

# Apprentissage semi-supervisé de fonctions d'ordonnement

Vinh Truong, Massih-Reza Amini

Laboratoire d'Informatique de Paris 6  
104, Avenue du Président Kennedy, 75016 Paris, France  
{truong, amini}@poleia.lip6.fr  
<http://www-connex.lip6.fr>

**Résumé.** Nous présentons dans cet article un algorithme inductif semi-supervisé pour la tâche d'ordonnement bipartite. Les algorithmes semi-supervisés proposés jusqu'à maintenant ont été étudiés dans le cadre strict de la classification. Récemment des travaux ont été réalisés dans le cadre transductif pour étendre les modèles existants en classification au cadre d'ordonnement. L'originalité de notre approche est qu'elle est capable d'inférer un ordre sur une base test non-utilisée pendant la phase d'apprentissage, ce qui la rend plus générique qu'une méthode transductive pure. Les résultats empiriques sur la base CACM contenant les titres et les résumés du journal *Communications of the Association for Computer Machinery* montrent que les données non-étiquetées sont bénéfiques pour l'apprentissage de fonctions d'ordonnement.

## 1 Introduction

Avec le développement des bibliothèques électroniques, il est devenu nécessaire de concevoir des méthodes automatiques pour la recherche de données pertinentes par rapport à une requête donnée. Pour de telles applications, il s'agit plus d'ordonner les exemples que de les discriminer.

La communauté d'apprentissage a formulé cette problématique à travers le nouveau paradigme d'apprentissage supervisé de fonctions d'ordonnement. Dans ce cas, il s'agit d'apprendre une correspondance entre un ensemble d'instances et un ensemble d'alternatives capable d'ordonner les alternatives par rapport à une instance donnée. Par exemple, dans le cas de la recherche documentaire (RD), une instance représente une requête et les alternatives sont les documents concernés par cette requête et le but est d'inférer un ordre partiel sur l'ensemble des alternatives de façon à ce que les documents pertinents par rapport à la requête soient mieux ordonnés que les documents non-pertinents.

Dans ce papier, nous nous plaçons dans le cadre de l'ordonnement bipartite dans lequel les instances sont soit positives soit négatives et où il s'agit d'ordonner les instances positives au-dessus des instances négatives. Ce cadre restreint englobe de nombreuses applications de la recherche d'information telle que le résumé automatique (Amini et al., 2005) ou la recherche