

**DATA MINING WITH BIO-INSPIRED**

**OPTIMIZATION ALGORITHMS**

by

**Tang Rui**

**Master of Science in Software Engineering**

2013



Department of Computer and Information Science  
Faculty of Science and Technology  
University of Macau

**DATA MINING WITH  
BIO-INSPIRED  
OPTIMIZATION ALGORITHMS**

by  
Tang Rui

A thesis submitted in partial fulfillment of the  
requirements for the degree of

Master of Science in Software Engineering

Department of Computer and Information Science

Faculty of Science and Technology

University of Macau

2013

Approved by \_\_\_\_\_

\_\_\_\_\_  
Supervisor

Date \_\_\_\_\_

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Macau, I agree that the Library and the Faculty of Science and Technology shall make its copies freely available for inspection. However, reproduction of this thesis for any purposes or by any means shall not be allowed without my written permission. Authorization is sought by contacting the author at

Address: ROOM NLG201, Choi Kai Yau Building (N), University of Macau, Av. Padre Tomás Pereira Taipa, Macau, China

Telephone: 00853-63540659

E-mail: tr77229591@126.com



Signature \_\_\_\_\_

Date \_\_\_\_\_

University of Macau

## **ABSTRACT**

### **DATA MINING WITH BIO-INSPIRED OPTIMIZATION ALGORITHMS**

by

Tang Rui

Thesis Supervisor: Dr. Simon Fong

Thesis co-Supervisor: Dr. Yang Xin-She

This thesis aims to study bio-inspired algorithms and to design several bio-optimization algorithms, bio-clustering methods and feature selection using optimization methods. We proposed a new bio-inspired optimization algorithm named Wolf Search Algorithm (WSA). WSA is a new type of swarm intelligence techniques and it is able to find solution to optimization of the continuous functions. In the proposed approach, the search agent is capable of doing global exploration, local exploitation and jump out of local capabilities.

Clustering using bio-optimization algorithms is a hybrid method, which is not just only an algorithm. It is a generic method, as we can choose any optimization algorithm and applies it into clustering to optimize clusters centroids. The objective function in our proposed method is the configuration of centroids. The optimization algorithm optimizes the objective function result to return the best centroids to the process of clustering. Then clusters are constructed around these best centroids.

Feature selection using optimization algorithms also is a generic method that integrates with optimization algorithms for optimizing the candidate feature sets in order to choose the optimal subset of feature from the whole set.

The experimental results show that our proposed algorithms and methods are very competitive when compared to other approaches.

**Keywords:** bio-inspired, Wolf Search Algorithm, WSA, clustering using optimization algorithm, feature selection using optimization algorithm

## TABLE OF CONTENTS

List of Figures .....	iii
List of Tables .....	v
List of Abbreviations .....	vii
Chapter 1: INTRODUCTION.....	1
1.1 Background.....	1
1.2 Research Objective and Structure.....	2
Chapter 2: LITERATURE SURVEY .....	3
2.1 Optimization .....	3
2.2 Clustering Algorithm.....	5
2.3 Feature Selection.....	7
Chapter 3: RESEARCH WOLF SEARCH ALGORITHM.....	9
3.1 Behavior Of Wolf .....	9
3.2 Algorithm Rules.....	10
3.3 Description Of Wolf Search Algorithm.....	13
3.3.1 Initialization.....	13
3.3.2 Merging.....	13
3.3.3 Preying.....	15
3.4 Validation and Comparison .....	20
3.4.1 Validation.....	20
3.4.2 Comparison.....	27
3.5 Choice of Parameters .....	34
3.5.1 Visual Distance .....	34
3.5.2 Population .....	36
3.5.3 Minus Steps.....	37
3.5.4 Escape threshold .....	39

3.5.5 The Difference in each Step between WSA and ACO .....	40
3.6 Discussion .....	43
 Chapter 4: CLUSTERING USING BIO-OPTIMIZATION .....	45
4.1 Objective Function and Fitness for Clustering using Bio-Optimization .....	45
4.2 Clustering Algorithm using Optimization. ....	47
4.2.1 Construct of Normal K-means .....	48
4.2.2 Constructs of K-means Combine with Optimization.....	51
4.2.2.1 Clustering with Cuckoo Algorithm.....	51
4.2.2.2 Clustering with Bat Algorithm.....	52
4.2.3 The Logic Flows for Clustering using Optimization .....	53
4.3 Experiments .....	59
4.3.1 Popular Data Sets.....	59
4.3.2 Popular Data Sets.....	66
4.4 Discussion .....	72
 Chapter 5: FEATURE SELECTION USING OPTIMIZATION ALGORITHM.....	73
5.1. Feature Selection.....	73
5.2. The Proposed Method.....	74
5.2.1 Search Agent (Wolf/Bat/PSO) Representation.....	74
5.2.2 Objective Function.....	76
5.2.3 Dimension Change.....	76
5.2.4 Algorithm.....	77
5.2.5 Experiment.....	78
 Chapter 6: CONCLUSION AND FUTURE WORK.....	88
 Bibliography .....	90

