

# Classification Non Supervisée pour Données Catégorielles

Pierre-Emmanuel JOUVE \*, Nicolas NICOLLOYANNIS \*

\*LABORATOIRE ERIC, Université Lumière - Lyon2

Bâtiment L, 5 av. Pierre Mendès-France

69 676 BRON cedex FRANCE

{pierre.jouve, nicolas.nicoloyannis}@eric.univ-lyon2.fr,

<http://eric.univ-lyon2.fr>

**Résumé.** La classification non supervisée (CNS) constitue l'une des problématiques centrales de l'Extraction de Connaissances à partir de Données (E.C.D.). Le cadre spécifique de la CNS pour données catégorielles a été l'objet de multiples travaux ces dernières années, les principaux challenges associés à ses recherches sont d'une part la définition de critères bien adaptés à ce cadre particulier et d'autre part la mise au point d'algorithme au coût calculatoire relativement faible. Le propos de cet article n'est pas de poursuivre dans ces directions de recherche mais plutôt de s'appuyer sur ces travaux afin de proposer une méthode efficace exhibant de nombreux avantages pour l'utilisateur et utilisable par un non spécialiste. Nous proposons et évaluons donc une méthode de CNS pour données catégorielles permettant la mise à jour relativement rapide d'une classification pertinente d'un ensemble d'objets, tout en facilitant la tâche de l'utilisateur : aucun paramètre obscur ni nombre final de classes à fixer, description compréhensible de la classification, possibilité d'intervention de l'utilisateur dans le processus de CNS...

## 1 Introduction

Nous présentons dans cet article une nouvelle méthode de classification non supervisée (CNS) pour données catégorielles dont le principal attrait réside dans la résolution de plusieurs problèmes auquel un utilisateur se trouve confronté en pratique (détermination du nombre de classes, compréhensibilité de la taxonomie construite...) tout en n'impliquant pas une perte de qualité dans les résultats fournis. L'ensemble des terminologies et formalismes que nous utiliserons afin d'introduire le processus de CNS pour données catégorielles proviennent de multiples références de la littérature que nous ne manquerons pas d'évoquer. La forme de cette présentation s'inspire quant à elle de [8] (ce choix est motivé par le souci de conserver une certaine uniformité avec certains travaux notoires du domaine). Afin de faciliter la présentation nous nous appuyons sur des exemples (notés EXEMPLE :) basés sur un jeu de données décrivant un ensemble de 3 votes de motions différentes par 9 nations lors de sessions à l'O.N.U. (voir Tableau 1 page suivante).