

Un cadre théorique pour la gestion de grandes bases de motifs¹

François Jacquenet, Baptiste Jeudy et Christine Largeron

Laboratoire Hubert Curien, UMR CNRS 5516, St-Etienne
prenom.nom@univ-st-etienne.fr

Résumé. Les algorithmes de fouille de données sont maintenant capables de traiter de grands volumes de données mais les utilisateurs sont souvent submergés par la quantité de motifs générés. En outre, dans certains cas, que ce soit pour des raisons de confidentialité ou de coûts, les utilisateurs peuvent ne pas avoir accès directement aux données et ne disposer que des motifs. Les utilisateurs n'ont plus alors la possibilité d'approfondir à partir des données initiales le processus de fouille de façon à extraire des motifs plus spécifiques. Pour remédier à cette situation, une solution consiste à gérer les motifs. Ainsi, dans cet article, nous présentons un cadre théorique permettant à un utilisateur de manipuler, en post-traitement, une collection de motifs préalablement extraite. Nous proposons de représenter la collection sous la forme d'un graphe qu'un utilisateur pourra ensuite exploiter à l'aide d'opérateurs algébriques pour y retrouver des motifs ou en chercher de nouveaux.

1 Introduction

La quantité d'information stockée dans les bases de données du monde entier n'a cessé de croître au cours du temps. Afin d'explorer ces mines potentielles de connaissance, des outils de fouille de données ont été conçus depuis plusieurs années. Ainsi, il est maintenant possible de fouiller de très grandes bases de données afin d'en extraire toute sorte de motifs modélisant de la connaissance. Selon les outils utilisés par les utilisateurs finaux pour leurs besoins, les motifs extraits peuvent être très variés. On peut citer par exemple les arbres de décision, les règles d'association, les concepts formels, etc. Toutefois, alors que la fouille de très grandes bases de données est devenue une tâche relativement aisée pour les utilisateurs finaux, ceux-ci sont maintenant confrontés à un nouveau problème : comment vont-ils pouvoir exploiter les grandes masses de motifs couramment extraites par les outils de fouille de données ? En fait, de la même façon qu'il était impossible il y a quelques années d'extraire manuellement de la connaissance à partir de grandes bases de données, il est de nos jours impossible de gérer de très grands volumes de motifs et les utilisateurs finaux sont donc en attente de nouveaux outils pour résoudre ce problème.

¹Ce travail a été partiellement soutenu par le projet BINGO de l'ACI Masse de données 2004 - 2007, financé par le ministère de la recherche.