

# Visualisation exploratoire des résultats d'algorithmes d'arbre de décision

Thanh-Nghi Do\*, Nguyen-Khang Pham\*\*, François Poulet\*\*\*

\*Equipe InSitu, INRIA Futurs, LRI, Bat.490, Université Paris Sud 91405 Orsay Cedex

Thanh-Nghi.Do@lri.fr

<http://www.lri.fr/~dtngnhi>

\*\*Equipe Texmex, IRISA, 35042 Rennes Cedex

pnguyenk@irisa.fr

\*\*\*ESIEA-Ouest, 38, rue des Docteurs Calmette et Guérin, 53000 Laval

francois.poulet@esiea-ouest.fr

<http://visu.egc.free.fr>

**Résumé.** Nous présentons une méthode d'exploration des résultats des algorithmes d'apprentissage par arbre de décision (comme C4.5). La méthode présentée utilise simultanément une visualisation radiale, focus+context, fisheye et hiérarchique pour la représentation et l'exploration des résultats des algorithmes d'arbre de décision. L'utilisateur peut ainsi extraire facilement des règles d'induction et élaguer l'arbre obtenu dans une phase de post-traitement. Cela lui permet d'avoir une meilleure compréhension des résultats obtenus. Les résultats des tests numériques avec des ensembles de données réelles montrent que la méthode proposée permet une bien meilleure compréhension des résultats des arbres de décision.

## 1 Introduction

Le volume de données stocké double actuellement tous les 9 mois (Lyman et al, 2003) et donc le besoin d'extraction de connaissances dans les grandes bases de données est de plus en plus important (Fayyad et al, 2004). La fouille de données (Fayyad et al, 1996) vise à traiter des ensembles de données pour identifier des connaissances nouvelles, valides, potentiellement utilisables et compréhensibles. Cette utilisabilité est fonction des buts de l'utilisateur donc seul l'utilisateur peut déterminer si les connaissances extraites répondent à ses attentes. Les outils de fouille de données doivent donc être interactifs et anthropocentrés. Notre approche consiste à impliquer plus fortement l'utilisateur dans le processus de fouille par des méthodes graphiques interactives dans un environnement de fouille.

De nombreuses méthodes de visualisation ont été développées dans différents domaines et utilisées pour l'analyse exploratoire et la fouille de données (Fayyad et al, 2001), (Keim, 2002). Les méthodes de visualisation peuvent être utilisées pour le pré-traitement de données (par exemple la sélection de données) ou en post-traitement (par exemple pour voir les résultats). Des méthodes récentes (Ankerst, 2001), (Do et Poulet, 2004a et b), (Munzner, 1997) essayent d'impliquer plus significativement l'utilisateur dans le processus de fouille de