

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

de la qualité moyenne de sa fratrie. En conséquence, on considère l'écart à la moyenne comme mesure de contraste d'un concept C : $|f(C) - \mu(\mathcal{F}_C)|$. A partir de cette mesure du contraste, C est remarquable si son écart à la moyenne est important : $|f(C) - \mu(\mathcal{F}_C)| \geq \delta * \sigma(\mathcal{F}_C)$.

3 Conclusion

Cette première étude doit se prolonger par une étude expérimentale afin d'estimer pratiquement l'intérêt de la notion et de comprendre les différences avec d'autres notions de concepts remarquables comme les *Jumping Emerging Pattern* (JEPs) (Dong et Li, 1999) et (Kane et al., 2015). En considérant les JEPs, notre mesure de contraste permet de sélectionner les JEPs remarquables par rapport à leur fratrie : les JEPs entourés de concepts non JEPs ou les non JEPs entourés de JEPs.

Références

- Davey, B. A. et H. A. Priestley (2002). *Introduction to Lattices and Order, Second Edition*. Cambridge University Press.
- Dong, G. et J. Li (1999). Efficient mining of emerging patterns : Discovering trends and differences. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 43–52.
- Ganter, B., G. Stumme, et R. Wille (2005). *Formal concept analysis : foundations and applications*, Volume 3626. springer.
- Kane, B., B. Cuissart, et B. Cremilleux (2015). Minimal Jumping Emerging Patterns : Computation and Practical Assessment. In *19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2015)*, Ho Chi Minh City, Vietnam.
- Novak, P. K., N. Lavrac, et G. I. Webb (2009). Supervised descriptive rule discovery : A unifying survey of contrast set, emerging pattern and subgroup mining. *J. Mach. Learn. Res.* 10, 377–403.

Summary

The purpose of pattern mining is to help experts understand their data. Starting from a dataset, our method is based on the lattice its associated formal concepts in order to explore their neighborhoods and discover remarkable, unexpected and therefore interesting concepts. In this article, we propose a new contrast measure based on the set of extensions of a concept and its neighbors. Given a measure of the quality of a concept, the contrast measure quantifies the variations in quality with respect to its neighborhood. When this deviation is sufficiently high, the concept is considered remarkable. We illustrate the capacity of our method to select a few hundred remarkable concepts from tens of thousands of frequent concepts mined from a graph-structured data set. This selection allows an expert to then conduct a localized analysis of the dataset.