

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Ecole nationale Supérieure d'Informatique (ESI)

## MEMOIRE

Présenté pour l'obtention du diplôme de Magister

En : Informatique

Spécialité : Informatique mobile

Par : BAGHLI Raouida

# Metaheuristiques pour la Classification non Supervisée

### **Mémoire soutenu devant le jury :**

M. ADMANE Lotfi : Maitre de Conférences à l'ESI (Président du jury)

M. BADACHE : Professeur Cerist

Mme SI TAYEB Fatima : Maitre de Conférences à l'ESI

Mme BENATCHBA Karima : Professeur à l'ESI (Directrice du mémoire)

---

## *Abstract*

The world today requires the handling of a large mass of information. This significant amount of information should not be lost and should be properly analyzed and exploited to satisfy the needs of competitiveness and performance increasingly growing in the field of business. This is where was born the need to resort to data mining, a semi-automatic process of data analysis. We can distinguish two main tasks of this process, predictive tasks and descriptive ones. The first attempts to predict future values as to the second aims to describe and exploit the current values. Unsupervised classification or clustering is a descriptive task considered as the most important one. It consists on a process that gathers objects in homogeneous classes where the most similar objects are in the same classes and the most distant ones are in different classes. This problem is NP-hard and the use of conventional methods to solve it is time consuming. Different metaheuristics have been used to overcome this problem. Among these metaheuristics, we are particularly interested in biomimetic and evolutionary methods. This work has two goals: First a state of the art on the application of these metaheuristics on clustering. Second, a comparative study on the behavior of three methods, namely evolutionary algorithms, artificial immune systems and bees' method, on different types of benchmarks.

**Key words:** Data-mining, Clustering, Metaheuristics, Biomimetic methods, Evolutionary methods, Artificial immune systems, Bees methods.

---

## Résumé

Le monde d'aujourd'hui nécessite la manipulation d'une grande masse d'information. Cette quantité importante d'information ne doit pas être perdue et doit être correctement analysée et exploitée afin de satisfaire les besoins de compétitivité et de performance de plus en plus croissant dans le domaine de l'entreprise. Et c'est de là qu'est née la nécessité de recourir au Data mining, un processus semi-automatique d'analyse de données. Nous pouvons distinguer deux tâches principales de ce processus, les tâches prédictives et les tâches descriptives. Les premières visent à prédire des valeurs futures quant aux secondes, elles ont pour objectif de décrire et d'exploiter au mieux les valeurs actuelles. La classification non supervisée ou clustering fait partie de ces dernières, c'est-à-dire les tâches descriptives, et peut être considérée comme l'une des plus importantes. C'est un processus de regroupement d'objets en classes homogènes où les données les plus similaires possibles sont dans la même classe et les données les plus éloignées possibles appartiennent à des classes différentes. Ce problème étant NP-difficile, l'utilisation de méthodes classiques pour le résoudre devient impossible. Différentes métaheuristiques sont alors apparues pour pallier au problème de clustering. Parmi ces métaheuristiques, nous nous sommes particulièrement intéressés aux méthodes biomimétiques et évolutionnaires. Notre travail consiste, dans une première étape, à faire un état de l'art sur leur application au clustering. Dans une seconde étape, il consiste à étudier et comparer le comportement de trois méthodes, à savoir les algorithmes évolutionnaires, les systèmes immunitaires artificiels et la méthode des abeilles, sur différents types de benchmarks et montrer l'influence du codage sur leur performance.

**Mots- clés :** Data-mining, Clustering, Métaheuristiques, Méthodes biomimétiques, Algorithmes évolutionnaires, Systèmes immunitaires artificiels, et la Méthode des abeilles.

