Al Al-Bayt University

Computer Science Department

Prince Hussein Bin Abdullah College for Information Technology

Two Level clustering hierarchies using Fuzzy clustering by Local Approximation of Memberships

خوارزمية للتقسيم الهرمي الثنائي باستخدام التجميع الغامض من خلال تقريب العضوية المحليه

by

Osama Sleman Yusuf Mashaqba

Supervisor: Dr. Khaled Batiha

Co. Supervisor: Dr. Wafa' Slaibi AL-Sharfat

This Thesis was submitted in Partial Fulfillment of the Requirements For the Master’s Degree of Science in Computer Science

Deanship of Graduate Studies

Al Al-Bayt University

نموذج رقم (1)

جامعة آل البيت
عمادة الدراسات العليا

نموذج تفويض

أنا الطالب أسامة سليمان يوسف مشاقبة

افوض جامعة آل البيت بتزويد نُسخ من رسالتي، للمكتبات أو المؤسسات أو الهيئات أو الأشخاص عند طلبهم حسب التعليمات النافذة في الجامعة.

التوقيع: .... التاريخ:

نموذج رقم (2)

جامعة آل البيت
عمادة الدراسات العليا

نموذج اقرار والتزام بقوانين جامعة آل البيت وانظمتها وتعليماتها لطلبة الماجستير والدكتوراه.

انا الطالب أسامة سليمان يوسف مشاقبة الرقم الجامعي: 1320901016

تخصص: علم حاسوب كلية: الحسين بن عبد الله لتكنلوجيا المعلومات

أعلنُ بأني قد التزمت بقوانين جامعة آل البيت وانظمتها وتعليماتها وقراراتها السارية المفعول المتعلقة بإعداد رسائل الماجستير والدكتوراه عندما قمت شخصياً بإعداد رسالتي / اطروحتي بعنوان:

Two Level Clustering Hierarchies using Fuzzy Clustering by Local Approximation of Memberships

خوارزمية للتقسيم الهرمي الثنائي باستخدام التجميع الغامض من خلال تقريب العضوية المحليه

وذلك بما ينسجم مع الأمانة العلمية المتعارف عليها في كتابة الرسائل والأطاريح العلمية. كما أنني أُعلن بأن رسالتي/ اطروحتي هذه غير منقولة أو مستلة من رسائل أو أطاريح أو كتب أو أبحاث أو أي منشورات علمية تم نشرها أو تخزينها في أي وسيلة اعلامية، وتأسيساً على ما تقدم فأنني اتحمل المسؤولية بأنواعها كافة فيما لو تبين غير ذلك بما فيه حق مجلس العمداء في جامعة آل البيت بإلغاء قرار منحي الدرجة العلمية التي حصلت عليها وسحب شهادة التخرّج مني بعد صدورها دون أن يكون لي الحق في التظلم أو الأعتراض أو الطعن بأي صورة كانت في القرار الصادر عن مجلس العمداء بهذا الصدد.

التوقيع .... التاريخ:

Committee Decision

This Thesis Two Level clustering hierarchies using Fuzzy clustering by Local Approximation of Memberships was Successfully Defended and Approved on 3djan. 2018.

Examination Committee **Signature**

Prof. Khaled Batiha , (Supervisor)
Associate Prof
Computer Science Department, Al al-Bayt University ....
batihakhalid@aabu.edu.jo

Dr.Wafa' Slaibi AL-Sharfat, (Co.Supervisor)
Assistant Prof ....
Information Systems Department, Al al-Bayt University
wafa@aabu.edu.jo

Dr. Faisal Sulaiman Al-Saqar, (Member)
Assistant Prof
Computer Science Department, Al al-Bayt University ....
Faisalss@aabu.edu.jo

Dr. Mohammad Albasheer, (Member)
Assistant Prof
Computer Science Department, Al al-Bayt University ....
mohdelb@aabu.edu.jo

Dr. Belal Abu Ata, (External Member)
Associate Prof
Information Systems Department, Yarmouk Univesity ....
Belalabuata@yu.edu.jo

Dedication

This thesis is dedicated to my parents, my brother, my sister, Dr. Khaled Batiha and Dr. Wafa'a AL-Shourfat for their courtliness, love, appreciating to support and encouragement me.

