

Agent Based Dynamic Partitional Clustering

D.M.M.B. Dehideniya

119154F



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Faculty of Information Technology

University of Moratuwa

November 2013

Agent Based Dynamic Partitional Clustering

D.M.M.B. Dehideniya

119154F



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Dissertation submitted to the Faculty of Information Technology, University of
Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the
Degree of MSc in Artificial Intelligence

November 2013

Declaration

I declare that this dissertation does not incorporate, without acknowledgment, any material previously submitted for a Degree or a Diploma in any University and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, to be made available for photocopying and for interlibrary loans, and for the title and summary to be made available to outside organizations.

Name of Student :  D. Umesh De Silva
Signature of Student
University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk
Date:

Supervised by

Name of Supervisor : Prof. A.S. Karunananda

Signature of Supervisor

Date:

Dedication

To

my father and mother

for



their love and inspiration throughout my life.
University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Acknowledgement

With a great pleasure, I present this dissertation on “Agent Based Dynamic Partitional Clustering” for partial fulfillment of my Degree M.Sc. in Artificial Intelligence. It would not be a reality without the assistance and dedication of numerous individuals.

First and foremost, I pay my sincere and heartiest gratitude to my supervisor Professor A.S. Karunananda for introducing this research area to me, for advice and guidance throughout my research work and preparation of this dissertation. Further I convey my earnest thanks to Professor G. Rzevski for his valuable comments during the early stage of my research work.

I express my heartfelt thanks to my parents and my sister, for their continuous motivation and inspiration during this research work, as usual in my life. Last but not least, I express my cordial thanks to my friends for their support and encouragement during this study.



University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Abstract

Most of the well established clustering algorithms assume that the underlying clustering structure of data set does not change over time. Hence, those algorithms fail to identify underlying cluster structures in currently available large scale dynamic data sources in an efficient manner. According to the literature, there were many attempts to address this issue by extending well established static clustering algorithms for dynamic context. For instance, Incremental K-Means and Incremental DBSCAN algorithms are two such attempts. Those methods update the model periodically according to changes in the database or change the model parameters whenever new data appear in an incremental manner. Additionally, Fuzzy Logic and Rough Set concepts are also employed to deal with the uncertainty in dynamic data sources. Other than these, Multi-agent technology has been used to address issues in data clustering in both static and dynamic contexts. PADMA and PAPHYRUS are two reported Multi-agent clustering systems in 1990s. Moreover, Chaimontree, Atkinson, and Goenen developed a Multi-agent based clustering system which provides the cluster configuration by the interaction and negotiation between cluster agents that represent each cluster in the data set.

Although there were many attempts to develop agent based clustering algorithms, but there are lesser number of reported works on identification of partitional clusters in a dynamic data source. The study presented in this thesis proposes a Multi-agent based approach to identify partitional clusters in a dynamic data source. Set of partitional clusters in a dynamic data source is identified by interactions and negotiations among the agents who represent data records in the data source. Then identified potential clusters are assigned to what are called Cluster agents. By interactions and negotiations between cluster agents and Data Record agents, the identified cluster configuration is continuously improved according to the internal cluster evaluation measures.

The proposed method is evaluated by synthetic data sets with different number of clusters in 2D and 3D spaces. Results indicate that the proposed method successfully identifies the clusters in those data sets with minimal human intervention.

Contents

	Page
Chapter 1 Introduction	1
1.1 Introduction	1
1.2 Background and Motivation	2
1.3 Problem Addressed	4
1.4 Aim and Objectives	4
1.5 Multi-agent Technology for Dynamic Data Clustering	4
1.6 Outline of the Thesis	5
1.7 Summary	5
Chapter 2 State of the Art of Dynamic Clustering	6
2.1 Introduction	6
2.2 An Overview of Dynamic Clustering	6
2.3 Major Approaches in Dynamic Clustering	7
2.4 Multi-agent Technology in Data Clustering	10
2.5 Limitations of Existing Clustering Algorithms	12
2.6 Summary	13
Chapter 3 An Overview : Multi-agent Technology	14
3.1 Introduction	14
3.2 An Overview of Multi-agent Technology	14
3.3 Belief-Desire-Intention Agent Architecture	15
3.4 Jadex	15
3.5 Tropos	17

