

Identification de concepts remarquables à partir de leur voisinage dans un treillis

Etienne Lehembre*, Jean-Luc Lamotte**, Christophe Couronne*, Bruno Cremilleux*, Jean-Luc Manguin*, Bertrand Cuissart*, Abelkader Ouali*, Albrecht Zimmermann*

*GREYC, CNRS UMR 6072, UNICAEN, Normandy Univ. Caen, France
prenom.nom@unicaen.fr,
<https://www.greyc.fr>

**CERMN, EA 4258 FR CNRS 3038 INC3M SF 4206 ICORE,
UNICAEN, Normandy Univ. Caen, France
prenom.nom@unicaen.fr
<http://cermn.unicaen.fr>

1 Introduction

Les mesures de contraste (Novak et al., 2009) évaluent la qualité d'un motif par rapport à l'ensemble des données sans tenir compte de similarité ou de proximité entre motifs. Pour deux motifs proches (par exemple, deux motifs ne se distinguant que par un ou deux items dans le cas des itemsets), un analyste s'attend ce que la différence entre les valeurs de la mesure d'intérêt entre ces deux motifs soit faible. Malheureusement, ces mesures ne prennent pas en compte les éventuelles variations locales de qualité. C'est pourquoi nous nous intéressons à l'influence du voisinage sur une mesure de contraste.

La théorie de l'ordre (Davey et Priestley, 2002) et l'*Analyse Formelle de Concepts* (AFC) qui en découle permettent de structurer les motifs extraits en treillis où chaque élément du treillis est un concept formel (Ganter et al., 2005). En regroupant les données associées à un concept, l'AFC permet de supprimer des informations redondantes et fait ressortir l'information commune au sous-ensemble de données.

2 Mesure de contraste exploitant les fratries

Notre mesure de contraste s'appuie sur la structure induite par une relation d'ordre partiel, ainsi que sur une mesure de qualité adaptée au problème posé sur les données. Les concepts frères étant structurellement proches, leurs qualités devraient se ressembler. Nous considérons donc comme remarquable un concept ayant une qualité différente des qualités de sa fratrie.

Soit C un concept formel, \mathcal{F}_C sa fratrie étant la couverture inférieure de la couverture supérieure de C , et f une *mesure de qualité* sur l'ensemble des concepts. On définit la *moyenne de la fratrie* $\mu(\mathcal{F}_C)$ par la moyenne des qualités de ses concepts et l'*écart-type de la fratrie* $\sigma(\mathcal{F}_C)$ par l'écart-type des qualités de ses concepts. Un concept est *remarquable* si sa qualité diffère