

**CLUSTERING CUSTOMER'S CURRENT PROFILE USING
HIERARCHICAL CLUSTERING**

This thesis is presented in partial fulfillment for the award of the

Bachelor of Electrical Engineering (Hons)

UNIVERSITI TEKNOLOGI MARA



MOHD HAIRI BIN MOHD ZIN

2004346976

Faculty of Electrical Engineering

UNIVERSITI TEKNOLOGI MARA

40450 SHAH ALAM,

SELANGOR DARUL EHSAN

ACKNOWLEDGEMENT

First of all thank to Allah for allowing this Industrial Project to complete. I would like to take this opportunity to express my highest gratitude and appreciation to my supervisor's, Dr Zuhaina binti Hj. Zakaria for his valuable advices on the framework of this dissertation, his careful review of the results and endless guidance during the completion of the work

Special thanks dedicated to my family who has been very encouraging and helpful towards the completion of this project as well as completing this course. Without their invaluable help and support, this project would never complete.

Last but not least, it is also a pleasure to gratefully acknowledge the support of friends, lecturers and colleagues at Universiti Teknologi MARA, Shah Alam who have help in this preparation of this thesis.

Mohd Hairi Bin Mohd Zin

UiTM Shah Alam

ABSTRACT

Clustering is a data mining activity that aims to differentiate groups inside a given set of objects, with respect to a set of relevant attributes of the analyzed objects. Generally, existing clustering methods start with a known set of objects, measured against a known set of attributes. But there are numerous applications where the attribute set characterizing the objects evolves. This paper proposed an incremental clustering method based on a hierarchical clustering, that is capable to re-partition the object set, when the attribute set increases.

The method starts from the partitioning into clusters that was established by applying the Hierarchical clustering (HC) before the attribute set changed. The current load profile can also indicate the type of consumers that connected to the feeder. In order to compare the performance of hierarchical clustering, a cophenet correlation coefficient was used. The closer the value of the cophenetic correlation coefficient is to one, the more accurately the clustering solution reflects the data.

TABLE OF CONTENT

DECLARATION	IV
ACKNOWLEDGEMENT	V
ABSTRACT	VI
TABLE OF CONTENT	VII
LIST OF FIGURES	X
LIST OF TABLES	XI
LIST OF ABBREVIATION	XII
CHAPTER 1	1
INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 PROJECT OBJECTIVES	3
1.3 SCOPE OF WORK	3
1.4 THESIS ORGANIZATION	5
CHAPTER 2	6
LITERATURE REVIEW	6
2.1 LOAD PROFILE	6
2.2 LOAD CURVE	8
2.2.1 CURRENT LOAD PROFILE	8
2.3 TYPE OF LOAD PROFILE	9
CHAPTER 3	10
CLUSTER ANALYSIS	10
3.2 CLUSTER ANALYSIS	11

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Clustering analysis is a fundamental but important tool in statistical data analysis. In the past, the clustering techniques have been widely applied in a variety of scientific areas such as pattern recognition, information retrieval, microbiology analysis, and so forth. Unsupervised classification, or clustering, is a data mining activity that aims to differentiate groups (classes or clusters) inside a given set of objects. The inferring process is carried out with respect to a set of relevant characteristics or attributes of the analyzed objects.

The resulting subsets or groups, distinct and non-empty, are to be built so that the objects within each cluster are more closely related to one another than objects assigned to different clusters. Central to the clustering process is the notion of degree of similarity (or dissimilarity) between the objects. Similarity and dissimilarity between objects are calculated using metric or semi-metric functions, applied to the attribute values characterizing the objects. A large collection of clustering algorithms is available in the literature, [1] and [2] contain comprehensive overviews of existing techniques.

Hierarchical clustering methods represent a major class of clustering techniques. There are two styles of hierarchical clustering algorithms. Hierarchical Clustering techniques can be divisive or agglomerative. A divisive method begins with all cases in one cluster. This cluster is gradually broken down into smaller and smaller clusters. Agglomerative techniques start with single member clusters. These are gradually fused until one large cluster is formed.