Analysis of MST based clustering algorithm with different threshold values

Dissertation

Submitted To:

Central Department of Computer Science & Information Technology

Tribhuvan University

Kirtipur, Kathmandu

Nepal

In partial Fulfillment of the requirements for the Degree of Master of Science in Computer Science & Information Technology

Submitted by:

Lalit Pant

February, 2016

Supervisor

Prof. Dr. Subarna Shakya

Co-supervisor

Mr. Arjun Singh Saud

Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

Student's Declaration

I hereby declare that I am the only author of this work and that no sources other than the listed here have been used in this work.

… … … … … … …
Lalit Pant

Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

Supervisor's Recommendation

I hereby recommend that the dissertation prepared under my supervision by **Mr. Lalit Pant** entitled "**Analysis of MST based clustering algorithm with different threshold values**" be accepted as in fulfilling partial requirement for the completion of Masters Degree of Science in Computer Science & Information Technology.

Prof. Dr. Subarna Shakya

Department of Electronics & Computer Engineering,
Institute of Engineering,
Pulchowk, Nepal

Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

LETTER OF APPROVAL

We certify that we have read this dissertation work and in our opinion it is appreciable for the scope and quality as a dissertation in the partial fulfillment of the requirements of Masters Degree of Science in Computer Science & Information Technology.

Evaluation Committee

Asst. Prof. Nawaraj Paudel
Head of Department
Central Department of Computer Science
& Information Technology
Tribhuvan University
Kirtipur

Prof. Dr. Subarna Shakya
(Supervisor)
Department of Electronics & Computer
Engineering, Institute of Engineering,
Pulchowk, Nepal

(External Examiner)

(Internal Examiner)

ACKNOWLEDGEMENT

I would never have been able to finish my dissertation without the guidance, support and encouragement of numerous people including my supervisor, co-supervisor, my friends, colleagues and support from my family. At the end of my thesis I would like to thank all those people who made this thesis possible and an unforgettable experience for me.

First, I would like to express my gratitude to my supervisor **Professor Dr. Subarna Shakya,** Institute of Engineering, Pulchowk Campus for his continuous support without which the thesis wouldn't have been possible to complete. His suggestions, guidance, thorough knowledge and expertise helped me immensely in understanding and developing this thesis. Also, the credit of the success of this dissertation work goes to my co-supervisor **Mr. Arjun Singh Saud**. I appreciate for his supervision and guidance.

Most importantly I would like to thank to respected Head of Department of Central Department of Computer Science and Information Technology, **Asst. Prof. Nawaraj Paudel** for his kind support, help and constructive suggestions. I am very much grateful and thankful to all the respected teachers Professor Dr. Shashidhar Ram Joshi, Mr. Dhiraj Kedar Pandey, Mr. Sarbin Sayami, Mrs. Lalita Sthapit, Mr. Bikash Balami and Mr. Jagdish Bhatta for providing me such a broad knowledge and inspirations. I am so much thankful to Mr. Deep Sharma for his continuous support throughout the thesis work and also like to thank my dear friends Ishwari, Rajendra, Dinesh, Chakra, and Bhim, for their cooperation.

Special thanks to my family for their endless motivation, constant mental support and love which have been influential in whatever I have achieved so far. All my class fellows are worthy of my gratefulness for their direct or indirect support in completion of my dissertation.

I have done my best to complete this research work. Suggestions from the readers are always welcomed, which will improve this work.

ABSTRACT

Clustering analysis has been an emerging research issue in data mining due to its variety of applications. Many algorithms are proposed so far, however each algorithm has been its own merits and demerits and cannot work for real situation. The MST based clustering algorithms have been widely used due to their ability to detect cluster with irregular boundaries. In this dissertation the clustering algorithm is inspired by MST.

In this dissertation the MST based clustering algorithm has been analyzed using different threshold value on MST and measured by validity index. Given the MST over data set, select or reject the edges of MST in process of forming the clusters, depending on the threshold value. Validity index is the ratio of intra cluster distance and inter cluster distance. Thresholds are taken by mean, standard deviation and mean + standard deviation of MST. These thresholds are evaluated by validity index. Smallest value of validity index is select for best clustering and best threshold value. The algorithm has been tested on the randomly generated data sets and as well as real world data sets.

Keywords: Clustering Algorithm, MST, Validity Index, Threshold Values

CONTENTS

DETAILS **PAGE NO**

CHAPTER 4
ALGORITHMS

CHAPTER 5
IMPLEMENTATION

CHAPTER 6
RESULT AND ANALYSIS

CHAPTER 7
CONCLUSION AND FUTURE WORK

LIST OF FIGURES

FIGURES **PAGE NO**

LIST OF TABLES

TABLES **PAGE NO**

LIST OF ABBREVIATIONS

MST	:	Minimum Spanning Tree
EMST	:	Euclidean Minimum Spanning Tree
SEMST	:	Standard Euclidean Minimum Spanning Tree
HEMST	:	Hierarchical Euclidean MST
MSDR	:	Maximum Standard Deviation Reduction
DHCA	:	Divisive Hierarchical Clustering Algorithm
MDHCA	:	Multi Divisive Hierarchical Clustering Algorithm.
LM	:	Larzlo and Mukherjee
SI	:	Sequential Initialization
VI	:	Validity Index