

# Motifs corrélés rares : caractérisation et nouvelles représentations concises exactes

Souad Bouasker, Tarek Hamrouni, Sadok Ben Yahia

Département des Sciences de l'Informatique, Faculté des Sciences de Tunis.  
{tarek.hamrouni, sadok.benyahia}@fst.rnu.tn

**Résumé.** Dans la littérature, les travaux se sont principalement focalisés sur l'extraction des motifs fréquents. Toutefois, récemment, la fouille des motifs rares s'est avérée intéressante puisque ces motifs permettent de véhiculer des connaissances concernant des événements rares, inattendus. Ils ont ainsi prouvé leur grande utilité dans plusieurs domaines d'application. Cependant, un constat important associé à l'extraction des motifs rares est d'une part leur nombre très élevé et d'autre part la qualité faible de plusieurs motifs extraits. Ces derniers peuvent en effet ne pas présenter des corrélations fortes entre les items les constituant. Afin de pallier ces inconvénients, nous proposons dans cet article d'intégrer la mesure de corrélation *bond* afin d'extraire seulement l'ensemble des motifs rares vérifiant cette mesure. Une caractérisation de l'ensemble résultant, des motifs corrélés rares, est alors réalisée en se basant sur l'étude des contraintes de nature différentes induite par la rareté et la corrélation. En outre, en se basant sur les classes d'équivalence associées à un opérateur de fermeture dédié à la mesure *bond*, nous proposons des représentations concises exactes des motifs corrélés rares.

## 1 Introduction et motivations

L'extraction des règles d'association est une technique très répandue dans la fouille de données et répond aux besoins des experts dans plusieurs domaines d'application. Plusieurs travaux se sont ainsi focalisés sur la dérivation des règles d'association à partir des motifs fréquents. Toutefois, l'utilisation de ces motifs ne constitue pas une solution intéressante pour certaines applications, telles que la détection d'intrusions, la détection des fraudes, l'audit des risques, l'identification des valeurs extrêmes dans les bases de données, l'analyse des données criminelles, l'analyse du désordre génétique à partir des données biologiques, l'analyse des maladies rares à partir des données médicales, l'analyse des données d'apprentissage en ligne, etc. (Booker, 2009; He et Xu, 2005; Koh et Rountree, 2010; Mahmood et al., 2010; Manning et al., 2008; Romero et al., 2010; Szathmary et al., 2010). En effet, dans de telles situations, un comportement fréquent peut être sans valeur ajoutée pour l'utilisateur final. Par contre, les événements peu fréquents sont les plus intéressants parce qu'ils indiquent qu'un événement inattendu, une exception par exemple (Taniar et al., 2008), est survenue. Une étude doit alors continuer afin de déterminer les causes possibles de ce changement peu commun du comportement normal. La fouille des motifs rares s'est alors avérée d'une réelle valeur ajoutée (Koh