Title of the Project

Submitted By

Student Name

Roll No

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

December 2014

Title of the Project

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering

Submitted By

Student Name

(Roll No.)

Guided By

Guide Name

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

December 2014

Certificate

This is to certify that the major project entitled **"Tile of the Project"** submitted by **Student Name (Roll No: 13MCECXX)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. K. P. Agrawal
Guide & Associate Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Prof. Vijay Ukani
Associate Professor,
Coordinator M.Tech - CSE
Institute of Technology,
Nirma University, Ahmedabad

Dr. Sanjay Garg
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr K Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Student Name**, Roll. No. **13MCECXX**, give undertaking that the Major Project entitled "**Title of the Project**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

_____

Signature of Student

Date:

Place:

Endorsed by

Guide Name

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. X. Y. Abcedef**, Associate Professor, Computer Science Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr K Kotecha**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

See that you acknowledge each one who have helped you in the project directly or indirectly.

- Student Name

13MCECXX

Abstract

Clustering aggregation problem is a kind of formal description for clustering ensemble problem and technologies for the solving of clustering aggregation problem can be used to construct clustering division with better clustering performance when the clustering performances of each original clustering division are fluctuant or weak. In this paper, an approach based on genetic algorithm for clustering aggregation problem, named as GeneticCA, is presented To estimate the clustering performance of a clustering division, clustering precision is defined and features of clustering precision are discussed In our experiments about clustering performances of GeneticCA for document clustering, hamming neural network is used to construct clustering divisions with fluctuant and weak clustering performances. Experimental results show that the clustering performance of clustering division constructed by GeneticCA is better than clustering performance of original clustering divisions with clustering precision as criterion.

Abbreviations

DBSCAN	Density-based spatial clustering of applications with noise.
QGIS	Quantum geographic information systems.
LEACH	Low Energy Adaptive Clustering Hierarchy.
FCM	Fuzzy C-Means

Contents

Certificate **iii**

Statement of Originality **iv**

Acknowledgements **v**

Abstract **vi**

Abbreviations **vii**

List of Figures **ix**

List of Tables **1**

1 Introduction **2**
1.1 Knowledge Discovery Process 2

2 Literature Survey **3**
2.1 Techniques :- 3

3 Demo Chapter **5**
3.1 Sample Comparison 5
3.1.1 Comparison Graph: 5

4 New Demo Chapter **6**
4.1 Sample Experiment 6
4.2 Equation 7

5 New Sample Chapter **8**
5.1 Sample Algorithm: 8

References **9**

List of Figures

3.1 comparison-1 graph 5
3.2 comparison-2 graph 5

4.1 demo figure 1 6

5.1 Sample Algorithm 8

List of Tables

2.1 My caption 3

Chapter 1

Introduction

1.1 Knowledge Discovery Process

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. The data sources can include databases, data warehouses, theWeb, other information repositories, or data that are streamed into the system dynamically.

Knowledge Discovery Process Steps [1] :-

- Cleaning the Data.
- Integrate Data.
- Select Data .
- Transformation of Data.
- Data mining.
- Pattern finding.

Chapter 2

Literature Survey

2.1 Techniques :-

Although a very rudimentary learning scheme, 1R does accommodate both missing values and numeric attributes. It deals with these in simple but effective ways. Missing is treated as just another attribute value so that, for example, if the weather data had contained missing values for the outlook attribute, a rule set formed on outlook would specify four possible class values, one for each of sunny, overcast, and rainy, and a fourth for missing.

Techniques shown in table:[2]

head1	**head2**	**head3**	**head4**	**head5**
a	b	c	d	e
l	m	n	o	t
v	w	x	y	z

Table 2.1: My caption

Sample Table shown in table 2.1 .

The diverse density is defined as the probability of the class labels of the bags in the training data, computed based on this probabilistic model. It is maximized when the reference point is located in an area where positive bags overlap and no negative bags are present, just as for the two geometric methods discussed previously. A numerical optimization routine such as gradient ascent can be used to find the reference point that maximizes the diverse-density measure. In addition to the location of the reference point,

implementations of diverse density also optimize the scale of the distance function in each dimension because generally not all attributes are equally important. This can improve predictive performance significantly.

