

L'analyse relationnelle de concepts pour la fouille de données temporelles – Application à l'étude de données hydroécologiques

Cristina Nica*, Agnès Braud*, Xavier Dolques*
Marianne Huchard**, Florence Le Ber*

*ICube, Université de Strasbourg, CNRS, ENGEES
prenom.nom@engees.unistra.fr, agnes.braud@unistra.fr
<http://icube-bfo.unistra.fr>

**LIRMM, Université de Montpellier, CNRS
huchard@lirmm.fr
<https://www.lirmm.fr>

Résumé. Cet article présente une méthode d'exploration de données temporelles, fondée sur l'analyse relationnelle de concepts (ARC) et appliquée à des données séquentielles construites à partir d'échantillons physico-chimiques et biologiques prélevés dans des cours d'eau. Notre but est de mettre au jour des sous-séquences pertinentes et hiérarchisées, associant les deux types de paramètres. Pour faciliter la lecture, ces sous-séquences sont représentées sous la forme de motifs partiellement ordonnés (po-motifs). Le processus de fouille de données se décompose en plusieurs étapes : construction d'un modèle temporel *ad hoc* et mise en œuvre de l'ARC ; extraction des sous-séquences synthétisées sous la forme de po-motifs ; sélection des po-motifs intéressants grâce à une mesure exploitant la distribution des extensions de concepts. Le processus a été testé sur un jeu de données réelles et évalué quantitativement et qualitativement.

1 Introduction

Les données temporelles ou spatio-temporelles, de plus en plus nombreuses et largement utilisées, ont conduit au développement d'un ensemble de méthodes, permettant l'extraction de motifs ordonnés fréquents, et relevant de la « fouille de données séquentielles » (Agrawal et Srikant, 1995). Considérons par exemple une base de données de maladies d'où sont extraits deux motifs séquentiels fréquents ((Migraine)(Fièvre)) et ((Gorge-enflammée)(Fièvre)). Ces deux motifs mettent en évidence les symptômes qui précèdent la fièvre, mais leur coexistence n'est pas prise en compte dans les approches classiques. Pour surmonter cette limitation, Casas-Garriga (2005) a proposé d'utiliser des motifs partiellement ordonnés (po-motifs). Les po-motifs ont pour intérêt d'être compacts, de contenir la même information que les ensembles de motifs qu'ils représentent, et d'être faciles à interpréter grâce à leur représentation sous forme de graphes orientés acycliques. Les motifs partiellement ordonnés ont déjà été étudiés selon différentes méthodes, telles que Orderspan (Fabrègue et al., 2015) et Frecpo