To every one who ask Allah to help

me in my thesis. For all their, ask my Allah to give them what they wish.

Acknowledgments

I would like to thank my supervisor Dr. Khaled Batiha and Co. Supervisor Dr. Wafa'a AL-Shourfat, for their sincere advice and guidance provided throughout my research and thesis preparation.

My amiable thanks and appreciation go to the defense committee for their valuable comments after reading this study; namely, Dr Faisal Sulaiman Al-Saqar, Dr. Mohammad Albasheer, and Dr. Belal Abo Ata.

Special thanks for my Father, Mother, Uncle Salem, my Aunt, and my best friend Raed Haraphsheh to their support and standing with me during my writing this thesis.

Finally i ask allah to keep my grandmother in his mercy and accept it in his sins.

Osama sleman mashaqba

Table of contents

1 **Introduction………………………………………………………………………… 1**

1.1 Introduction………………………………………………………………………... 1

1.2 Wireless sensor network (WSN)………………………… ………………………….. 1

1.3 Fuzzy Logic……………………………………………………………………….. 3

1.4 Motivation ………………………………………………………………………… 3

1.5 Problem Statement ………………………………………………………………... 4

1.6 Contributions ……………………………………………………………………… 4

1.7 Thesis Structure …………………………………………………………………… 5

2 **Research Background ……………………………………………………………….. 6**

2.1 Introduction ……………………………………………………………………….. 6

2.2 Clustering …………………………………………………….............................. 6

2.3 Fuzzy clustering …………………………………………………………………... 7

2.3.1 C-means …………………………………………………………………… 8

2.3.2 Hard C-means (K-means) ……………………………….......................... 8

2.3.3 Fuzzy Clustering by Local Approximation Membership (FLAME) …………... 9

2.4 K-nearest algorithm neighbors ……………………………………...................... 11

2.5 Data set ……………………………………………………………...................... 11

2.5.1 Iris Data set …………………………………………………………….. 11

2.5.2 Dummy Data set ………………………………………………………. 15

3 **Literature Review …………………………………………………….................... 16**

3.1 FLAME Clustering ………………………………………………………………. 16

3.2 Fuzzy Clustering …………………………………………………………………. 16

3.3 K-nearest Clustering ……………………………………………………………… 18

3.4 Link-state routing algorithm ……………………………………………………….. 19

4 Methodology 20

4.1 Introduction 20

4.2 System Stages 20

5 Experimental Results 27

5.1 Introduction 27

5.2 Clustering Measurements 27

5.2.1 Error Rate 27

5.2.2 Accuracy Rate 28

5.2.3 Precision 28

5.2.4 Distance Measurements 29

5.2.4.1 Manhattan Distance 29

5.2.4.2 Euclidean Distance 29

5.3 Dataset 29

5.4 Experimental Results 29

5.4.1 The First Experiment: Internal Result 30

5.4.2 The Second Experiment: Accuracy, Error Rate, precision using IRIS dataset 32

5.4.2.1 Error Rate 32

5.4.2.2 Precision 34

5.4.2.3 Accuracy 35

5.4.3 The Third Experiment: Distance Measurement 36

5.4.4 The Fourth Experiment: Using Random Data 38

6 Conclusion and Future Work 40

6.1 Conclusion 40

6.2 Future Work 41

References 42

List of Figures

Figure 3-1	FLAME algorithm example	10
Figure 4-1	System Stages	21
Figure 4-2	(a) Example of our research method	23
	(b) Example of our research method	24
Figure 5-1	Error rate comparison for internal result	31
Figure 5-2	Accuracy comparison for internal result	31
Figure 5-3	Error Rate comparison	33
Figure 5-4	Precision comparison	34
Figure 5-5	Accuracy comparison	35
Figure 5-6	Distance Measurement comparison for 50 and 25 nodes	37
Figure 5-7	Distance Measurement comparison for 70 nodes	37
Figure 5-8	Error rate and Accuracy comparison	38