Chapter 3

Demo Chapter

3.1 Sample Comparison

3.1.1 Comparison Graph:

figure 3.1 and 3.2 shows comparison:[2] [3]

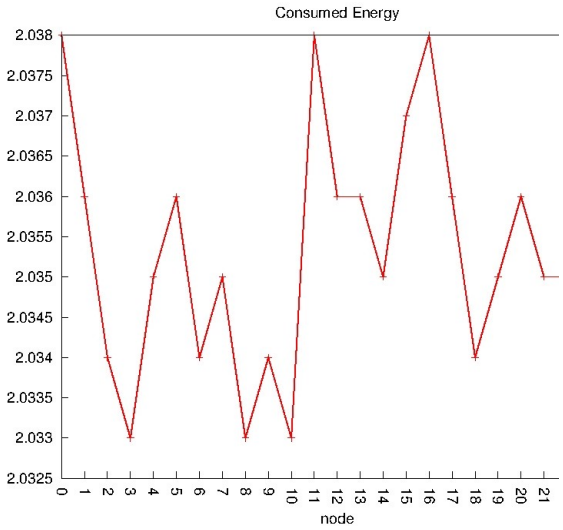

Figure 3.1: comparison-1 graph

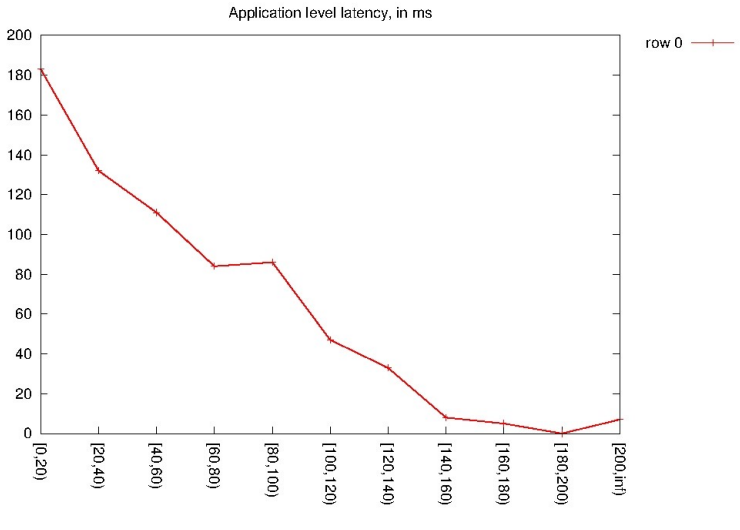

Figure 3.2: comparison-2 graph

Chapter 4

New Demo Chapter

4.1 Sample Experiment

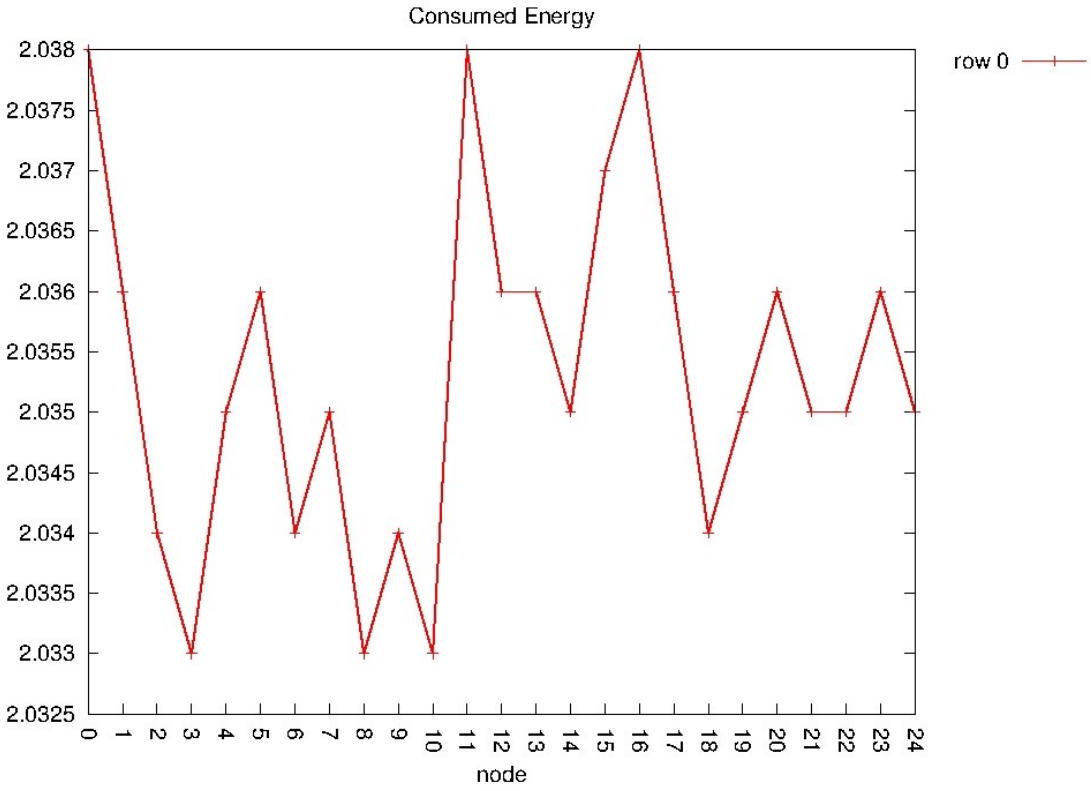

Figure 4.1: demo figure 1

4.2 Equation

Example Equations [2]

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4}}}} \tag{4.1}$$

$$\frac{(x_1 x_2) \times (x_1' x_2')}{(y_1 y_2 y_3 y_4)} \tag{4.2}$$

Equation 4.1 is a Continued fractions.

Equation 4.2 is Multiplication of two numbers.

$\sqrt{\frac{a}{b}}$

$\sqrt[n]{1 + x + x^2 + x^3 + \ldots}$

Chapter 5

New Sample Chapter

5.1 Sample Algorithm:

Algorithm 15.3 Gibbs Sampling

for $i = 1, \ldots, n$ **do** [bootstrapping]
    Initialize $P(Y_i | \mathbf{X}, \mathbf{Y}_{\mathcal{N}_i})$ using local features X_i and the features $\mathbf{X}_{\mathcal{N}_i}$ and *observed* labels $\mathbf{Y}^{\ell}_{\mathcal{N}_i} \subseteq \mathbf{Y}_{\mathcal{N}_i}$ of its neighbors
end for
for $i = 1, \ldots, n$ **do** [initialize samples]
    $\mathcal{S}_i \leftarrow \emptyset$
end for
