

Une Approche d'Extraction de Motifs Graduels (Fermés) Fréquents Sous Contrainte de la Temporalité

Jerry Lonlac*,** Yannick Miras** Aude Beauger**
Marie Pailloux* Jean-Luc Peiry** Engelbert Mephu Nguifo*

*CNRS, UMR 6158, LIMOS, Université Clermont Auvergne, F-63173 Aubière, France
{lonlac, pailloux, mephu}@isima.fr

**CNRS, UMR 6042, GEOLAB, Université Clermont Auvergne, F-63000 Clermont-Ferrand
{yannick.miras, aude.beauger, jean-luc.peiry}@univ-bpclermont.fr

Résumé. La fouille de motifs graduels a pour but la découverte de co-variations fréquentes entre attributs numériques dans une base de données. Plusieurs algorithmes d'extraction automatique de tels motifs ont été proposés. La principale différence entre ces algorithmes réside dans la sémantique de variation considérée. Dans certains domaines d'application, on trouve des bases de données dont les objets sont munis d'une relation d'ordre temporel. Ainsi, du fait de leur sémantique de variation, les algorithmes de la littérature sont inadaptés pour de telles données. Dans ce contexte, nous proposons une approche de fouille de motifs graduels sous contrainte d'ordre temporel, qui réduit le nombre de motifs générés. Une étude expérimentale sur des bases de données paléoécologiques permet d'apprendre les groupements d'indicateurs qui modélisent l'évolution de la biodiversité. Les connaissances apportées par ces groupements montre l'intérêt de notre approche pour le domaine environnemental.

1 Introduction

Les motifs graduels qui capturent les corrélations d'ordre de la forme "plus/moins X, plus/moins Y" jouent un rôle important dans plusieurs applications du monde réel où le volume de données numériques à gérer est important, c'est le cas de données biologiques, ou de données médicales. Les algorithmes de fouille de données sont le plus souvent utilisés pour extraire automatiquement de tels motifs (Berzal et al., 2007; Masseglia et al., 2008; Di-Jorio et al., 2008, 2009; Laurent et al., 2009; Do et al., 2015).

En effet, Berzal et al. (2007) utilisent pour la première fois les méthodes de fouille de données à travers une adaptation de l'algorithme *Apriori* pour extraire les motifs graduels et évaluent le support de ces motifs en considérant tous les couples d'objets possibles. Dans Masseglia et al. (2008), les auteurs introduisent les motifs séquentiels graduels pour rendre compte de la force de modification (accélération). Deux autres méthodes d'extraction sont proposées dans Di-Jorio et al. (2008) et Di-Jorio et al. (2009), la différence entre elles étant liée au mode de calcul du support. En effet, dans Di-Jorio et al. (2008), étant donné un motif graduel, les auteurs proposent une heuristique permettant d'éliminer les objets qui empêchent le maximum