



شبكة المعلومات الجامعية

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



شبكة المعلومات الجامعية
@ ASUNET



شبكة المعلومات الجامعية التوثيق الالكتروني والميكروفيلم



شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الالكتروني والميكرو فيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأفلام قد أعدت دون أية تغيرات



يجب أن

تحفظ هذه الأفلام بعيدا عن الغبار

في درجة حرارة من ١٥-٢٥ مئوية ورطوبة نسبية من ٢٠-٤٠%

To be Kept away from Dust in Dry Cool place of
15-25- c and relative humidity 20-40%

بعض الوثائق الأصلية تالفة

بالرسالة صفحات لم ترد بالاصل

٢٩٧١

Cairo University
Institute of Statistical Studies and Research
Department of Computer and Information Sciences

Fuzzy Rough Set Approaches For Data Clustering

A thesis submitted in partial fulfillment of requirements for the degree of
master of computer science

Submitted by

Usama Mokhtar Hassen Hasaneen

Supervised by

prof. Dr. Atef Abd El Moneim

Mathematical Statistics Department
Institute of statistical studies
and Research

Dr . Hesham Ahmed Hefny

Associate professor, Department
of Computer and Information
Sciences Cairo University
Institute of statistical studies and Research
Cairo University

April 2009

Cairo University
Institute of Statistical Studies and Research
Department of Computer and Information Sciences
Cairo University
Egypt

Approval sheet

Fuzzy Rough Set Approaches For
Data Clustering

A Thesis submitted to the Department of
Computer and Information Sciences
In partial fulfillment of the
Requirements for the degree of
Master of Computer Science

Submitted by

Usama Mokhtar Hassen Hasaneen

Approved by the Examining Committee

Prof. Dr. Atef Mohamed Abd El Moneim



Prof. Dr. Ashraf Hassen Abd El wahab



Dr. Abo El Ela Hassaneen El Oteefy



Dr. Hesham Ahmed Hefny



Statement

I certify that this work has not been accepted in substance for any academic degree and is not being concurrently submitted in candidature for any other degree.

Any portions of this thesis for which I am indebted to other sources are mention and explicit references are given.

Student name: Usama Mokhtar Hassan.

Signature:

Abstract

One of the most important methods in analysis of large data sets is clustering. Cluster analysis is a technique for classifying data, i.e., to divide a given dataset into a set of classes or clusters. The goal is to divide the dataset in such a way that two cases from the same cluster are as similar as possible and two cases from different clusters are as dissimilar as possible.

In this thesis, we propose modification to one of modern clustering which is Fuzzy Rough C-Means (FRCM). The algorithm named at Merging Fuzzy Rough C-Means (MFRCM) Clustering Algorithm. This algorithm can achieve high accuracy result when it compare with other trademark clustering algorithm such FCM and FRCM it self. Aim of This algorithm to cluster data with a minimum number of errors in clustering process.

Acknowledgement

First of all, I would like to express my appreciation to Dr. Hesham Ahmed Hefny who has been supporting my studies, guiding my thesis, and constant involvement in guiding me towards my goal, also for his a lot of kindness, patience and effectual support during the work. My research could not have been finished reasonably without insightful advice from him.

My thanks, and appreciations to Prof. Dr. Atef Abd Elmoneim must be stated for his support and cooperation

Finally, I am deeply grateful to my family, especially my father, my mother and my wife for their love, support and patience during the writing of the thesis.

TABLE OF CONTENTS

Contents	Page
1. Abstract	II
2. Acknowledgments	III
3. List of chapters	V
4. List of figures	X
5. List of tables	XII
6. List of Algorithms	XIII

List of Chapters:

