

Analyse Relationnelle de Concepts pour l'exploration de données relationnelles

Xavier Dolques*, Florence Le Ber*, Marianne Huchard**, Clémentine Nebut**

*ICUBE, Université de Strasbourg/ENGEES, CNRS, Strasbourg,
prénom.nom@engees.unistra.fr

**LIRMM, Université de Montpellier 2 et CNRS, Montpellier,
prénom.nom@lirmm.fr

Résumé. L'Analyse Relationnelle de Concepts (ARC) est une extension de l'Analyse Formelle de Concepts (AFC), une méthode de classification non supervisée d'objets sous forme de treillis de concepts. L'ARC supporte en plus la gestion de relations entre objets de différents contextes ce qui permet d'établir des liens entre les concepts des différents treillis. Cette particularité lui permet d'être plus intuitive à utiliser pour extraire des connaissances à partir de données relationnelles et de donner des résultats plus riches. Malheureusement lorsque les jeux de données présentent de nombreuses relations, les résultats obtenus sont difficilement exploitables et des problèmes de passages à l'échelle se posent. Nous proposons dans cet article une adaptation possible de l'ARC pour explorer les relations de manière supervisée pour augmenter la pertinence des résultats obtenus et réduire le temps de calcul. Nous prenons pour exemple des données hydrobiologiques ayant trait à la qualité des milieux aquatiques.

1 Introduction

L'Analyse Formelle de Concepts (Ganter et Wille, 1999), notée de manière abrégée AFC, est une méthode de classification automatique d'objets décrits par des attributs au travers d'une relation binaire. Le résultat d'une telle classification est un treillis de concepts (appelé aussi treillis de Galois (Barbut et Monjardet, 1970)) où chaque concept regroupe tous les objets ayant en commun un ensemble d'attributs. On peut naviguer dans le treillis de manière simple et intuitive, des concepts les plus spécifiques (les concepts regroupant beaucoup de caractéristiques partagées par peu d'objets) aux moins spécifiques (les concepts regroupant beaucoup d'objets partageant peu de caractéristiques).

L'AFC est exploitée dans différents domaines en tant que méthode d'extraction de connaissances et les différentes publications sur le sujet, notamment Carpineto et Romano (2004); Valtchev et al. (2004), ont permis d'en identifier les forces et les limites. Certaines de ces limites ont pu être contournées en utilisant différentes approches.

L'Analyse Relationnelle de Concepts (ARC) (Huchard et al., 2007) est une extension de l'AFC qui permet de prendre en compte non seulement les caractéristiques des objets, mais aussi les relations que les objets entretiennent entre eux. L'ARC consiste à appliquer itérativement un algorithme de l'AFC pour gérer les données relationnelles : les objets sont décrits