, l'optimisation et les statistiques, avec comme applications le traitement automatique du langage, la vision par ordinateur et la recherche de données.



Le vendredi 17 février 2012, 10h45, pavillon Aisenstadt, salle 5340

Information : Pierre L'Ecuyer, <http://www.iro.umontreal.ca/~lecuyer/>