

Un raisonnement approximatif pour l'apprentissage supervisé de règles

Amel Borgi

Unité de recherche SOIE, ENSI, Tunis
Institut National des Sciences Appliquées et de la Technologie
Centre Urbain Nord de Tunis
BP 676, 1080 Tunis Cedex, Tunisie
Amel.Borgi@insat.rnu.tn

Résumé. Le cadre de ce travail est celui de la méthode d'apprentissage supervisé SUCRAGE qui se base sur la génération automatique de règles de classification. Ces règles sont exploitées par un moteur d'inférence classique : seules les règles dont les prémisses sont vérifiées par la nouvelle observation à classer sont déclenchées. Ce moteur a été étendu à une inférence approximative qui permet de déclencher les règles pas trop éloignées de la nouvelle observation. Nous proposons une utilisation originale du raisonnement approximatif non plus comme un mode d'inférence mais comme un moyen d'affiner l'apprentissage. Le raisonnement approximatif est utilisé pour générer de nouvelles règles dont les prémisses sont élargies : les imprécisions des observations sont alors prises en compte et les problèmes liés à la discrétisation des attributs continus sont atténués. Notre approche a été testée avec différentes bases d'apprentissage et confrontée à une application réelle dans le domaine du traitement d'images.

1 Introduction

Le cadre général de notre travail est celui de l'apprentissage supervisé pour la fouille de données, et plus précisément l'apprentissage supervisé par génération de règles (Duch et al., 2004) (Zhou 2003). La simplicité et la facilité d'interprétation des règles de production font qu'elles constituent une forme de représentation des connaissances largement utilisée dans les systèmes d'apprentissage (Duch et al., 2004) (Haton et al. 1991) (Holmes et al., 2002) (Mikut et al., 2005) (Prentzas et al., 2005) (Rakotomalala, 2005).

Parmi les nombreux travaux portant sur des méthodes d'apprentissage inductif par génération de règles, nous pouvons citer ceux portant sur les arbres de décision (Quinlan, 1986, 1993) (Breiman et al., 1984) (Rakotomalala, 2005), ou les méthodes de type « graphes d'induction » (Zighed et al., 2002). La fonction de classement y est donnée sous forme d'arbre ou de graphe, elle se traduit aisément en une base de règles. Dans ces méthodes dites *mono-attributs* ou *monothétiques*, les prémisses des règles sont construites étape par étape. A chacune d'elles, on ajoute une condition sur un meilleur attribut, et le choix du meilleur attribut se fait selon le pouvoir discriminant relativement aux classes de cet attribut pris seul. D'autres méthodes, qualifiées de *multi-attributs* ou *polythétiques* présentent l'avantage de sélectionner en bloc les attributs qui apparaissent dans les prémisses des règles (DiPalma et al. 97). Parmi ces approches nous pouvons citer des méthodes symboliques explorant de