

Fouille de Motifs Graduels Fermés Fréquents Sous Contrainte de la Temporalité

Jerry Lonlac^{*,**} Benjamin Negrevergne^{**} Yannick Miras^{***} Aude Beauger^{***}
Engelbert Mephu Nguifo^{*}

^{*}CNRS, UMR 6158, LIMOS, Université Clermont Auvergne, F-63173 Aubière, France

^{**}LAMSADE, CNRS UMR 7243, Université Paris Dauphine
{benjamin.negrevergne}@dauphine.fr

^{***}CNRS, UMR 6042, GEOLAB, Université Clermont Auvergne,
F-63000 Clermont-Ferrand

{jerry.lonlac_konlac, engelbert.mephu_nguifo, yannick.miras, aude.beauger}@uca.fr

1 Introduction

La fouille de motifs graduels a pour but la découverte de co-variations fréquentes de la forme "plus/moins X, plus/moins Y" 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 paléocéologiques 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.

2 Fouille de motifs graduels sous contrainte temporelle

Soit une base de données numériques Δ contenant un ensemble d'objets $\mathcal{D} = \{d_1, \dots, d_n\}$ décrit par un ensemble d'attributs $\mathcal{I} = \{i_1, \dots, i_m\}$. Nous dénotons par $d_j[i_k]$ la valeur de l'attribut i_k sur l'objet d_j . Un item graduel est défini sous la forme i^* , où i est un attribut de \mathcal{I} et $*$ $\in \{\geq, \leq\}$ est un opérateur de comparaison. Un itemset (motif) graduel $s = (i_1^{*1}, \dots, i_k^{*k})$ est un ensemble non vide d'items graduels. Une séquence d'objets $\langle d_1, \dots, d_s \rangle$ respecte s si $\forall p \in [1, s-1], \forall l \in [1, k],$ nous avons $d_p[i_l] *_{l} d_{p+1}[i_l]$.

Le calcul du support d'un motif graduel dans une base de données Δ revient à mesurer à quel point le motif est présent dans Δ . Dans ce travail, nous considérons la sémantique de variation qui définit le support d'un motif graduel s comme la longueur de la plus longue séquence d'objets respectant le motif s .