

Fouille Interactive de Motifs à Haute Utilité

Arnold Hien**, Samir Loudni**, Maxime Garfagni*,
Albrecht Zimmermann***, Nicolas Beldiceanu**

*IMT Atlantique, Brest, France

maxime.garfagni@imt-atlantique.net

** TASC – DAPI, IMT-Atlantique, LS2N – CNRS, Nantes, France
{arnold.hien, samir.loudni, nicolas.beldiceanu}@imt-atlantique.fr

***CNRS - UMR GREYC, Normandie Univ., UNICAEN, Caen, France
albrecht.zimmermann@unicaen.fr

Résumé. Ces dernières années, l'extraction de motifs à haute utilité (ou HUI, High Utility Itemset) a fait l'objet de nombreux travaux. La fouille des HUIs permet en effet la découverte de motifs dont les items présentent des pondérations traduisant leur importance dans les transactions qui les contiennent. Les HUIs sont alors des motifs présentant une utilité ou une importance certaine pour l'utilisateur. Dans les différentes approches de l'état de l'art, leur extraction nécessite de fournir en plus du jeu de données, un ensemble d'utilités associées aux items et aux transactions. Si l'existence du jeu de données est une évidence, celle des utilités peut éventuellement être problématique dans certaines situations. Dans cet article, nous proposons une méthode interactive d'extraction de motifs à haute utilité. Notre approche permet d'apprendre les utilités des items et d'améliorer itérativement l'ensemble des HUIs extraits.

1 Introduction

Ces dernières années, plusieurs travaux se sont intéressés à la fouille de HUIs (Yao et al., 2004) qui consiste à trouver des motifs dits à haute utilité dans les jeux de données. Cette utilité peut être mesurée en terme de risque, profit, poids, quantité ou autre en fonction des préférences de l'utilisateur. La fouille des HUIs peut alors être vue comme une généralisation de la fouille de motifs fréquents dans lequel les items de chaque transaction sont associés à un poids (utilité externe) représentant leur importance, ainsi qu'à une utilité interne représentant une caractéristique associée comme la quantité (Yao et al., 2004; Yao et Hamilton, 2006). Cette approche de fouille peut ainsi être utilisée dans les applications quotidiennes comme l'analyse des flux de clics (Chu et al., 2008), des achats (Li et al., 2008) et comportements des utilisateurs (Shie et al., 2013) ou dans la génétique (Zihayat et al., 2017).

Différentes méthodes plus ou moins efficaces ont été proposées pour l'extraction de HUIs (Liu et Qu, 2012; Fournier-Viger et al., 2015; Tseng et al., 2016; Qu et al., 2019; Duong et al., 2022; Diop, 2022), mais elles nécessitent toutes de définir au préalable les utilités des items au moment de la fouille. Cependant, déterminer l'utilité des items n'est pas trivial, car souvent définies de manière arbitraire, rendant leur exploitation difficile dans un contexte applicatif.