



N° d'ordre :.....

**UNIVERSITE DE M'SILA**  
**FACULTE DES MATHÉMATIQUES ET DE**  
**L'INFORMATIQUE**

Département d'Informatique

# Mémoire

Présenté pour l'obtention du diplôme de Magistère

Spécialité : STIC

Option : SIC

Par ARBOUCHE NASSIMA

## SUJET

Fouille de données biomédicales complexes  
Application à la segmentation d'images médicales

Soutenu publiquement le 24/06/2010 devant le jury composé de :

Mr. A.ZIDANI	Université de Batna(MC)	Président.
Mr. Abd elouaheb Moussaoui	UFAS(MC)	Rapporteur.
Mr. M.C.BABAHENINI	Université de Biskra(MC)	Examineur.
Mr. F.CHERIF	Université de Biskra(MC)	Examineur.
Mr. M.BOURAHLA	Université de M'sila(MC)	Examineur.

Promotion 2006 – 2007

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Fouille de Données</b>	<b>4</b>
1.1 Introduction . . . . .	4
1.2 Définition de la fouille de donnée ou «Data Mining» . . . . .	4
1.3 Entrepôt de données « Datawarehouse » . . . . .	5
1.4 Le DM dans le processus d'ECD . . . . .	5
1.4.1 La phase de préparation . . . . .	5
1.4.1.1 L'acquisition . . . . .	6
1.4.1.2 Prétraitement des données . . . . .	6
1.4.2 Phase Fouille de données . . . . .	7
1.4.3 Phase d'interprétation et d'évaluation . . . . .	7
1.5 Méthodes et techniques du data mining . . . . .	7
1.5.1 Les méthodes du DM . . . . .	8
1.5.1.1 Les modèles prédictifs . . . . .	9
1.5.1.2 Méthodes de structuration (segmentation) . . . . .	9
1.5.1.3 La description . . . . .	9
1.5.2 Les techniques utilisées . . . . .	10
1.5.2.1 Les techniques classiques . . . . .	10
1.5.2.2 Les techniques sophistiquées . . . . .	10
1.5.2.2.1 Algorithmes de segmentation . . . . .	10
1.5.2.2.2 Règles d'association . . . . .	11
1.5.2.2.3 Algorithmes de classification, estimation et prédiction	12
1.5.2.2.4 Métaheuristiques pour l'extraction de connaissances	15
1.6 Conclusion . . . . .	16
<b>2 Fouille de Données Complexes</b>	<b>18</b>
2.1 Introduction . . . . .	18
2.2 Données complexes . . . . .	18
2.3 Fouille de données complexes . . . . .	20
2.3.1 Préparation des données complexes . . . . .	22
2.3.1.1 Prétraitement des données complexes . . . . .	22
2.3.2 Une remise en cause des outils traditionnels . . . . .	23
2.3.3 Les travaux réalisés . . . . .	23
2.3.4 La description de quelques fouilles de données . . . . .	24
2.3.4.1 La fouille de données textuelles . . . . .	24
2.3.4.2 La fouille de séquences . . . . .	24

2.3.4.3	La fouille de données spatiales . . . . .	24
2.3.4.4	Web mining . . . . .	25
2.4	Fouille de données biomédicales complexes : . . . . .	25
2.4.1	Diversité des données biomédicales . . . . .	25
2.4.2	Caractéristiques et spécificités des données biomédicales . . . . .	26
2.4.3	Définition de la fouille de données biomédicales . . . . .	27
2.4.4	Les travaux réalisés . . . . .	28
2.5	Conclusion . . . . .	29
<b>3</b>	<b>Méthodes de Segmentation d'Images</b>	<b>30</b>
3.1	Introduction . . . . .	30
3.2	Segmentation d'images . . . . .	31
3.2.1	La segmentation de l'IRM cérébrale . . . . .	31
3.2.1.1	Spécificités de la segmentation de l'IRM cérébrale . . . . .	32
3.3	Les différentes approches de segmentation . . . . .	33
3.3.1	État de l'art du point de vue « Traitement d'images » . . . . .	34
3.3.1.1	Approches Région . . . . .	34
3.3.1.1.1	Le suillage . . . . .	34
3.3.1.1.2	La croissance de région . . . . .	35
3.3.1.1.3	Morphologie mathématique . . . . .	36
3.3.1.2	Approches contours . . . . .	37
3.3.1.2.1	Méthodes dérivatives . . . . .	37
3.3.1.2.2	Modèles déformables . . . . .	38
3.3.1.2.3	Espace-échelle . . . . .	39
3.3.2	État de l'art du point de vue «Reconnaissance des formes» . . . . .	39
3.3.2.1	Les méthodes de classification . . . . .	41
3.3.2.1.1	Classification supervisée . . . . .	41
3.3.2.1.2	Classification semi supervisée . . . . .	41
3.3.2.1.3	Classification non supervisée (Automatique) . . . . .	42
3.3.2.2	Méthodes supervisées . . . . .	42
3.3.2.2.1	Approches probabilistes . . . . .	42
3.3.2.3	Méthodes non supervisées (Automatiques) . . . . .	44
3.3.2.3.1	La famille des C-moyennes . . . . .	45
3.3.2.3.2	Les métaheuristiques pour l'extraction des connaissances . . . . .	50
3.3.2.3.3	Autres méthodes . . . . .	54
3.4	Conclusion . . . . .	54
<b>4</b>	<b>Fusion de Données</b>	<b>55</b>
4.1	Introduction . . . . .	55
4.2	Éléments théoriques . . . . .	56
4.2.1	Représentation de Connaissances ambiguës « imparfaites » . . . . .	56
4.2.1.1	Théorie des probabilités . . . . .	56
4.2.1.1.1	Cadre bayésien . . . . .	57
4.2.1.2	Théorie des ensembles flous . . . . .	57
4.2.1.3	Théorie des possibilités . . . . .	61
4.2.1.4	Mesure de possibilité . . . . .	61
4.2.1.4.1	Distribution de possibilité . . . . .	61

