

Extraction d'itemsets graduels de haute utilité

Priscile Audrey Fongue Assondji*, Jerry Lonlac**, Norbert Tsopze*

*Département d'informatique, Université de Yaoundé 1, Cameroun
{fongueaudrey0,tsopze.norbert}@gmail.com

** IMT Nord Europe, IMT, Univ. Lille, Centre for Digital Systems, F-59000 Lille, France
jerry.lonlac@imt-nord-europe.fr

1 Contexte et problématique

Les itemsets/motifs graduels présentés sous la forme "plus/moins A, plus/moins B, ..." permettent d'exprimer des covariations entre les différents attributs qui décrivent les données. Ils ont fait l'objet de plusieurs études et de nombreux algorithmes (Lonlac et Nguifo, 2020) ont été proposés pour les extraire à partir des données quantitatives. Le nombre de motifs extraits est souvent élevé et de nombreuses mesures (support, saisonnalité, émergence) ont été proposées pour évaluer ces motifs. Par ailleurs, les motifs de haute utilité permettent d'exprimer d'autres intérêts de l'utilisateur dans la recherche des motifs à travers le concept d'utilité. Ces deux formes de motifs (graduel et de haute utilité) permettent d'exprimer d'une part des covariations et d'autre part l'intérêt de l'utilisateur. Ce papier aborde le problème d'extraction de motifs dits graduels de haute utilité permettant de présenter à l'utilisateur comment une covariation de certains items impactera sur son intérêt dans une base de données quantitative. Nous combinons la mesure d'intérêt utilité avec la gradualité pour extraire les motifs graduels de haute utilité. Pour ce faire, l'approche méthodologique consiste à transformer la base de données originale Δ en une nouvelle base Δ' qui stocke les écarts entre les valeurs du même attribut entre différentes transactions. L'algorithme EFIM (Zida et al., 2017) est modifié pour extraire des motifs graduels de haute utilité de la base Δ' .

2 Fouille d'itemsets graduels de haute utilité

Nous définissons un itemset graduel de haute utilité comme un itemset graduel ayant une utilité supérieure à un seuil fixé par l'utilisateur.

Le principe général de l'extraction d'itemsets graduels de haute utilité suit deux étapes : le codage de la base de données et l'extraction d'itemsets à partir de la base de données encodée.

1. **Codage de la base de données** : Il s'agit de transformer la base de données originale Δ en une base de données Δ' contenant les mêmes attributs que Δ mais dont les occurrences sont les différences entre celles de Δ . Pour une colonne c , les valeurs sont calculées de la manière suivante : $\Delta'_{k,j}^c = \Delta_j^c - \Delta_k^c$. Une valeur positive de $\Delta'_{k,j}$ représente une augmentation de la quantité lorsqu'on passe de k à j , tandis qu'une valeur négative traduit une diminution.