	Page
3.6 Summary	18
Chapter 4 Agent Based Approach for Dynamic Clustering	19
4.1 Introduction	19
4.2 Hypothesis	19
4.3 Proposed Solution	19
4.3.1 Input and Output	20
4.3.2 Process	20
4.4 Features of the Proposed Solution	21
4.5 Hypothetical Example	21
4.6 Summary	23
Chapter 5 Design	24
5.1 Introduction	24
5.2 Top level Architecture of the ABDC System	24
5.3 Multi-agent Environment	25
5.3.1 Design Details of the Data Record Agent	26
5.3.2 Design Details of the Cluster Agent	27
5.4 Interaction Protocols	29
5.5 Data Retrieval Component	31
5.6 Cluster Result Visualization Component	31
5.7 Summary	32
Chapter 6 Implementation	33
6.1 Introduction	33
6.2 Implementation of the Proposed Multi-agent Environment	33



	Page
6.2.1 Implementation of Agents in the Multi-agent Environment	33
6.2.2 Implementation of the Data Record Agent	34
6.2.3 Implementation of the Cluster Agent	35
6.3 Implementation of Messages used in Interactions Between Agents	36
6.4 Implementation of the Data Retrieval Component	37
6.5 Implementation of Cluster Result Visualization Component	38
6.6 Summary	39
Chapter 7 Evaluation	40
7.1 Introduction	40
7.2 Ability of Identifying the Cluster Configuration of a Dynamic Data Source	40
7.3 Performance of the Proposed Method in 2D and 3D space	42
7.4 Accuracy Comparison with K-Mean Algorithm	44
7.5 Summary	45
Chapter 8 Conclusion and Further Work	46
8.1 Introduction	46
8.2 Major Achievements	46
8.3 Significance of the Proposed Method	47
8.4 Problems Encountered	48
8.5 Limitations of the Study	49
8.6 Further Work	49
References	50
Appendix A : Sample Agent Definition File - Cluster Agent	52
Appendix B : Application Definition File	53

List of Figures

	Page
Figure 2.1 : Cycle of Dynamic Soft Computing Clustering	9
Figure 3.1 : Jadex Abstract Architecture	16
Figure 3.2 : Main Components of Jadex BDI Agent	17
Figure 4.1 : Scatter Plot of a Hypothetical Data Set	21
Figure 4.2 : Cluser Configurations at Various Stages of the Cluster Identification Process	22
Figure 5.1 : Top Level Architecture of the Proposed System	24
Figure 5.2 : Proposed Multi-agent Environment	25
Figure 5.3 : Goal Diagram of the Data Record Agent	26
Figure 5.4 : Goal Diagram of the Cluster Agent	28
Figure 5.5 : Initial Membership Request-Response Interaction Protocol	30
Figure 5.6 : Inter-Cluster Merging Request-Response Protocol	30
Figure 6.1 : Jadex Control Window	34
Figure 6.2 : Graphical User Interface of Data Source Agent	37
Figure 6.3 : GUI of the Cluster Result Visualization Component	38
Figure 7.1 : Changes in Cluster Configurations while New Data Records Added	41
Figure 7.2 : Variations in the Number of Clusters with Time	42
Figure 7.3 : Cluster Result of 2D Data Set No.1	42
Figure 7.4 : Cluster Result of 2D Data Set No. 2	43
Figure 7.5 : Cluster Result of 2D Data Set No.3	43
Figure 7.6 : Scatter Plot Matrix of 3D Data Set No. 2	44
Figure 7.7 : Cluster Result of 2D Data Set No. 4	45

List of Tables

	Page
Table 5.1 : Agent Specification of the Data Record agent	27
Table 5.2 : Agent Specification of the Cluster Agent	28
Table 5.3 : Agent Specification of the Data Source Agent	31
Table 6.1 : Definition of the Cluster Merge Request Message	36
Table 7.1 : Details of Synthetic Data Set	40
Table 7.2 : Comparison of Cluster Results of the Proposed Method and K-means Algorithm	44



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk