

Utilisation de métadonnées pour l'aide à l'interprétation de classes et de partitions

Abdourahamane Baldé*, Yves Lechevallier*,
Brigitte Trousse**, Marie-Aude Aaufaure***

* INRIA Rocquencourt (Projet AxIS)
Domaine de Voluceau Rocquencourt, B.P. 105, F-78153 Le Chesnay Cedex, France
{abdourahamane.balde, yves.lechevallier}@inria.fr

** INRIA Sophia Antipolis (Projet AxIS)
Route des Lucioles, B.P. 93, F-06902 Sophia Antipolis Cedex, France
brigitte.trousse@inria.fr

***Supélec - Plateau du Moulon - Service Informatique
F-91192 Gif-sur-Yvette Cedex
marie-aude.aufaure@supelec.fr

Résumé. Les résultats des méthodes de fouille de données sont difficilement interprétables par un utilisateur n'ayant pas l'expertise requise. Dans ce papier nous proposons un outil permettant aux utilisateurs d'interpréter les résultats issus des méthodes de classification non supervisée. Cet outil est basé sur des métadonnées utilisées pour formaliser le processus d'interprétation automatique. Ces métadonnées vont servir à l'utilisateur pour comprendre dans quelles circonstances les données originales ont été collectées et de quelle manière elles ont été agrégées puis classifiées. L'intérêt de ce travail porte sur la souplesse qu'auront les utilisateurs à pouvoir interpréter facilement les classes obtenues. Nous développons notre approche basée sur l'utilisation des métadonnées. Nous traduirons notre méthodologie par un exemple concret.

1 Introduction

La fouille de données définie comme étant l'extraction à partir de données brutes de connaissances potentiellement exploitables, n'en demeure pas moins un processus complexe dès lors qu'il s'agit d'interpréter les résultats fournis. Les techniques de fouille de données représentent une étape fondamentale du processus d'Extraction de Connaissances dans les Bases de Données connu sous le nom ECD ou KDD (Knowledge Discovery in Databases) (Han 2001).

Dans ce papier nous nous intéressons à l'une de ces techniques : la classification non supervisée. Celle-ci est définie comme un ensemble de processus aptes à être exécutés sur ordinateur pour constituer des hiérarchies de classes ou de simples partitions établies à partir de tableaux de données (Jambu 1978). Les règles d'interprétation des structures classificatoires obtenues (hiérarchies, partitions, etc.) à l'issue de ces classifications n'ont pas la simplicité des méthodes descriptives uni-dimensionnelles.