## Table des matières

<i>Abstract</i> .....	I
<i>Résumé</i> .....	II
<i>Liste des figures</i> .....	VI
<i>Liste des algorithmes</i> .....	VII
INTRODUCTION GENERALE .....	1
Chapitre 1 : DATA MINING ET KDD .....	4
1.1. Introduction .....	5
1.2. Définition du data mining .....	5
1.3. KDD (Knowledge Discovery in Databases) .....	7
1.3.1. Modèles académiques .....	7
1.3.2. Modèles industriels .....	9
1.3.3. Modèles hybrides .....	10
1.3.4. Comparaison entre les différentes catégories de modèle .....	10
1.4. Tâches du Data mining .....	12
1.4.1. Tâches descriptives .....	12
1.4.2. Tâches prédictives .....	13
1.5. Méthodes du data mining .....	13
1.6. Domaines d'application du data mining .....	14
1.7. Conclusion .....	14
Chapitre 2 : CLUSTERING .....	15
2.1. Introduction .....	16
2.2. Notions de base .....	16
2.2.1. Distance .....	16
2.2.2. Distance sur $R^m$ .....	17
2.3. Classification hiérarchique .....	17
2.3.1. BIRCH .....	19
2.3.2. CURE .....	19
2.4. Classification par partitionnement .....	20
2.4.1. K-Means .....	20
2.4.2. K-MEDOIDS .....	21
2.5. Classification basée sur la densité .....	23
2.5.1. DBSCAN .....	25
2.5.2. DENCLUE .....	26
2.6. Classification basée sur les grilles .....	27

2.6.1. STING .....	28
2.6.2. CLIQUE .....	28
2.7. Conclusion.....	28
Chapitre 3 : METHODES BIOMIMETIQUES POUR LE CLUSTERING .....	30
3.1. Introduction .....	31
3.2. Recuit Simulé .....	31
3.3. Recherche Tabou .....	33
3.4. Fourmis artificielles.....	35
3.4.1. ANTCLASS.....	35
3.4.2. ANTCLUST .....	39
3.4.3. AntTree.....	43
3.5. Classification par optimisation basée sur la biogéographie .....	47
3.5.1. Principe de la BBO .....	47
3.5.2. ALGORITHME DE LA BBO .....	48
3.6. Algorithmes d'abeilles basés sur la recherche de nourriture.....	49
3.6.1. ALGORITHME DE COLONIE ARTIFICIELLE D'ABEILLES (ABC) .....	49
3.6.2. ALGORITHME DES ABEILLES (BA) .....	50
3.7. Conclusion.....	51
Chapitre 4 : METHODES EVOLUTIVES POUR LE CLUSTERING .....	52
4.1. Introduction .....	53
4.2. Algorithmes évolutionnaires .....	53
4.2.1. SCHEMA TYPE D'UN ALGORITHME EVOLUTIONNAIRE .....	53
4.2.2. PRESSION DE SELECTION ET DERIVE GENETIQUE .....	54
4.2.3. POLITIQUES DE SELECTION POUR LE CROISEMENT.....	54
4.2.4. METHODE DE CROISEMENT .....	57
4.2.5. MUTATION .....	58
4.2.6. SELECTION POUR LE REMPLACEMENT.....	59
4.2.7. ALGORITHMES GENETIQUES.....	60
4.2.8. CONVERGENCE DES ALGORITHMES EVOLUTIONNAIRES .....	60
4.2.9. ALGORITHMES EVOLUTIONNAIRES POUR LA CLASSIFICATION NON SUPERVISEE .....	60
4.3. Les systèmes immunitaires artificiels .....	64
4.3.1. SYSTEME IMMUNITAIRES HUMAIN.....	64
4.3.2. INTERET POUR LES SYSTEMES IMMUNITAIRES .....	64
4.3.3. SYSTEMES IMMUNITAIRES ARTIFICIELS POUR L'OPTIMISATION .....	65
4.3.4. RESEAUX IMMUNITAIRES ET INTERACTION ENTRE LYMPHOCYTES.....	67
4.3.5. SYSTEMES IMMUNITAIRES ARTIFICIELS POUR LE CLUSTERING .....	68

---

4.3.6. CODAGE PAR ANTICORPS .....	70
4.4. Conclusion.....	71
Chapitre 5 : TESTS ET RESULTATS.....	72
5.1. Introduction.....	73
5.2. Présentation des benchmarks .....	73
5.2.1. FONCTION DE FIABILITE .....	73
5.2.2. BENCHMARKS UTILISES.....	74
5.3. Algorithmes Evolutionnaires .....	75
5.3.1. CODAGE PAR CENTRE .....	76
5.3.2. CODAGE PAR PARTITIONNEMENT .....	77
5.3.3. COMPARAISON ENTRE LES CODAGES.....	77
5.4. Colonies d'abeilles (ABC) .....	79
5.5. Algorithme évolutionnaire VS ABC .....	80
5.5.1. CODAGE PAR CENTRE .....	80
5.5.2. CODAGE PAR PARTITIONNEMENT .....	81
5.6. Système immunitaire artificiel.....	82
5.7. Conclusion.....	84
Conclusion générale .....	85
Bibliographie .....	87