## LIST OF FIGURES

<i>Number</i>	<i>Page</i>
Figure 1: Brownian Motion Movement .....	10
Figure 2: Procedure of Preying Passively .....	16
Figure 3: A Snapshot of WSA in Action .....	18
Figure 4: For All Wolves Flow Chart .....	19
Figure 5: For One Wolf Flow Chart .....	19
Figure 6: The Visualization of The Schaffer's F6 Function\ .....	21
Figure 7: The Wolves' Locations at Initial Step Which Are Distributed at Random.....	21
Figure 8: The Wolves' Locations at Final Step, Merged at The Global Minimum.....	22
Figure 9: The Overview of Rastrigin's Function .....	23
Figure 10: The Wolves' Locations at Initial Step.....	23
Figure 11: The Wolves' Locations at Final Step .....	24
Figure 12: The Overview of Rosenbrock Function .....	25
Figure 13: The Initial Phase of Wolves .....	26
Figure 14: The Final Phase of WSA.....	26
Figure 15: WSA Average Fitness for Each Generation.....	41
Figure 16: ACO Average Fitness for Each Generation .....	41
Figure 17: Fitness Difference between WSA and ACO .....	42
Figure 18: Initial Centroid and Result .....	47
Figure 19: Flowchart of an Ordinary K-means Clustering Algorithm .....	49
Figure 20: A Generalized Flowchart Consists of Three Types of Constructs .....	50
Figure 21: Flow Chart of C-WSA Algorithm.....	54
Figure 22: Flow Chart of C-bat Algorithm.....	55
Figure 23: Flow Chart of C-cuckoo Algorithm .....	56
Figure 24: Flow Chart of C-Firefly Algorithm.....	57
Figure 25: Flow Chart of C-ant Algorithm.....	58
Figure 26: Clustering results by different algorithms for mouse dataset.....	62
Figure 27: Visualization of Clustering Results.....	69

Figure 28: Flow Chart of Feature Selection using Bio-Optimization.....78





## LIST OF TABLES

<i>Number</i>	<i>Page</i>
Table 1: The WSA Pseudo Code .....	12
Table 2: Comparison of Convergence cycles Using Fixed Tolerance Threshold .....	30
Table 3: Comparison of optimization Fitness Using Fixed Number of Iterative cycles.....	32
Table 4: Comparison of visual distance.....	35
Table 5: Comparison of visual populations .....	36
Table 6: Comparison of previous steps.....	38
Table 7: Comparison of previous steps.....	39
Table 8: Comparison between WSA and ACO .....	40
Table 9: Datasets.....	59
Table 10: Parameter setting of C-firefly and C-cuckoo.....	60
Table 11: Comparison of Algorithms w.r.t. Objective Function Value and CPU Time for Iris Dataset .....	63
Table 12: Comparison of Algorithms w.r.t. Objective Function Value and CPU Time for Wine Dataset.....	64
Table 13: Comparison of Algorithms w.r.t. Objective Function Value and CPU Time for Libra Dataset.....	64
Table 14: Comparison of Algorithms w.r.t. Objective Function Value and CPU Time for Haberman' survival Dataset .....	65
Table 15: Comparison of Algorithms w.r.t. Objective Function Value and CPU Time for synthetic Dataset .....	65
Table 16: Comparison of Algorithms w.r.t. Objective Function Value and CPU Time for Mouse Dataset.....	66
Table 17: Parameters used in the clustering algorithms .....	67
Table 18: Web based data description .....	68
Table 19: Comparison in PB dataset.....	70
Table 20: Comparison in IU dataset .....	71

Table 21: Comparison in IP dataset .....	71
Table 22: Comparison in SB dataset.....	71
Table 23: Test Dataset .....	78
Table 24: Parameter setting .....	79
Table 25: Feature selection using WSA (Naive Naïve Bayes) .....	80
Table 26: Feature selection using PSO (Naive Naïve Bayes).....	80
Table 27: Feature selection using Bat (Naive Naïve Bayes) .....	81
Table 28: Feature selection using WSA (BP network).....	81
Table 29: Feature selection using PSO (BP network) .....	82
Table 30: Feature selection using BA (BP network) .....	82
Table 31: Feature selection using WSA (Decision Tree) .....	83
Table 32: Feature selection using PSO (Decision Tree).....	83
Table 33: Feature selection using BA (Decision Tree).....	84
Table 34: Feature selection using Cfsubsetval(Decision Tree).....	84
Table 35: Feature selection using Cfsubsetval(Pattern Network).....	84
Table 36: Feature selection using Cfsubsetval(Naïve Bayes) .....	85
Table 37: Feature selection using WSA (Pattern Network).....	85
Table 38: Feature selection using WSA (Decision Tree) .....	85
Table 39: Feature selection using WSA (Naïve Bayes).....	86
Table 40: Feature selection using Bat (Decision Tree).....	86
Table 41: Feature selection using Bat (Naïve Bayes).....	86
Table 42: Feature selection using BAT (Pattern Network) .....	86
Table 43: Feature selection using PSO (Decision Tree).....	87
Table 44: Feature selection using PSO (Naïve Bayes) .....	87
Table 45: Feature selection using PSO (Pattern Network) .....	87

## LIST OF ABBREVIATIONS

<b>GA</b>	Genetic Algorithm
<b>PSO</b>	Particle Swarm Optimization
<b>SA</b>	Simulated Annealing
<b>TS</b>	Tabu Search
<b>EP</b>	Evolutionary Programming
<b>ES</b>	Evolutionary Strategy
<b>ACO</b>	Ant Colony Optimization
<b>FA</b>	Firefly Algorithm
<b>CA</b>	Cuckoo Algorithm
<b>BA</b>	Bat Algorithm
<b>WSA</b>	Wolf Search Algorithm
<b>C-WSA</b>	Clustering using WSA
<b>C-Bat</b>	Clustering using BA
<b>C-Cuckoo</b>	Clustering using CA
<b>C-Ant</b>	Clustering using ACO

## ACKNOWLEDGMENTS

I wish to express my deep appreciation to my supervisor, Dr. Simon Fong, and my co-supervisor, Dr. Yang Xin-She, for your patience, support, guidance and encouragement in all the time of research and writing of this thesis.

I also thank my friends and my family who have provided constant support and encouragement to me through this process.