List of Tables

Table 2-1 Iris data set samples 12

Table 5-1 Measurement Parameters 30

Table 5-2 Error rate, Accuracy, precision in iris data set applying in this research 33

Table 5-3 Error Rate using Iris data set 34

Table 5-4 Precision using Iris data set 35

Table 5-5 Accuracy using Iris data set 36

Table 5-6 Distance Measurement 31

Table 5-7 Error rate and Accuracy comparison using random data 38

Two Level Clustering Hierarchies using Fuzzy Clustering by Local Approximation of Memberships

A Master Thesis By
Osama Sleman Yusuf Mashaqba

Supervisor: Dr. Khaled Batiha
Co. Supervisor: Dr. Wafa' Slaibi AL-Sharfat

Computer Science Department, Al al-Bayt University, 2018

Abstract

The simplicity, speed, and effectiveness, have made fuzzy algorithm widely used in clustering and classification of algorithms. Many algorithms such as C-means and K-means are used in clustering to split nodes to cluster and elect one main head cluster to manage the communication between the nodes in its cluster. In addition to that, there should be algorithms that make communication between node faster and more effective by using the definition of multi-layer and determining exactly where the node will be in each cluster. In this study, we presented Two Level clustering hierarchies by using Fuzzy clustering by Local Approximation of Memberships (TLCH FLAME), TLCH FLAME is presented to be used in clustering problems, and it’s better than the well-known clustering technique like K-means and C-means in the context of clustering the non-linear network clusters. In this study the TLCH FLAME was tested on iris data sets and random generated data sets on a multicore system, through a stage of experiments and by using the data sets we compare our experimental results with other methods. The results have shown that our research methods possess nearly the best result for both datasets than K-means and C-means.

Key Word: Clustering, FLAME, Fuzzy Clustering.

خوارزمية للتقسيم الهرمي الثنائي باستخدام التجميع الغامض من خلال تقريب العضوية المحليه

رسالة ماجستير قُدمت من قبل
اسامة سليمان يوسف مشاقبة

المشرف الرئيسي: د. خالد بطيحة
المشرف المشارك : د. وفاء الشرفات

قسم علم الحاسوب، جامعة آل البيت، 2018م

ملخص

من الاسباب التي جعلت خوارزمية Fuzzy تستخدم على نطاق واسع في تجميع وتصنيف العقد في الشبكة البساطة والسرعة والفعالية. يتم استخدام العديد من الخوارزميات مثل C-means و K-means في التجميع وتقسيم العقد إلى كتل واختيار عقدة رئيسية واحدة لإدارة الاتصال بين العقد في نفس التجمع. بالإضافة إلى ذلك، يجب أن تكون هناك خوارزميات تجعل الاتصال بين العقد أسرع وأكثر فعالية باستخدام مفهوم تعددية الطبقات وتحديد بالضبط أين ستكون كل عقدة وفي اي تجمع. في هذه الدراسة، قدمنا تسلسلين هرميين على مستوى المجموعات باستخدام التجميع الضبابي من خلال تقريب العضوية المحلية (TLCH FLAME)، تم تطوير TLCH FLAME لاستخدامه في حل مشاكل تجمع الكتل داخل الشبكة، ويعد TLCH FLAME أفضل من بعض التقنيات في سياق التقسيم والتجميع في الشبكات الغير خطية. في هذه الدراسة تم اختبار TLCH FLAME على مجموعتين من البيانات: الاولى هي IRIS dataset والثانية مجموعة من البيانات التي ولدت عشوائيا من خلال نظام متعدد النوى Multi Core ، من خلال مرحل من التجارب وباستخدام مجموعات البيانات السابقة قارنا نتائجنا التجريبية مع خوارزميات تجميع أخرى. وقد أظهرت النتائج أن أسلوب البحث لدينا أمتلك أفضل نتيجة بين كل من K-means والـ C-means.

