

Allier CSPs et motifs locaux pour la découverte de motifs sous contraintes n-aires

Mehdi Khiari, Patrice Boizumault, Bruno Crémilleux

GREYC, CNRS - UMR 6072, Université de Caen,
Campus Côte de Nacre,
F-14032 Caen Cedex, France
{Prenom.Nom}@info.unicaen.fr

Résumé. Dans cet article, nous étudions la relation entre la découverte de motifs sous contraintes et les CSPs (Constraint Satisfaction Problems) afin de définir des contraintes de plus haut niveau qui sont précieuses pour mener à bien des tâches de fouille de données. Pour cela, nous proposons une approche de modélisation et d'extraction de motifs sous contraintes n-aires exploitant les motifs locaux. L'utilisateur définit un ensemble de contraintes n-aires et un solveur de CSP génère l'ensemble des solutions. Notre approche profite des progrès récents sur l'extraction de motifs locaux et permet de modéliser de manière concise et élégante tout ensemble de contraintes combinant plusieurs motifs locaux, permettant ainsi la découverte de motifs répondant mieux aux buts finaux de l'utilisateur. Les expériences menées montrent la faisabilité de notre approche.

1 Introduction

Un problème majeur dans les processus d'Extraction de Connaissances dans les Bases de Données (ECBD) est le nombre conséquent de motifs produits rendant leur utilisation difficile. Ainsi, les motifs les plus significatifs sont souvent noyés au milieu d'informations triviales ou redondantes. D'autre part, l'intérêt d'un motif dépend souvent d'autres motifs et beaucoup de modèles comme les classifieurs ou les méthodes de clustering requièrent d'exploiter et de considérer simultanément plusieurs motifs. Contribuer à la découverte de motifs de plus haut niveau, fondés sur la combinaison de motifs tels que les motifs locaux (cf. section 2.1), permet ainsi la découverte de motifs plus proches des buts finaux de l'utilisateur.

Il existe plusieurs approches visant à réduire les motifs produits, telles que le paradigme de l'extraction sous contraintes Ng et al. (1998), l'utilisation des représentations condensées Calders et al. (2005) ou encore la compression des jeux de données en exploitant le principe du MDL Siebes et al. (2006). En ciblant la recherche de l'information à extraire suivant les centres d'intérêt de l'utilisateur, le paradigme des motifs contraints a pour but d'améliorer la qualité des motifs extraits Ng et al. (1998). Ce problème est plutôt bien maîtrisé via des approches génériques de découverte de *motifs locaux* sous contraintes De Raedt et al. (2002); Soulet et Crémilleux (2005). On appelle *contraintes locales* les contraintes visant à extraire les motifs locaux. Le caractère local de la contrainte provient du fait que, vérifier si un motif satisfait ou pas une contrainte donnée, ne dépend pas des autres motifs. Dans la pratique, même si le