

Symétries et Extraction de Motifs Ensemblistes

Said Jabbour*, Mehdi Khiari**, Lakhdar Sais*, Yakoub Salhi *, Karim Tabia*

*CRIL UMR CNRS 8188, Université d'Artois, F-62300 Lens,
{jabbour, sais, salhi, tabia}@cril.univ-artois.fr,
<http://www.cril.univ-artois.fr>

**Zero To One Technology, Campus Effiscience, F-14460 Colombelles,
mehdi.khiari@zto-technology.com,
<http://zto-technology.com>

Résumé. Les symétries sont des propriétés structurelles qu'on détecte dans un grand nombre de bases de données. Dans cet article, nous étudions l'exploitation des symétries pour élaguer l'espace de recherche dans les problèmes d'extraction de motifs ensemblistes. Notre approche est basée sur une intégration dynamique des symétries dans les algorithmes de type Apriori permettant de réduire l'espace des motifs candidats. En effet, pour un motif donné, les symétries nous permettent de déduire les motifs qui lui sont symétriques et vérifiant par conséquent les mêmes propriétés. Nous détaillons notre approche en utilisant l'exemple des motifs fréquents. Ensuite, nous la généralisons au cadre unificateur de Mannila et Toivonen pour l'extraction des motifs ensemblistes. Les expériences menées montrent la faisabilité et l'apport de notre approche d'élagage basé sur les symétries.

1 Introduction

Le but de cet article est d'introduire un cadre d'élagage basé sur les symétries pour la fouille de données. Dans d'autres domaines, tels que la programmation par contraintes et la satisfaisabilité propositionnelle, les symétries sont souvent exploitées pour élaguer l'espace de recherche et améliorer les performances des solveurs. Notre but est de montrer l'apport d'un élagage à base de symétries dans un cadre de fouille de données. Dans cet article, nous nous intéressons au cas de l'extraction de motifs ensemblistes et nous prenons comme exemple d'application l'algorithme APRIORI (Agrawal et Srikant, 1994).

Les symétries sont un concept fondamental en informatique, mathématiques, physiques et plein d'autres domaines. Elles existent dans divers problèmes réels. Les symétries sont communément exploitées dans la résolution de problèmes combinatoires, tels que les problèmes d'ordonnement. Par exemple, dans un problème d'ordonnement où certaines machines peuvent être interchangeables, partant d'un ordonnancement valide, on peut en obtenir un autre valide en permutant ces machines. Exploiter les symétries permet de réduire le coût de la recherche de solution, en évitant d'explorer des branches symétriques de l'espace de recherche. Beaucoup de travaux ont ainsi porté sur les symétries dans les problèmes de satisfaction de contraintes (CSP) (e.g. (Puget, 1993; Gent et Smith, 2000)), satisfaisabilité propositionnelle