

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 :

Amroun Karima

Thème

**Découverte des dépendances fonctionnelles floues dans des
modèles relationnels sous imprécision.**

Devant le jury de soutenance composé de:

Mr. MEZGHICHE Mohamed	Professeur	UMBB	Président.
Mr. DJOUADI Yassine	Maître de conférence	UMMTO	Rapporteur.
Mme. AISSANI-MOKHTARI Aicha	Professeur	USTHB	Examinatrice.
Mme. ALI MAZIGHI Zaïa	Professeur	USTHB	Examinatrice.

Année Universitaire : 2007/2008

Table des matières

Liste des définitions.....	i
Table des figures.....	iv
Liste des algorithmes.....	v

Chapitre I : Introduction

I.1- Motivations.....	1
I.2- Problématique.....	3
I.3- Contributions.....	4
I.4- Plan du mémoire.....	5

Chapitre II : Les bases de données relationnelles

II.1- Le modèle relationnel (principes et notations formelles).....	7
II.2- Cohérence des données.....	8
II.2.1- Décomposition relationnelle.....	8
II.2.2- Les dépendances fonctionnelles (DFs).....	11
II.2.2.1- Axiomes d'Armstrong.....	11
II.2.2.2- Propriétés des DFs et définitions.....	12
A) Equivalence d'ensembles de DFs.....	13
B) Couverture minimale.....	13
C) Notion de clé.....	13
II.2.3- Formes normales (1 ^{ère} , 2 ^{ème} et 3 ^{ème}).....	13
II.2.4- Quatrième forme normale et Dépendances multivaluée.....	14
II.2.4.1- Propriétés des DMVs.....	15
II.2.4.2- Règles d'inférence pour DF et DMV.....	15
II.2.5- Dépendances de jointure (DJ) et Cinquième forme normale.....	16
II.3- Conclusion.....	17

Chapitre III : Extraction de Connaissances, Data Mining et Inférence des Dépendances Fonctionnelles

III.1- Extraction de Connaissances à partir des Données.....	18
III.1.1- Définitions de l'extraction de connaissances à partir des données	19
III.1.2- Etapes du processus d'extraction de connaissances à partir des données (ECD).....	19
III.1.2.1- Consolidation	21
III.1.2.2- Sélection et pré-traitement.....	21
III.1.2.3- Fouille de données.....	21
III.1.2.4- Interprétation.....	22
III.2- Fouille de Données (Data Mining).....	22
III.2.1- Les tâches du Data Mining	23
III.2.1.1- La classification	23
III.2.1.2- L'estimation	23
III.2.1.3- La prédiction	24
III.2.1.4- Les associations	24
III.2.1.5- La segmentation.....	24
III.2.2- Les méthodes et techniques de Data Mining.....	24
III.3- Les règles d'association	25
III.3.1- Modèle général des règles d'association	26
III.3.2- Sémantique d'une règle d'association	27
III.3.3- Inférence des règles d'association	28
III.3.4- L'algorithme Apriori.....	29
Notations utilisées par Apriori	29
Génération des candidats pour Apriori avec la procédure Apriori-Gen	31
III.3.5- Critiques de la méthode des règles d'association.....	33
III.4- Etat de l'art sur la découverte de DFs	34
III.4.1- Intérêt de la découverte des dépendances fonctionnelles.....	35
III.4.2- Approches de découverte des DFs.....	36
III.4.2.1- Algorithme naïf.....	36
III.4.2.2- Algorithme Dep-miner	37
Phase de prétraitement de l'algorithme Dep-miner	37
Découverte des ensembles en accord de r	39
Dérivation des compléments des ensembles maximaux	41

Calcul des parties gauches des DFs.....	42
Inférence des DFs minimales non triviales.....	44
III.4.2.3- Algorithme TANE.....	45
Les partitions dans TANE	45
La procédure COMPUTE-DEPEDENCIES.....	47
La procédure PRUNE	48
La procédure GENERATE-NEXT-LEVEL	49
Apports de TANE	49
III.5- Conclusion	50

Chapitre IV : Théorie des ensembles flous, bases de données et dépendances fonctionnelles floues

IV.1- Théorie des ensembles flous.....	51
IV.1.1- Historique sur la théorie des ensembles flous	51
IV.1.2- Eléments de base de la théorie des ensembles flous.....	52
A)- Le noyau.....	54
B)- Le support	55
C)- La hauteur	55
Sous-ensemble flou normalisé	55
D)- Coupe de niveau α (α -coupe).....	55
Ensemble des éléments propres d'une coupe.....	56
IV.1.3- La cardinalité d'un ensemble flou.....	57
IV.1.4- Opérations ensemblistes floues.....	57
IV.1.4.1- Intersection et Union d'ensembles flous	57
Normes et co-normes triangulaires.....	58
IV.1.4.2- Le complément	59
IV.1.4.3- Le produit Cartésien.....	59
IV.1.4.4- Les relations floues et la composition de relations.....	60
Propriétés des relations floues	61
Relations floues particulières.....	61
IV.1.4.5- L'égalité	61
IV.1.4.6- L'inclusion	62
IV.1.5- Les implications floues	62