4.2.1.4.2	Mesure de nécessité . . . . .	62
4.2.1.5	Théorie des fonctions de croyance . . . . .	63
4.2.1.5.1	Fonction de masse . . . . .	63
4.2.1.5.2	Fonction de croyance . . . . .	64
4.2.1.5.3	Fonction de plausibilité . . . . .	64
4.2.1.5.4	Propriétés des fonctions de croyance . . . . .	64
4.2.1.5.5	Relation entre les trois théories . . . . .	65
4.2.1.6	Autres approches . . . . .	65
4.3	Fusion d'information . . . . .	65
4.3.1	Caractéristiques générales des données . . . . .	66
4.3.2	Système de fusion et type d'architecture . . . . .	67
4.3.3	Modèles de fusion . . . . .	67
4.3.4	Classification des types de fusion . . . . .	69
4.3.5	Opérateurs de fusion . . . . .	69
4.3.5.1	Classification des opérateurs de fusion . . . . .	70
4.3.5.2	Quelques propriétés . . . . .	71
4.3.6	Méthodes numériques de fusion . . . . .	71
4.3.6.1	Fusion en théorie des probabilités . . . . .	71
4.3.6.1.1	Etape de fusion . . . . .	71
4.3.6.1.2	Etape de décision . . . . .	71
4.3.6.2	Fusion floue et possibiliste . . . . .	72
4.3.6.2.1	Etape de fusion . . . . .	72
4.3.6.2.2	Etape de décision . . . . .	72
4.3.6.3	Fusion en théorie des croyances . . . . .	73
4.3.6.3.1	Etape de fusion . . . . .	73
4.3.6.3.2	Etape de décision . . . . .	73
4.4	Fusion des données médicales . . . . .	74
4.4.1	Les applications médicales de la fusion de données . . . . .	76
4.4.2	Fusion d'image médicale . . . . .	76
4.4.2.1	Introduction de l'information spatiale . . . . .	77
4.5	Conclusion . . . . .	78
<b>5</b>	<b>Contribution</b> . . . . .	<b>79</b>
5.1	Introduction . . . . .	79
5.2	Approche proposée . . . . .	80
5.2.1	Quelle est la fusion envisagée ? . . . . .	81
5.2.2	Les types de données à représenter ? . . . . .	82
5.2.2.1	Modélisation de la donnée issue du dossier médical . . . . .	82
5.2.2.1.1	Variables linguistiques . . . . .	83
5.2.2.2	Modélisation de l'image . . . . .	84
5.2.2.2.1	Méthode supervisée ou non ? . . . . .	85
5.2.2.2.2	Classification floue ou non floue ? . . . . .	85
5.2.2.2.3	l'algorithme utilisé ? . . . . .	85
5.2.2.3	Informations supplémentaires . . . . .	93
5.2.2.4	Comment envisage-t-on la fusion ? . . . . .	94
5.2.3	Niveau de fusion ? . . . . .	94