Chapter 1: Introduction.....	1
Chapter 2: Overview of data clustering.....	5
2.1 Introduction.....	5
2.2 Concept of data mining.....	6
2.3 Concept of data clustering	7
2.4 Clustering vs. classification	8
2.5 Cluster definition and analysis.....	9
2.5.1 Cluster definitions	9
2.5.2 Cluster analysis	9
2.5.3 Objectives of cluster analysis.....	10
2.6 Components of clustering task.....	10
2.7 Challenges of clustering task	11
2.8 Applications of data clustering	14
2.9 similarities measurement	15
2.10 Clustering techniques.....	18
2.10.1 Hierarchal clustering.....	18
2.10.2 Partitioning clustering.....	20
2.10.2.1 K-Medoids methods	21
2.10.2.2 K-Means methods.....	22
2.10.2.3 Probabilistic clustering	22
2.10.3 Density Based clustering.....	23
2.10.4 Grid Based methods	23
2.10.5 Other methods	24
2.11 Summary.....	24

Chapter 3: Fuzzy and Rough sets :An overview.....	25
3.1 Introduction.....	25
3.2 Fuzzy sets overview.....	25
3.2.1 Classical Sets.....	25
3.2.2 Classical Sets operation	27
3.2.3 Fuzzy sets.....	28
3.2.4 Membership function.....	30
3.2.5 Operation on fuzzy sets.....	32
3.2.6 Properties of the operations in Fuzzy sets.	34
3.2.7 T-norm and S-norm operators.....	35
3.2.8 Important notes.	35
3.2.9 Features on membership function.....	36
3.2.10 Linguistic variables.	36
3.2.11 Linguistic hedges.	38
3.2.12 Relations.....	39
3.2.12.1 Crisp relations and compositions	39
3.2.12.2 Fuzzy relations and compositions.....	40
3.2.12.2.1 ..Operation on fuzzy relations	41
3.2.12.2.2The composition on fuzzy relations	44
3.2.13 Applications of Fuzzy Systems.....	43
3.2.14 Advantages of Fuzzy Systems.	44
3.2.15 Problems and limitations of fuzzy systems.	44
3.3 Rough set: An overview.....	45
3.3.1 Information systems.....	46
3.3.2 Decision systems.....	47
3.3.3 Approximation Spaces.....	48

3.3.3.1	Equivalence relation.....	48
3.3.3.2	Properties of equivalence relation.....	48
3.3.3.3	Equivalence class	48
3.3.3.4	Universe partitioning.....	49
3.3.3.4	Concepts	49
3.3.4	Set approximation	49
3.3.4.1	Inconsistent information tables	50
3.3.4.2	Lower and upper approximations	50
3.3.5	The operations in rough sets	53
3.3.6	Properties of the operations in rough sets.....	53
3.3.7	Classification of approximation space.....	54
3.3.8	Accuracy measure of rough set.....	56
3.3.9	Rough membership	57
3.3.10	Definability and rough sets	57
3.3.11	Properties of approximations	59
3.3.12	Rough Set Applications	60
3.3.13	Advantages of Rough Sets	64
3.3.14	disadvantages of Rough Sets	62
3.3.15	Crisp, fuzzy, and rough sets: an illustration	63
3.4	Fuzzy Rough sets and Rough Fuzzy sets.....	65
3.4.1	Fuzzy equivalence classes.....	65
3.4.2	Fuzzy Rough sets....	66
3.4.3	Rough Fuzzy sets....	67
3.5	Summary.....	68

4. Chapter 4: Data clustering techniques : Related work	69
4.1 Introduction.....	69
4.2 Hard C-Means clustering algorithm	69
4.3 Fuzzy C-Means clustering algorithm.....	71
4.4 Rough C-Means clustering algorithm.....	74
4.5 Fuzzy Rough C-Means clustering algorithm.....	78
4.6 Rough Fuzzy C-Means Clustering algorithm	80
4.7 Summary.....	82
 Chapter 5: The proposed Fuzzy Rough data clustering technique	86
5.1 Introduction	83
5.2 Merging algorithm.....	84
5.2.1 Merging algorithm steps	86
5.2.2 Updating cluster center algorithm	86
5.3 MFRCM steps.....	87
5.4 Clustering validity indices	87
5.5.1 Crisp form.....	88
5.5.2 Fuzzy form.....	88
5.5.3 Fuzzy Rough form..	89
5.5 Summary.....	89
 6. Tests and Results.....	90
6.1 Introduction.....	90
6.2 Implementation Environment	91
6.3 Parameters values.....	91
6.4 Experiments and results.....	92