Ain Shams University
College of Women for
Arts, Science and Education
Physics Department

CONTRIBUTIONS TO DATA MINING USING ANT COLONY OPTIMIZATION ALGORITHMS

By

SAMAH REFAT ABDELHAI

Thesis Submitted
For
The Degree of Doctor of Science
(Physics – Computer Science and its Applications)

Under Supervision of

Dr.* Ashraf H. Abdelwahab, *Professor
Computers and Systems Dept.
Electronics Research Institute

Dr. Amira Z.Dakroury, *Professor*
University College of Women for Arts, Science and Education
Ain Shams University

Dr. Hesham A. Hefny, *Associate Prof.*

Computers and Information Dept.
Institute of Statistical Research and Studies
Cairo University

Dr. Mohammed E. El-Telbany, *Associate Prof.*

Computers and Systems Dept.
Electronics Research Institute
(2010)

Ain Shams University
College of Women for
Arts, Science and Education
Physics Department

CONTRIBUTIONS TO DATA MINING USING ANT COLONY OPTIMIZATION ALGORITHMS

By

SAMAH REFAT ABDELHAI

Thesis Submitted
For
The Degree of Doctor of Science
(Physics – Computer Science and its Applications)

Approved by the Examining Committee:

Dr. Mohammed Z. Abd-Elmgid	**Advisor.**
Dr. Samia A. Mashali	**Advisor.**
Dr. Ashraf H. Abdelwahab, *Professor*	**Thesis main advisor.**
Dr. Amira Z.Dakroury, *Professor*	**Thesis main advisor.**
Dr. Hesham A. Hefny, Associate Proffeso	**Member.**
Dr. Mohammed E. El-Telbany, Associate Proffesor	**Member.**

University College of Women for Arts, Science and Education
Ain Shams University

2010

جامعة عين شمس
كلية البنات للآداب و العلوم والتربية
قسم الطبيعة

مساهمات في التنقيب في البيانات باستخدام خوارزميات مستعمرات النمل

مقدمة من الباحثة:

سماح رفعت عبد الحي

للحصول على درجة الدكتوراه فى فلسفة العلوم

تخصص (تطبيقات الحاسب الألي في علم الفيزياء)

المشرفون

د/أشرف حسن عبد الوهاب

أستاذ الذكاء الاصطناعي – معهد بحوث الالكترونيات– المركز القومي للبحوث .

د/ أميرة ذكي دكروري

أستاذ فيزياء الجوامد – قسم الفيزياء–كلية البنات– جامعة عين شمس .

د/ هشام أحمد حفني

أستاذ مساعد الذكاء الحسابي – معهد الدراسات و البحوث الأحصائية– جامعة القاهرة .

د/ محمد السعيد التلباني

أستاذ مساعد الذكاء الاصطناعي – معهد بحوث الالكترونيات–المركز القومي للبحوث .

٢٠10

ABSTRACT

In the past few decades, cluster analysis has played a central role in a variety of fields ranging from engineering (machine learning, artificial intelligence, pattern recognition, mechanical engineering, electrical engineering), computer sciences (web mining, spatial database analysis, textual document collection, image segmentation), to bioinformatics and medical sciences (genetics, biology, microbiology, paleontology, psychiatry, pathology). Clustering is often used as a tool for preliminary and descriptive data analysis and for unsupervised classification. Its main purpose is to identify homogeneous groups by finding similarities between objects regarding their characterizing attributes.

Swarm Intelligence (SI) is an innovative distributed intelligent paradigm for solving optimization problems that originally took its inspiration from the biological examples by swarming, flocking and herding phenomena in vertebrates. Particle Swarm Optimization (PSO) incorporates swarming behaviors observed in flocks of birds, schools of fish, or swarms of bees. Ant Colony Optimization (ACO) deals with artificial systems that are inspired from the

foraging behavior of real ants, which are used to solve discrete optimization problems.

Data mining and *Swarm intelligence* may seem that they do not have many properties in common. However, recent studies suggest that they can be used together for several real world data mining problems especially when other methods would be too expensive or difficult to implement.

This thesis deals with the application of swarm intelligence methodologies in data mining especially for data clustering. In our thesis PSO_Ant_Clustering algorithm represents the proposed clustering technique in which Ant Based Clustering is integrated by Particle Swarm Optimization for data clustering. The PSO_Ant_Clustering algorithm allows the parameters controlling the clustering sensitivity to be adapted as the PSO search algorithm proceeds. The proposed hybrid model consists of a set of particles, each of which has an Ant_clustering parameters set. The ants use these parameters to guide its search for favorable dropping/picking locations. The main goal of this is to create clusters of higher quality.

We have examined different types of problems. Using the traditional clustering techniques (such as K-Means and

Expectation Maximization), traditional Ant Based Clustering and the proposed technique PSO_Ant_Clustering algorithm are applied to the problems at hand. It has been shown that the PSO_Ant_Clustering algorithm has superior results than that of all such other techniques.

LIST OF PUBLICATIONS

1. El-Telbany M., Refat S., Abdelwhab A., Hefny H., and El Dakroury A., "*From Evolution to Swarm Intelligence: Data Clustering Survey*", International Journal of Intelligent Computing and Information Sciences, vol. 7, No. 2, pp. 219-236, July 2007.
2. El-Telbany M., Refat S., Sultan H. Aljahdali , "*Natural_Inspired Data Clustering: A Hybrization Between Ant Clustering and Particle Swarm Optimization*", in ISCA 18th International Conference on Software Engineering and Data Engineering, Las Vegas , USA, June 2009
3. Refat S., El-Telbany M., Abdelwhab A., El Dakroury A., and Hefny H., "Integrated Ant Based Clustering with Particle Swarm Optimization for Data Clustering", International Journal of Intelligent Computing and Information Sciences, vol 10., No 1. , pp. 291-302 , Jan 2010.

ACKNOWLEDGEMENT

My parents over the years showed an amount of patience that is hard for me to fathom. They consistently gave me the freedom to chart my own path and stood by me, and without their support none of this would have been possible. There are many people to thank for contributing to this dissertation and to my success in graduate study. Foremost is professor Ashraf H. Abdulwahab and Professor Amira Z.EL-Dakruory, who have been my principles guide throughout my graduate career. I have benefited enormously from their scientific curiosity, attention to details, and research standard. They have considerably changed my view to the world. Dr. Hesham A. Hefny and Dr. Mohammed E. El-Telbany have provided invaluable technical assistance over the years, improving my writing style, and gives me the moral support needed to complete this dissertation. Finally, I cannot adequately express my gratitude to any one, who has helped me and encouraging me in my research work.

TABLE OF CONTENTS

LIST OF FIGURES