

# GPOID : Extraction de Motifs Graduels pour les Bases de Données Imprécises

Michael Chirmeni Boujike\*, Jerry Lonlac\*\*  
Norbert Tsopze\*, Engelbert Mephu Nguifo\*\*\*

\*Département d'informatique - Université de Yaoundé 1

Sorbonne University, IRD, UMMISCO, F-93143, Bondy, France

michael.chirmeni@facsciences-uy1.cm, norbert.tsopze@facsciences-uy1.cm

\*\* IMT Lille Douai, IMT, Univ. Lille, Center for Digital Systems, F-59000 Lille, France

jerry.lonlac@imt-lille-douai.fr

\*\*\* LIMOS CNRS UMR 6158 - Univ. Clermont-Auvergne, F-63000 Clermont-FD

engelbert.mephu\_nguifo@uca.fr

**Résumé.** Ces dernières années, les motifs graduels ont attiré l'attention de la communauté de la science des données et plusieurs algorithmes ont été conçus pour extraire ces motifs à partir de différents modèles de données. Sur certaines données, comme les données imprécises, l'un des biais dans les algorithmes traditionnels est le fait qu'ils définissent la gradualité comme une augmentation/diminution de valeur. Par conséquent, certains motifs extraits ne sont qu'un effet de bruit dans les données. Pour remédier à ce problème, nous proposons dans cet article, une méthode qui introduit dans le processus de fouille, un seuil graduel à partir duquel considérer une gradualité. Les expérimentations sur différentes bases de données montrent que notre proposition réduit les temps de calcul et le nombre de motifs générés en focalisant sur les motifs d'intérêt. De plus, elle extrait les motifs graduels dans certains cas où les approches traditionnelles échouent.

## 1 Introduction

Le raisonnement humain est le plus souvent basé sur les données imprécises ou incomplètes ; en effet, il est facile pour un humain de déterminer si une personne est de petite ou de grande taille sans pour autant connaître sa taille exacte ; ce qui n'est pas le cas pour un ordinateur car ce dernier traite les données exactes. Transmettre les facultés du raisonnement humain à un ordinateur a été initié par (Zadeh et al., 1996) dont le but était de faire traiter les données imprécises par un ordinateur. Ainsi, des travaux basés sur la logique floue (Zadeh et al., 1996) et la logique classique (Agrawal et Srikant, 1994) ont été réalisés pour l'extraction des connaissances dans les bases de données catégorielles mais rencontrent de nombreuses difficultés dans les bases de données numériques. Récemment, un nouveau type de motif a émergé (motifs graduels) pour l'extraction de connaissance dans ces bases de données numériques et plusieurs algorithmes efficaces furent proposés pour extraire automatiquement ces motifs à