

Représentation condensée de règles d'association multidimensionnelles

Alexandre Bazin*, Aurélie Bertaux** et Christophe Nicolle***

Le2i, Université de Bourgogne
21000 Dijon

*contact@alexandrebazin.com

**aurelie.bertaux@iut-dijon.u-bourgogne.fr

***cnicolle@u-bourgogne.fr

Résumé. La fouille de règles d'association est un problème qui a donné lieu à une littérature foisonnante, notamment dans les données binaires bidimensionnelles classiques. En particulier, la relation entre les ensembles fermés et les règles d'association est bien connue. Tel n'est pas le cas dans les données multidimensionnelles. Dans ce papier, nous montrons que la connaissance des n -ensembles fermés d'un tenseur booléen multidimensionnel est suffisante pour inférer la confiance de toutes les règles d'association multidimensionnelles.

1 Introduction

Le calcul de règles d'association (Agrawal et al. (1993)) est un problème important en fouille de données qui a donné lieu à une littérature foisonnante. Dès sa naissance, et pour pallier au grand nombre de motifs produits, l'accent a été mis sur la recherche d'ensembles réduits de règles contenant une information jugée intéressante. Comme souvent lorsque deux critères sont à optimiser – ici le nombre de règles et l'information contenue –, l'un d'eux prend le pas. Ainsi, dans le domaine des règles d'association, le nombre de règles est souvent vu comme primordial.

Le problème de la représentation de la totalité des règles – et donc de l'ensemble de l'information – est celui qui nous intéresse ici. Dans le cas de données binaires bidimensionnelles, le premier à être considéré, la question n'est plus ouverte. Nous savons que les règles peuvent être représentées de façon minimale par les ensembles fermés. Ce résultat, basé sur le fait que les fermés sont des représentants uniques de leurs classes d'équivalence vis à vis du support, a donné lieu à de nombreuses combinaisons avec les mesures d'intérêt utilisées pour réduire le nombre de règles au détriment de l'information.

Dans ce papier, nous considérons le cas des données binaires multidimensionnelles. Bien qu'il ait été moins étudié que le cas bidimensionnel, des moyens de réduire le nombre de règles ont déjà été proposées en généralisant la mesure d'intérêt la plus connue : la fréquence. Cependant, à notre connaissance, aucun résultat n'existe sur des représentations condensées de l'entièreté des règles. Nous nous proposons ici d'y remédier en montrant que, tout comme dans le cas bidimensionnel, les ensembles n -fermés d'une transformation d'un tenseur booléen