

Extraire les motifs minimaux efficacement et en profondeur

Arnaud Soulet*, François Rioult**

*Université François-Rabelais de Tours, LI EA 6300,
Campus de Blois, 41000 Blois
arnaud.soulet@univ-tours.fr

**Université de Caen Basse-Normandie, GREYC UMR 6072,
Campus II Côte de Nacre, 14000 Caen
francois.rioult@unicaen.fr

Résumé. Les représentations condensées ont fait l'objet de nombreux travaux depuis 15 ans. Tandis que les motifs maximaux des classes d'équivalence ont reçu beaucoup d'attention, les motifs minimaux sont restés dans l'ombre notamment à cause de la difficulté de leur extraction. Dans ce papier, nous présentons un cadre générique concernant l'extraction de motifs minimaux en introduisant la notion de système minimisable d'ensembles. Il permet de considérer des langages variés comme les motifs ensemblistes ou les chaînes de caractères, mais aussi différentes métriques dont la fréquence. Ensuite, pour n'importe quel système minimisable d'ensembles, nous introduisons un test de minimalité rapide permettant d'extraire en profondeur les motifs minimaux. Nous démontrons que l'algorithme proposé est *polynomial-delay* et *polynomial-space*. Des expérimentations sur les benchmarks traditionnels complètent notre étude.

1 Introduction

La minimalité est un concept essentiel de l'extraction de motifs. Pour une fonction f et un langage \mathcal{L} , un motif X est minimal si sa valeur pour f est distincte de celle de chacun de ses sous-ensembles. La collection de tous les motifs minimaux constitue une représentation condensée de \mathcal{L} adéquate à f : il est possible de retrouver $f(Y)$ pour n'importe quel motif $Y \in \mathcal{L}$. Typiquement, l'ensemble des itemsets libres (Boulicaut et al., 2000) est une représentation condensée de tous les itemsets (dans ce cas, f et \mathcal{L} sont respectivement la fréquence et le langage des itemsets). Bien sûr, il est souvent plus efficace d'extraire les motifs minimaux plutôt que la totalité des motifs. En plus, les motifs minimaux ont de nombreuses applications utiles notamment en extraction de connaissances : la production de bases de règles d'association, la construction de classifieurs ou la génération des traverses minimales. La minimalité a été étudiée dans le cas de différentes fonctions (comme la fréquence (Calders et al., 2004) ou les fonctions condensables (Soulet et Crémilleux, 2008)) et avec des langages variés (dont les itemsets (Boulicaut et al., 2000) et les séquences (Lo et al., 2008)). Bien que la minimalité ait des avantages évidents (Li et al., 2006), elle reste peu explorée en comparaison de la maximalité (i.e., les motifs fermés). En particulier, à notre connaissance, il n'existe pas de cadre suffisamment général comme celui proposé par Arimura et Uno (2009) pour les maximaux.