Chapter 1

Introduction

1.1 Introduction

Clustering is a technique that collects and puts the similar data into group. Data grouping depends mainly on data similarity, thus, alike data are grouped together.. The concept of clustering network is considered to solve many of the limitations on the network communication. Clustering in Wireless Network is one of the most important research issues. Many fields have been used the clustering algorithm such as image processing, speech recognition, psychology, disciplines of biology, and archeology…etc. There are many algorithms proposed for clustering, such as clustering by density, priority, and grid. These algorithms are divided into two subcategories: proceed from the overall clustering, such as CURE algorithm, and starting from the individual, such as CHAMELEON algorithm (Chao Qu, etal., 2013).

1.2 Wireless Sensor Network (WSN):

Wireless Sensor Network Sometimes called wireless sensor and actor network (WSN). WSN are spatially distributed autonomous sensors. WSN contains nodes that can move completely freely in space. These nodes can be seen as hosts and routers (Wang, 2008). As hosts, nodes need to provide user-oriented services; Where as routers, nodes need to run the routing protocols and be functioning like any router in the network. Call it head cluster (central node), to determine each node where they belong in this network We adopted on a head cluster for analysis, and processing, each cluster head acts like manger on its own cluster.

Many models for WSN have been proposed, we can notice that there are characteristics for WSN in most models (Wang, 2008):

- all of them contain a large number of sensor nodes, and some nodes (one or more) that manage the network (cluster head).
- All sensor nodes main job is to collect raw data, and provide it to the head cluster node with or without some primitive preprocessing.
- The master node (head cluster) collects data from all sensors, do the analyze and process to these data,
- The master node could be the best node in its cluster due to some factor that determined by the model (Wang, 2008).

To ensure effective performance in sensor network, we should use protocols that play role in determined the head cluster, a sensor network can contains one or several head cluster that can be determined by geographical coverage or cost effectiveness consideration, in case we have one head cluster, they collect data from all sensor node on network, on case we have several head cluster each one do collecting and processing distributed among a group of sensor node that working together.

In a hierarchical cluster structure, the first layer of network there are group of head cluster that collect information from neighbor node, in second layer the selection of the head cluster depending on many factor like power consumption, network topology, and the optimization objectives. The data aggregation at cluster heads could reduce data traffic on the network.

1.3 Fuzzy Logic

Fuzzy Logic is an extension of Boolean logic that is used to help making decision in computer-based. In classic boolean either the element has a full membership or not be a member, where as in fuzzy logic the member can be a crossbar between the two values (0,1), thus it allows partial membership of elements in a collection.

In fuzzy clustering, every node has its own membership to cluster, these memberships can be determined by simple assumption to be subdivided into subgroup, instead of fully belong to one cluster, maybe the cluster contain node that have condition a lesser degree than node in the center of cluster, recently fuzzy clustering have been taken on mind because its ability to split one type of data for several cluster. The most widely used fuzzy clustering algorithm is fuzzy C-means algorithm (Bezdak, 1981), this method allow the node to belong to two or more clusters.

1.4 Motivation

Clustering, is important for several tasks like machine learning, image processing, and data mining (Chao, etal, 2013). A number of research methods have been working to build clusters for many networks, each clusters have their own advantages and disadvantages (Chao, etal, 2013).

A new clustering techniques have been proposed in the DNA biomedical field using Fuzzy Clustering by Local Approximation Memberships (FLAME). The result for this proposed algorithm make motivation for us to implement Two Level cluster hierarchies by use Fuzzy clustering by Local Approximation of Memberships (FLAME).

1.5 Problem Statement:

It is very hard to select a specific rule for the number of node in each cluster, to solve this problem and make clustering very improvement in network, as result we need an algorithm that can strongly manage the cluster and choose the best device to manage the cluster and another head clusters to help the main head cluster on collecting and processing network data on its own cluster.

Consequently, the present study gives the network the best management for the node and identify groups within the group to ensure greater similarity.

1.6 Contributions:

The main contributions of this study is to Improve the accuracy rate, and reduce the error rate of the network using FLAME. Reduce data traffic collusion by electing multi head cluster instead of one head cluster. Then use the definition of hierarchies clustering and applied it using FLAME.

