

# Découverte de proportions analogiques dans les bases de données : une première approche

William Correa Beltran,  
Hélène Jaudoin,  
Olivier Pivert

Université de Rennes 1 – Irisa  
Lannion, France

{William.Correa\_Beltran@irisa.fr, Helene.Jaudoin@irisa.fr, Olivier.Pivert@irisa.fr}

**Résumé.** Cet article présente un nouveau cadre pour la découverte de connaissances basé sur la notion de proportion analogique qui exprime l'égalité des rapports entre les attributs de deux paires d'éléments. Cette notion est développée dans le contexte des bases de données pour découvrir des parallèles dans les données. Dans un premier temps, nous donnons une définition formelle des proportions analogiques dans le cadre des bases de données relationnelles, puis nous étudions le problème de l'extraction des proportions analogiques. Nous montrons qu'il est possible de suivre une approche de clustering pour découvrir les classes d'équivalence de paires de  $n$ -uplets dans le même rapport de proportion analogique. Ce travail constitue une première étape vers l'extension des langages d'interrogation de base de données avec des requêtes « analogiques ».

## 1 Introduction

Les travaux de cette dernière décennie dans le domaine de la découverte de connaissances, comme ceux notamment de (Han et al. (2007)) et de (Tiwari et al. (2010)), témoignent du vif intérêt pour le problème de l'extraction des ensembles d'items fréquents, des motifs séquentiels, des motifs structurels (dans les données de type arbres, graphes ou treillis), et la recherche de méthodes efficaces pour extraire ces motifs. Dans cet article, nous nous intéressons à la notion de proportion analogique, essentiellement étudiée dans le domaine de l'intelligence artificielle, pour extraire de nouveaux types de motifs dans les bases de données. Les proportions analogiques relient quatre objets  $A$ ,  $B$ ,  $C$ ,  $D$  du même type dans une assertion de la forme «  $A$  est à  $B$  ce que  $C$  est à  $D$  ». Ils permettent d'exprimer l'identité (ou la proximité) des rapports existant entre deux paires d'éléments. Des exemples typiques de cette notion en langage naturel sont : « le veau est à la vache ce que le poulain est à la jument », « l'aurochs est au bœuf ce que le mammouth est à l'éléphant ». Ces relations permettent d'exprimer que ce qui distingue  $A$  de  $B$  est comparable à ce qui distingue  $C$  de  $D$ . Les exemples ci-dessus montrent la diversité (et la potentielle complexité) des sémantiques possibles