

# Fouille de motifs sans seuil par optimisation multi-objectifs : Application aux règles d'association

Charles Vernerey\*, Samir Loudni\*, Nouredine Aribi\*\* Yahia Lebbah\*\*

\* MT Atlantique, 44307 Nantes, France  
{samir.loudni, charles.vernerey}@imt-atlantique.fr  
\*\* Université Oran1, Lab. LITIO, 31000 Oran, Algeria  
{lebbah.yahia, aribi.nouredine}@univ-oran1.dz

**Résumé.** Cet article propose un nouveau modèle pour extraire les motifs Pareto dominants à l'aide de la programmation par contraintes. Notre modèle exploite le principe de la représentation condensée pour réduire l'espace de recherche. Nous démontrons que notre approche peut être utilisée pour découvrir des règles d'association intéressantes pour l'utilisateur sans avoir à fixer de seuil. Des expérimentations menées sur un jeu de données génomique ont démontré l'intérêt de cette approche pour l'extraction de règles d'association.

## 1 Introduction

L'extraction de motifs est une tâche importante en fouille de données, l'objectif étant d'extraire des motifs qui peuvent être interprétés par des experts du domaine ou utilisés comme descripteurs dans d'autres tâches comme par exemple la classification. Depuis la publication de l'article précurseur Agrawal and Srikant [1994], deux problèmes ont limité l'usage de cette approche : 1) comment fixer des seuils qui sont nécessaires dans plusieurs contraintes et 2) comment traiter des résultats qui contiennent parfois des millions de motifs. Utiliser l'approche *top-k* présente un inconvénient majeur du fait qu'il est difficile de choisir la valeur de  $k$ . Faire un post traitement des résultats à l'aide des représentations condensées [Pasquier *et al.*, 1999] ou la fouille d'ensemble de motifs [De Raedt and Zimmermann, 2007] ne fait que repousser le problème.

Depuis une dizaine d'années plusieurs approches ont été proposées pour découvrir des interactions plus complexes entre les motifs. Un exemple d'interaction est l'optimisation multi-objectifs (MO) où plusieurs objectifs (souvent antagonistes) doivent être optimisés simultanément. Peu d'approches ont été proposées concernant la découverte de motifs et la MO. Ghosh and Nath [2004] ont proposé une approche MO où des algorithmes génétiques ont été utilisés. van Leeuwen and Ukkonen [2013] ont proposé un algorithme pour la fouille de sous-groupes skylines. La notion de motif skyline a été exploitée dans [Ugarte *et al.*, 2017] pour extraire des motifs de haut niveau par rapport à plusieurs mesures.

Cet article s'inscrit dans la voie qui vise à connecter la programmation par contraintes (PPC) à la fouille de motifs [Guns *et al.*, 2011]. Nous proposons un nouveau modèle PPC compact et flexible pour découvrir des motifs Pareto optimaux (a.k.a. skypatterns) par rapport