

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université M'hamed BOUGARA de BOUMERDES



Faculté des Sciences

Département d'Informatique

## MEMOIRE DE MAGISTER

**Spécialité : Système informatique et génie des logiciels**

**Option : Spécification de Logiciel et Traitement de l'Information**

**Ecole Doctorale**

**Présenté par :**

**Alouane Basma**

**Thème**

**Recherche de partitions floues optimales par segmentation floue pour  
la fouille de données quantitatives.**

Devant le jury de soutenance composé de:

Mr. MEZGHICHE Mohamed	Professeur	UMBB	Président.
Mr. DJOUADI Yassine	Maître de conférence	UMMTO	Rapporteur.
Mr. AHMED NACER Mohamed	Professeur	USTHB	Examineur.
Mr. BABA-ALI Ahmed Riadh	Maître de conférence	USTHB	Examineur.

Année Universitaire : 2007/2008

## Table de matière

Chapitre I : Introduction .....	I.
Introduction.....	1
Chapitre II : Extraction de connaissance et fouille de données	
II.1 Extraction de connaissances à partir des données.....	4
II.1.1 Définition générale.....	5
II.1.2 Les étapes d'un processus d'Extraction de Connaissances à partir des données.....	6
II.2 Fouille de données (Data Mining).....	11
II.2.1 Historique.....	11
II.2.2 Définition de la fouille de données.....	12
II.2.3 Les méthodes de fouille de données.....	14
II.3 Les règles d'association.....	17
II.3.1 Cadre informel.....	17
II.3.2 Cadre formel.....	18
II.3.3 La découverte des règles d'association.....	21
II.3.3.1 Extraction des itemsets fréquents.....	21
II.3.3.2 Génération des règles d'association.....	27
II.3.4 Les améliorations de l'algorithme Apriori.....	32
II.3.5 Réduction de l'ensemble de règles d'association.....	33
II.3.5.1 Approche orientée données.....	33
II.3.5.2 Approche orientée utilisateur.....	39
II.3.6 Domaine d'applications.....	40
II.3.7 Types des données considérées.....	41
II.4 Règles d'association quantitatives.....	41
II.4.1 Problème des règles d'association quantitatives.....	44
II.4.2 Constat.....	45
II.5 Conclusion.....	46
III.1 La théorie des sous ensembles flous.....	47
III.1.1 Ensemble classique et Ensemble flou.....	47
Ensemble classique.....	47
III.1.1.2 Ensemble flou.....	48
III.1.2 Fonction d'appartenance.....	50
Le type.....	51
Le noyau.....	51
La hauteur.....	51
Les coupes de niveau $\alpha$ .....	52
Le support.....	52
III.1.3 La cardinalité.....	53
III.1.3.1 La cardinalité scalaire $\sum$ -count.....	53
III.1.3.2 La cardinalité floue de Zadeh.....	53
III.1.3.3 Cardinalité relative d'un ensemble flou.....	53
III.1.4 Opérations sur les sous-ensembles flous.....	54
III.1.4.1 L'égalité.....	54
III.1.4.2 L'inclusion.....	54
III.1.4.3 L'union :.....	54
III.1.4.4 L'intersection.....	55
III.1.4.5 Le complément.....	56

