

PGP-mc : extraction parallèle efficace de motifs graduels

Anne Laurent*, Benjamin Negrevergne***, Nicolas Sicard**, Alexandre Termier***

*LIRMM - UM2- CNRS UMR 5506 - 161 rue Ada - 34392 Montpellier Cedex 5

laurent@lirmm.fr

<http://www.lirmm.fr>

**LRIE - EFREI - 30-32 av. de la République, 94 800 Villejuif

nicolas.sicard@efrei.fr

<http://www.efrei.fr>

***LIG - UJF-CNRS UMR 5217 - 681 rue de la Passerelle, B.P. 72, 38402 Saint Martin d'Hères

{benjamin.Negrevergne@imag.fr,Alexandre.Termier@imag.fr}

<http://www.liglab.fr>

Résumé. Initialement utilisés pour les systèmes de commande, les règles et motifs graduels (de la forme “plus une personne est âgée, plus son salaire est élevé”) trouvent de très nombreuses applications, par exemple dans les domaines de la biologie, des données en flots (e.g. issues de réseaux de capteurs), *etc.* Très récemment, des algorithmes ont été proposés pour extraire automatiquement de tels motifs. Cependant, même si certains d’entre eux ont permis des gains de performance importants, les algorithmes restent coûteux et ne permettent pas de traiter efficacement les bases de données réelles souvent très volumineuses (en nombre de lignes et/ou nombre d’attributs). Nous proposons donc dans cet article une méthode originale de recherche de ces motifs utilisant la *multi-threading* pour exploiter au mieux les multiples coeurs présents dans la plupart des ordinateurs et serveurs actuels. L’efficacité de cette approche est validée par une étude expérimentale.

1 Introduction

En fouille de données, la recherche de motifs fréquents est un sujet de recherche très actif. Initialement étudiés dans le cas de données transactionnelles, les algorithmes ont vite été étendus aux structures de données complexes (séquences, arbres, graphes *etc.*) Han et Kamber (2006). De manière générale, tous les algorithmes doivent explorer un espace de recherche très important. Ils sont donc très coûteux en temps de calcul, la complexité des calculs augmentant avec celle des structures de données à traiter. Une grande partie des travaux a été menée pour mettre au point des algorithmes de plus en plus efficaces permettant de fouiller de très grandes bases de données aux structures potentiellement complexes.

Récemment, un nouveau type de règles et motifs a été introduit : les *motifs graduels* (ou *itemssets graduels*). Ce problème vise à découvrir, à partir de bases de données numériques, des motifs du type “plus un individu est âgé, plus son salaire est élevé” et trouve de très nombreuses applications pour les bases de données numériques comme par exemple les données