Résumé

Ces travaux concernent la problématique générale de la classification, et divers domaines d'applications en informatique : l'apprentissage automatique, le data mining et l'analyse d'image. Deux grands types de résultats ont été obtenus : ceux produisant des algorithmes, leur implémentation, leur analyse de performance (théorique et sur test), et ceux concernant des résultats négatifs de complexité, bornant inférieurement les performances de n'importe quel algorithme sur ces problèmes. Les algorithmes concernés par la première catégorie sont des algorithmes d'induction de classes de formules particulières, des algorithmes de plus proche voisins, des algorithmes de réduction de données (sélection de variables et/ou d'exemples), et des algorithmes de segmentation d'images. Un certain nombre de ces algorithmes utilisent des propriétés de concentration de variables aléatoires. Les résultats concernés par la deuxième catégorie sont essentiellement des preuves d'inapproximabilité de problèmes d'optimisation, qui admettent pour certains des extensions dans divers modèles d'apprentissage automatique, avec divers objectifs : minimisation de taille de formules (Formes Normales Disjonctives, ensembles de Clauses de Horn, etc.), minimisation de l'erreur sous contrainte, minimisation de critères alternatifs à l'erreur, minimisation de taille de jeux de données, etc.

Mots-clés: Analyse d'Images, Apprentissage Automatique, Data Mining.

Abstract

This work deals with classification related topics, with applications to domains such as machine learning, data mining, and image processing. Two categories of results are presented : those leading to algorithms, their implementations and performance analyzes, and those producing complexity-theoretic results, lowerbounding the performances of any potential algorithm for these or related problems. The algorithms concerned by the first category of results are induction algorithms (for various classes of concept representation), nearest neighbor algorithms, data reduction algorithms (prototype/feature selection), and image segmentation algorithms. Some of these algorithms use results on the concentration phenomena of random variables. The complexity-theoretic results are essentially inapproximability results for various optimization problems, some of which admit extensions in formal models of learning with various objectives : minimization of the output size (Disjunctive Normal Form formulas, variously restricted subsets of Horn Clauses, etc.), constraint-based error minimization, optimization of error replacement criteria, dataset size reduction, etc.

Keywords: Data Mining, Image Processing, Machine Learning.