

Extraction de Règles d'Association Quantitatives Application à des Données Médicales

Cyril Nortet* Ansaf Salleb**
Teddy Turmeaux* Christel Vrain*

*LIFO Rue Léonard de Vinci BP 6759 45067 Orléans Cedex 02
{Cyril.Nortet, Teddy.Turmeaux, Christel.Vrain}@lifo.univ-orleans.fr

**IRISA/INRIA, Projet Dream, Campus de Beaulieu, Rennes
Ansaf.Salleb@irisa.fr

Résumé. L'extraction de règles d'association est devenue aujourd'hui une tâche populaire en fouille de données. Cependant, l'algorithme Apriori et ses variantes restent dédiés aux bases de données renfermant des informations catégoriques.

Nous proposons dans cet article QUANTMINER, qui est un outil que nous avons développé dans le but d'extraire des règles d'association gérant variables catégoriques et numériques. L'outil que nous proposons repose sur un algorithme génétique permettant de découvrir de façon dynamique les intervalles des variables numériques apparaissant dans les règles.

Nous présentons également une application réelle de notre outil sur des données médicales relatives à la maladie de l'athérosclérose et donnons des résultats de notre expérience pour la description et la caractérisation de cette maladie.

Mots clé : Fouille de Données, Règle d'Association, Attribut Numérique, Discrétisation, Attribut Catégorique, Algorithme Génétique.

1 Introduction

L'extraction de règles d'association est devenue aujourd'hui une tâche populaire en fouille de données. Elle a pour but de dégager des relations intelligibles entre des attributs dans une base de données. Une règle d'association (Agrawal et al. 1993) est une implication $\mathcal{C}_1 \Rightarrow \mathcal{C}_2$, où \mathcal{C}_1 et \mathcal{C}_2 expriment des conditions sur les attributs de la base de données. La qualité d'une règle est classiquement évaluée par un couple de mesures *support* et *confiance*, définis comme suit :

- $\text{Support}(\mathcal{C})$, où \mathcal{C} exprime des conditions sur les attributs, est le nombre de n-uplets (lignes de la table) qui satisfont \mathcal{C} .
- $\text{Support}(\mathcal{C}_1 \Rightarrow \mathcal{C}_2) = \text{Support}(\mathcal{C}_1 \wedge \mathcal{C}_2)$
- $\text{Confiance}(\mathcal{C}_1 \Rightarrow \mathcal{C}_2) = \frac{\text{Support}(\mathcal{C}_1 \wedge \mathcal{C}_2)}{\text{Support}(\mathcal{C}_1)}$

Étant donnés deux seuils, MINSUPP et MINCONF, une règle est dite *solide*, si son support et sa confiance dépassent MINSUPP et MINCONF respectivement.

Quoique très utilisés, l'algorithme *Apriori* et ses variantes restent dédiés aux bases de données renfermant des informations catégoriques (dites aussi symboliques ou encore