

Extraction complète efficace de chemins pondérés dans un a-DAG

Nazha Selmaoui-Folcher*, Frédéric Flouvat*
Chengcheng Mu*, Jérémy Sanhes*, Jean-François Boulicaut**

*PPME - Université de la Nouvelle Calédonie
BP R4, 98851, Nouméa, Nouvelle Calédonie
prenom.nom@univ-nc.nc,

**LIRIS - INSA de LYON

Bâtiment Blaise Pascal F-69621 Villeurbanne cedex, France
jean-francois.boulicaut@insa-lyon.fr

Résumé. Un nouveau domaine de motifs appelé chemins pondérés condensés a été introduit en 2013 lors de la conférence IJCAI. Le contexte de fouille est alors un graphe acyclique orienté (DAG) dont les sommets sont étiquetés par des attributs. Nous avons travaillé à une implémentation efficace de ce type de motifs et nous montrons que l'algorithme proposé était juste mais incomplet. Nous établissons ce résultat d'incomplétude et nous l'expliquons avant de trouver une solution pour réaliser une extraction complète. Nous avons ensuite développé des structures complémentaires pour calculer efficacement tous les chemins pondérés condensés. L'algorithme est amélioré en performance de plusieurs ordres de magnitude sur des jeux de données artificiels et nous l'appliquons à des données réelles pour motiver qualitativement l'usage des chemins pondérés.

1 Introduction

Avec les avancées technologiques en terme d'acquisition des données scientifiques (images satellitaires, capteurs, etc.), les scientifiques s'intéressent de plus en plus à des applications importantes en terme de surveillance et suivi de l'environnement. Les données collectées sont généralement hétérogènes, multiéchelles, spatiales et temporelles (série temporelle d'images satellites, aériennes, modèles numériques de terrain, nature du sol ...) et sont destinées à comprendre et prédire des phénomènes résultant de processus complexes et d'origine pluridisciplinaire (données climatiques, géologiques, ...). L'explosion de cette information spatiale, temporelle et des systèmes d'informations géographiques nécessitent l'investissement dans des méthodes d'extraction de connaissances et nous nous intéressons à celles qui reposent sur la détection de motifs locaux comme, par exemple, la découverte de motifs séquentiels (Agrawal et Srikant, 1995; Mannila et al., 1997; Masseglia et al., 1998) ou de motifs plus complexes comme des sous-graphes Inokuchi et al. (2000) ou des sous-arbres Zaki (2002). Nos besoins concernent l'étude spatiale et temporelle des évolutions d'objets et de leurs interactions. Les objets peuvent être caractérisés par plusieurs attributs et leurs évolutions que