repeat[sampling]
    $\pi \leftarrow \textsc{GenPerm}(n)$ # generate permutation π over $1, \ldots, n$
    for $i = 1, \ldots, n$ **do**
        Sample $s \sim P(Y_i | \mathbf{X}, \mathbf{Y}_{\mathcal{N}_i})$
        $\mathcal{S}_i \leftarrow \mathcal{S}_i \cup s$
        Update $P(Y_i | \mathbf{X}, \mathbf{Y}_{\mathcal{N}_i})$ using $\mathcal{S}_i$
    end for
until convergence or maximum iterations reached
for $i = 1, \ldots, n$ **do** [compute marginals]
    Remove first T samples (i.e., burn-in) from $\mathcal{S}_i$
    for $y \in \mathcal{Y}_i$ **do**
        $P(Y_i = y) \leftarrow \frac{1}{|\mathcal{S}_i|} \sum_{s \in \mathcal{S}_i} \mathbb{1}[y = s]$
    end for
end for

Figure 5.1: Sample Algorithm

Sample Algorithm shown in figure 5.1 Gibbs Sampling.

References

[1] C. C. Aggarwal, *Social Network Data Analytics*. Springer, 2011.

[2] D. Vidhate and P. Kulkarni, "Cooperative machine learning with information fusion for dynamic decision making in diagnostic applications," pp. 70–74, Aug 2012.

[3] Y. Jin and B. Sendhoff, "Pareto-based multiobjective machine learning: An overview and case studies," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 38, pp. 397–415, May 2008.