

Exploitation des techniques de fouille de données pour la compression de contraintes table

Soufia Bennai *, Kamal Amroun **, Samir Loudni ***

*LIMED - Faculté des Sciences Exactes, Université de Bejaia, Algérie
sofia.bennai21@gmail.com

**LIMED - Faculté des Sciences Exactes, Université de Bejaia, Algérie
kamalamroun@gmail.com

***GREYC (CNRS UMR 6072), Université de Caen Normandie, France
samir.loudni@unicaen.fr

Résumé. Dans ce papier, nous proposons une amélioration de l'étape de compression de la méthode proposée par Gharbi et al. (2014) pour la compression des contraintes table. En plus de la fréquence des motifs, nous exploiterons leur couverture dans le FP-Tree afin de faciliter la création des tables fragmentées. Pour décider si un motif fréquent est nécessaire à la compression, nous proposons d'exploiter la métrique de "taux de compression" au lieu du "calcul de gain". Cela permet d'obtenir une compression plus élevée et une meilleure résolution du problème. Les résultats expérimentaux sont encourageants.

1 Introduction

Un problème de satisfaction de contraintes (CSP) peut être défini comme un ensemble de variables avec des domaines finis de valeurs possibles et un ensemble de contraintes définies sur des sous-ensembles de variables. Les contraintes définies en extension (contraintes table) sont largement utilisées (bases de données, problèmes de configuration, etc.). Une liste de tuples des valeurs autorisées ou non-autorisées est donnée. La taille des contraintes table peut être très importante, ce qui peut constituer un obstacle au processus de résolution. De nombreux travaux ont été proposés dans la littérature pour la compression des contraintes table dans le but de faciliter la résolution du CSP correspondant. Nous pouvons citer Sliced tables Gharbi et al. (2014), micro-structure based compression Jabbour et al. (2015), etc.

Dans ce papier, nous proposons une amélioration de l'étape de compression de la méthode proposée par Gharbi et al. (2014) (appelée dans ce papier FPTCM) et qui exploite l'arbre de préfixe FP-Tree pour énumérer les motifs fréquents nécessaires à la compression des contraintes table. Notre première amélioration consiste à construire le FP-Tree autrement. Au lieu de considérer uniquement la fréquence des motifs dans la table des fréquences et dans le FP-Tree, nous exploitons aussi la notion de couverture d'un motif qui représente l'ensemble des indices des tuples où un motif apparaît. Cela permettra de réduire le temps de construction de la table compressée. La deuxième amélioration consiste à calculer le taux de compression estimé qu'on peut obtenir en compressant les tuples dont les indices apparaissent dans la couverture d'un