

Un automate pour la génération complète ou partielle des concepts du treillis de Galois

Ganaël Jatteau, Rokia Missaoui
M. Sarifuddin

Département d'informatique et d'ingénierie
Université du Québec en Outaouais
C.P. 1250, succursale B, Gatineau
Québec, J8X 3X7
{jatg01, Rokia.Missaoui, M.Sarifuddin}@uqo.ca

Résumé. Cet article se situe dans le domaine de l'analyse formelle de concepts et du treillis de concepts (treillis de Galois) lequel est un cadre théorique intéressant pour le regroupement conceptuel des données et la génération des règles d'association. Puisque la prospection de données (*data mining*) est utilisée comme support à la prise de décision par des analystes rarement intéressés par la liste exhaustive (souvent très longue) des concepts et des règles, l'élaboration d'une solution approximative sera dans la plupart des cas un compromis satisfaisant et relativement moins coûteux qu'une solution exhaustive. Dans cet article, on propose une approche appelée CIGA (*Closed Itemset Generation using an Automata*) de génération partielle ou complète de concepts par la construction et le parcours d'un automate à états finis. La génération des concepts permet l'identification des "itemsets" fermés fréquents, étape cruciale pour l'extraction des règles d'association.

1 Introduction

L'analyse formelle de concepts (treillis de Galois) est un cadre théorique intéressant pour la prospection de données puisqu'elle permet la génération de concepts et de règles d'association. Un concept formel est un couple complet qui associe un ensemble d'objets (extension) à un ensemble d'attributs (intention) permettant ainsi de regrouper les objets qui ont des caractéristiques communes.

Dans plusieurs applications de prospection de données, la production d'un ensemble *exhaustif* de connaissances (règles d'association, concepts) peut être très coûteuse et comporter plusieurs éléments absolument peu pertinents pour un utilisateur donné. Aussi, il serait avantageux d'offrir des mécanismes de génération d'un sous-ensemble de ces connaissances qui pourraient si nécessaire inciter l'utilisateur soit à solliciter l'affichage d'autres connaissances ou à demander des détails sur les associations et les concepts issus d'un ensemble plus restreint de données.

La découverte des règles d'association se fait généralement en deux étapes : (i) la détermination de l'ensemble des "itemsets" fréquents (i.e., ceux dont le support dépasse un seuil déterminé), puis (ii) la génération des règles d'association à partir des "itemsets" fréquents obtenus à la première étape.