A)- S-implications	63
B)- R-implication	64
Quelques caractéristiques des implications floues.....	65
Sémantique associée aux implications floues	66
IV.2- La théorie des possibilités	67
A)- Incertitude	67
B)- Imprécision	67
IV.2.1- Mesures de possibilité d'événement usuel	68
IV.2.1.1- Propriétés des mesures de possibilités	69
IV.2.1.2- Distribution de possibilité	69
Représentation d'une distribution de possibilité par une α -coupe.....	70
IV.2.1.3- Cas extrêmes Ignorance totale et Certitude totale.....	70
IV.2.2- Mesure de nécessité d'un événement usuel	71
Propriétés des mesures de nécessité	72
IV.2.3- Possibilité et Probabilité.....	72
IV.2.4- Possibilités et ensemble flou	73
IV.2.5- Possibilité/ Nécessité d'événements flous.....	73
IV.2.5.1- Possibilité d'événements flous	73
IV.2.5.2- Nécessité d'événement flou	74
IV.2.5.3- Propriétés de mesures de possibilité/nécessité floues	74
IV.2.6- Intérêt de la théorie des possibilités	74
IV.3- Les principaux modèles de bases de données floues	75
IV.3.1- Modèle de Pons	76
IV.3.2- Modèle de Buckles-Petry	76
IV.3.3- Modèle d'Umano-Fukami	78
IV.3.4- Modèle de Prade-Testemale.....	79
IV.3.5- Modèle GEFRED de Médina et al.....	80
IV.3.6- Le Modèle Relationnel Possibiliste Généralisé.....	82
Sémantique de l'opérateur de restriction flou	83
a) Interprétation certaine (nécessaire) d'une base de données possibiliste	85
b) Interprétation possible d'une base de données possibiliste	85
IV.4- Les Dépendances Fonctionnelles Floues (DFFs)	88
IV.4.1- Principes des dépendances fonctionnelles floues	88
A)- Le principe de relation floue	88
B)- Le principe de valeurs imprécises.....	88

C)- Le principe d'égalité floue.....	89
D)- Le principe de quantificateur flou.....	89
IV.4.2- Les modèles de Dépendances Fonctionnelles Floues	90
A)- Approche de Kiss et al	90
B)- Approche de Raju et Mazumder	91
C)- Approche de Prade et Tastemale.....	91
D)- Approche de Bhuniya et Niyogi.....	92
E)- Approche de Cubero et al.....	92
F)- Approche de Chen et al.....	93
G)- Approche de Bosc et al	94
IV.5- Etat de l'art sur la découverte des DFFs	94
Algorithme de Wang	95
La procédure COMPUTE-DEPEDENCIES.....	96
Limites de l'approche de Wang.....	96
IV.6- Conclusion	97

Chapitre V : Contributions

V.1- Préparation des données cibles	99
V.1.1- Choix du modèle cible et fuzzification des données.....	99
V.1.2- Traitement des valeurs manquantes.....	101
V.2- Présentation de l'approche stratifiée.....	103
V.2.1- Proposition d'une nouvelle définition dépendance fonctionnelle floue.....	104
V.2.2- Axiomes d'Armstrong.....	105
V.2.3- Processus global de découverte des α -DFFs (algorithme général α -FFD-Extract).....	108
A)- Définitions formelles	108
B)- Principe général du processus de découverte des α -DFFs	110
C)- Algorithme α -FFD-Extract	111
E)- Validation des parties gauches de taille i de chaque attribut K	113
F)- Elagage de l'ensemble des parties gauches candidates de taille $i+1$	116
G)- Mise en forme des α -DFFs validées.....	116
H)- Comparaison de la complexité algorithmique de $\alpha_FFD_Extract$ et celle de l'algorithme de WANG	117
V.3- Exemple d'application	117

V.3.1- Traitement des données manquantes (Missing Values).....	118
Application de la transformation de DUBOIS et PRADE.....	118
Base de données après fuzzification :.....	121
a) Découverte des α-DFFs minimales non triviales à un niveau $\alpha = 0.3$:.....	122
b) Découverte des α-DFFs minimales non triviales à un niveau $\alpha = 1.0$:.....	125
c) Analyse des résultats obtenus dans les deux cas :.....	127
V.4- Conclusion.....	128

Conclusion et perspectives

Bilan et Contributions.....	129
Perspectives.....	131

Annexe 1 : Fermeture et treillis

A.1.1- Ensemble ordonné et relation de couverture.....	132
A.1.2- Treillis et fermeture.....	136

Annexe 2 : correction et complétude du système de règles d'inférence des α -FFDs

A.2.1- Preuve de correction :.....	139
A.2.2- Preuve de complétude.....	140

Bibliographie.....	141
---------------------------	------------