5.2.4	Cadre théorique de fusion ? . . . . .	94
5.2.5	Opérateur de fusion ? . . . . .	95
5.2.5.1	Opérateur proposé . . . . .	97
5.2.5.2	Pourquoi avoir choisi des pondérations dynamiques ? . . . . .	98
5.2.6	Critère de décision . . . . .	98
5.2.7	Architecture de fusion . . . . .	100
5.2.8	Présentation de l'algorithme général . . . . .	100
5.3	Conclusion . . . . .	101
<b>6</b>	<b>Résultats et validation</b>	<b>102</b>
6.1	Introduction . . . . .	102
6.2	Choix des données biomédicales . . . . .	102
6.3	Description des volumes traités . . . . .	104
6.3.1	Fantôme du cerveau . . . . .	104
6.3.2	Images simulées . . . . .	104
6.3.3	Images réelles . . . . .	104
6.4	Critères de validation . . . . .	105
6.5	Résultats de la classification . . . . .	106
6.5.1	Résultats de classification sur des données de synthèse . . . . .	106
6.5.2	Résultats de classification sur des données réelles . . . . .	109
6.5.3	Comparaison des résultats obtenus . . . . .	111
6.6	Conclusion . . . . .	115
	<b>Conclusion et perspectives</b>	<b>116</b>
	<b>Appendices</b>	
<b>A</b>	<b>Anatomie du cerveau</b>	<b>119</b>
A.1	Introduction . . . . .	119
A.2	Système nerveux central . . . . .	119
A.2.1	L'encéphale . . . . .	120
A.2.1.1	Tronc cérébral . . . . .	121
A.2.1.2	Le cervelet . . . . .	121
A.2.1.3	Le cerveau . . . . .	122
A.2.1.4	Les méninges . . . . .	123
A.2.1.4.1	Le liquide céphalo-rachidien . . . . .	123
A.2.1.4.2	La substance grise . . . . .	124
A.2.1.4.3	La matière blanche . . . . .	124
<b>B</b>	<b>Les images IRM</b>	<b>125</b>
B.1	Introduction . . . . .	125
B.2	L'imagerie par résonance magnétique (IRM) . . . . .	126
B.2.1	Principe de fonctionnement . . . . .	126
B.2.2	Notions élémentaires sur le magnétisme . . . . .	127
B.2.2.1	Champ magnétique et moment magnétique . . . . .	127
B.2.2.2	Moment magnétique et aimantation . . . . .	127
B.2.2.3	Propriétés magnétiques de la matière . . . . .	128

## Résumé

L'objectif de ce travail est d'élargir le champ d'utilisation des méthodes de fouille de données pour qu'elles soient adaptées aux spécificités et aux caractéristiques des données complexes comme les données médicales. Dans ce contexte, et appliqué aux données biomédicales, nous proposons une architecture de fusion dans un cadre possibiliste de données de type hétérogène (donnée numérique ou linguistique et une image de modalité IRM). La théorie des ensembles flous a été retenue pour la représentation de la donnée numérique issue du dossier médical ; quand à la représentation de l'image, une hybridation entre l'algorithme AntClust basé sur une population d'agents fourmis et l'algorithme flou FCM a été proposée.

Pour pallier aux défauts du FCM à savoir, l'initialisation et la sensibilité aux données aberrantes, AntClust servira de phase d'initialisation, découvrant ainsi automatiquement le nombre de classes, quand à la sensibilité aux données aberrantes, elle est traitée lors de la phase de fusion avec la donnée issue du dossier médical.

**Mots clés : Mots-clés : Fouille de données complexes, modalité IRM, Segmentation d'images, fusion de données, FCM, fourmis artificielles, antclust**

## Abstract

The aim of this work is to broaden the scope of use of methods of data mining to be adapted to the specificities and characteristics of complex data such as medical data. In this context, and applied to biomedical data, we propose a merger in a possibilistic two heterogeneous data types (numeric or linguistic data and an image modality MRI). The theory of fuzzy sets was chosen to represent the digital data from medical records, when displaying the image, Hybrids between AntClust algorithm based on a population of ant agents and fuzzy FCM algorithm was proposed. To overcome the shortcomings of FCM namely, initialization and sensitivity to outliers, AntClust serve as initialization phase, automatically discovering the number of classes, when the sensitivity to outliers, it is treated during the phase Merge with data from the medical record.)

**Keywords : complex data mining, MRI, clustering, data fusion, FCM, artificial ants, antclust**

## ملخص

الهدف من هذا العمل هو توسيع نطاق استخدام أساليب استخراج البيانات إلى أن تتكيف مع خصوصيات وسمات البيانات المعقدة مثل البيانات الطبية. في هذا السياق، تم تطبيقها على البيانات الطبية، اقترحنا دمج نوعين من البيانات غير المتجانسة (بيانات رقمية أو لغوية، و صورة الرنين المغناطيسيIRM) في نظرية «possilistic» .

تم اختيار نظرية مجموعات الضبابية لتمثيل البيانات الرقمية الآتية من السجلات الطبية، أما بالنسبة لصورة IRM، اقترحنا التنسيق بين الخوارزمي AntClust المستنبطة من (مستعمرات النمل الاصطناعي) وخوارزمي المتوسطات الضبابية FCM. للتغلب على أوجه القصور في FCM، ألا وهي التهيئة والحساسية للقيم المتطرفة، وضع AntClust بمثابة مرحلة التهيئة، واكتشاف عدد الطبقات تلقائيا، أما بالنسبة لحساسية القيم المتطرفة، فيتم التعامل معها خلال مرحلة الدمج مع بيانات السجل الطبي).

## الكلمات الرئيسية

التنقيب في البيانات المعقدة، صور الرنين المغناطيسي، التقسيم، دمج البيانات، وخوارزمي المتوسطات الضبابية FCM، النمل الاصطناعي، Antclust.