III.1.4.6	Le produit cartésien.....	56
III.1.4.7	Normes et conormes triangulaires.....	57
III.1.5	Raisonnement à partir des ensembles flous.....	58
III.1.5.1	Variable linguistique .....	58
III.1.5.2	Proposition floues.....	60
III.1.6	Règles floues .....	61
III.2	Règles d'association floues .....	63
III.2.1	Approche ensembliste .....	63
III.2.1.1	Principe générale .....	63
III.2.1.2	Evaluation algébrique du support et de la confiance.....	64
III.2.1.3	Approche sémantique .....	70
III.2.2	Approche logique (Dubois et Prade).....	76
III.3	Constat.....	78
III.4	Conclusion.....	79
IV.1	Principe fondamental de la segmentation .....	80
IV.1.1	Différents domaines d'application de segmentation .....	82
IV.1.2	Processus de segmentation .....	82
IV.2	Méthodes de segmentation.....	83
IV.2.1	La segmentation hiérarchique .....	85
IV.2.1.1	Méthodes ascendantes ou agglomératives.....	85
IV.2.1.2	Méthodes descendantes .....	87
IV.2.2	La segmentation par partition.....	87
IV.2.2.1	Méthode basé sur la densité .....	88
IV.2.2.2	Méthode basée sur les grilles .....	89
IV.2.2.3	Méthodes basés sur la théorie des graphes.....	90
IV.2.2.4	Méthodes basés sur la minimisation d'une fonction objective .....	91
IV.3	Segmentation floue.....	93
IV.3.1	Algorithme de C-moyennes floues (CMF).....	93
IV.3.2	Avantages et inconvénients de l'algorithme (CMF) .....	98
IV.3.3	Les algorithmes dérivés de l'algorithme CMF.....	99
IV.3.3.1	Algorithme de Gustafson et kessel.....	99
IV.3.3.2	Algorithme de Gath et Geva (FMLE) .....	100
IV.4	Le nombre de groupes et les indices de validités .....	101
IV.4.1	Le nombre de groupes .....	101
IV.4.2	Les indices de validités dédiés à la segmentation floue .....	103
IV.5	Conclusion.....	107
V.1	Principe général.....	108
V.1.1	Adaptation de l'Algorithme CMF .....	110
V.1.2	Justification du choix de l'algorithme CMF .....	114
V.2	Détermination du nombre optimal de groupes.....	116
V.2.1	Proposition basée sur le support.....	116
V.2.2	Proposition utilisant l'indice de validité $V_{PC}$ .....	118
V.2.3	Proposition utilisant l'indice de validité $V_{FS}$ .....	118
V.3	Proposition pour la découverte des règles d'association floues.....	120
V.3.1	Sémantique des règles d'association floues .....	120
V.3.2	Proposition d'une mesure de cardinalité floue.....	121
V.3.3	Proposition d'une t-norme.....	121
V.3.4	Présentation détaillée de l'algorithme de découverte des règles d'association floues	121
V.3.5	Evaluation des règles d'association floues.....	125

V.4	Comparaison des ensembles flous .....	126
V.4.1	Mesure de ressemblance.....	127
V.4.2	L'agrégation des mesures de ressemblances.....	128
	Si on veut donnée une valeur général de degré de ressemblance $S$ définit sur $\Omega$ , qui satisfait les propriétés des mesure de ressemblance, [111] [112] [114] propose d'utiliser une t-norme comme opérateur d'agrégation.....	128
V.5	Conclusion.....	128
VI.1	Présentation des bases de données .....	130
VI.1.1	La base de données KDD'99.....	130
	Attributs d'une connexion dans la base de données KDD'99.....	130
	Attributs retenus pour le cas de notre étude .....	132
VI.1.2	La base de données adult <sup>1</sup> .....	133
VI.2	Résultats obtenus.....	135
VI.2.1	Le nombre de partitions trouvées par chaque méthode.....	141
➤	Base KDD'99 .....	141
VI.2.2	Génération des règles d'association floues .....	142
VI.2.2.1	Règles d'association générées à partir de KDD'99.....	143
VI.2.2.2	Règles d'association floues générées à partir de Adult.....	146
VI.3	Comparaison des règles d'association floues.....	148
VI.3.1	Construction du modèle de validation.....	148
VI.3.2	Principe de comparaison par rapport au modèle de validation .....	150
VI.3.3	Evaluation des règle générées à partir de la base KDD'99.....	151
VI.3.4	Evaluation des règles générée à partir de la base adult .....	156
VI.3.5	Interprétation des résultats .....	161
VI.4	Le temps d'exécution .....	162
VI.5	Conclusion.....	163

### Abstract

The original problem of research of association rules was to extract some correlations from binary data.

Noting that the data are often quantitative, the problem has been extended for quantitative attributes by partitioning attribute domain and consequently mapping the quantitative problem to a binary one. However, such mapping generates sharp boundary problem.

In order to avoid such problem, fuzzy sets have been considered. All existing approaches, assume that fuzzy sets are empirically given. For this purpose, an original approach is proposed in this work in order to generate automatically fuzzy partitions. Our approach is based on fuzzy clustering method. We also suggest how to find automatically the optimal number of the fuzzy sets by using validity indices.

**Key words:** Data mining, association rules, fuzzy sets, fuzzy association rules, fuzzy clustering, validity indices, fuzzy resemblance measure.

## Résumé

Le problème original de recherche de règles d'association consistait à extraire certaines corrélations à partir de données binaires.

Constatant que souvent les données sont quantitatives, le problème a été étendu. L'idée consiste à ramener le problème à un cas binaire. Cependant une telle transformation cause le problème de valeurs aux limites.

Afin de pallier à ce problème, les ensembles flous ont été proposés. Toutes les approches existantes dans la littérature considèrent que les ensembles flous sont donnés d'une manière empirique. A cette fin, nous proposons dans notre mémoire une approche originale qui permet de générer automatiquement les partitions floues. Nous proposons aussi deux méthodes pour trouver le nombre de partitions floues.

**Mots clés :** Fouille de données, Règles d'association, ensembles flous, Règles d'association floues, Segmentation floue, Indice de validité, Mesure de ressemblance floue.