1.7 Thesis Structure:

This thesis is organized as the following:

Chapter 2 includes research background.

Chapter 3 includes literature review of previous studies.

Chapter 4 clearly shows how we applied the methodology of Clustering algorithm and what happens during this process.

Chapter 5 includes result.

Chapter 6 concludes the main ideas of the present study and develops future achievements.

Chapter 2
Research Background

2.1 Introduction:

The classification of node in the network requires the use of classification techniques so as to detect to which group the node belongs on the network, the fuzzy clustering is a form of clustering in which every node in the network can be assigned to multiple cluster, the fuzzy clustering was developed by Dunn in 1973 and improved by Bezdek (J.C. Bezdek, etal., 1981). In this study we use Fuzzy clustering by Local Approximation of Memberships algorithm to construct two layer of network topology, the experiments with our study applied on the iris dataset that used on most Clustering method .

2.2 Clustering:

Clustering is the task of making groups of data from data sets. These group are called clusters. Each cluster has nodes that are more similar than the other nodes in other clusters. A cluster is also called data or node segmentation because the clustering does the partition for large data or nodes into similar clusters according to their similarity. However, different research methods have been employed for different clustering models, these models are:

- Connectivity based clustering model: also known as hierarchical clustering. The main idea of this model is grouping data or nodes based on their distance; the closer the node the more related together than the node further away (Jain, etal., 2012).

- Centroid based clustering model: in these models: each cluster is represented by a single central vector. This vector is not a member of any data set. The best example of this model is k-means, the k-means clustering gives a formal definition for an optimized problem by finding the k-centers of the assigned data or node to the nearest k-center (Jain, etal., 2012).
- Hard clustering model: each node is either belong to a cluster or will not belong to any cluster (Jain, etal., 2012).
- Soft clustering model: also known as fuzzy clustering. In this model each data or node belongs to more than one cluster (Jain, etal., 2012).
- Density-based clustering model: in density-based clustering, clusters are built by the densest node than the reminder nodes. These nodes, which are usually far away, are considered to be border points (Jain, etal., 2012).
- Strict partitioning clustering model: each node will belong to exactly one cluster (Jain, etal., 2012).
- Strict partitioning clustering with outlier model: each node will belong to exactly one cluster, or can belong to no cluster that called outer (Jain, etal., 2012).

We used in our approach connectivity based clustering model, soft clustering model, and strict partitioning clustering with outlier model.

2.3 Fuzzy clustering:

Fuzzy clustering is a type of clustering in which every node in the network can be assigned to multiple cluster. In regular clustering, each node is assigned only one cluster.

Cluster identified on application requirement or data used. These measures can be distance measurement, density, and connectivity. The fuzzy term was used with the 1965 suggestion proposal by Lotfi Zadeh (Zadeh, 1965).

The fuzzy clustering contains zero and one as classical clustering and it contains other case, the objective of this technique is not to force nodes to be on specific cluster.

2.3.1 C-means:

The fuzzy C-means (FCM) is the most fuzzy clustering algorithm used in the network. The FCM has been developed by Dunn (Dunn, 1973) and improved by Bezdak (Bezdak, 1981). In FCM each node has its own membership to a specific cluster. Each cluster has its own cluster head which is updated iteratively (Chattopadhyay, 2011).

2.3.2 Hard C-means (K-means):

K-Means or Hard C-Means algorithm is one of the simplest algorithms that solve the well-known data clustering problem. The k-mean tries to find the number of center data (cluster head) that are defined by the user. Number of center data is also determines the number of clusters. If these center data are placed on different locations, they give different results. So the best choice is to place it far away from each other. The k-means has been presented by Mac Queen (Queen, 1967).

The difference between K-means and C-means is that in K-means each data points assigned to one cluster, while for the C-means, each data point is assigned to all available clusters. The disadvantages of both algorithms are being unable to deal with outliers and noise data. Also, they have failed with non-linear data set.

2.3.3 Fuzzy Clustering by Local Approximation Membership (FLAME):

Fuzzy Clustering by Local Approximation Membership (FLAME) is a cluster algorithm that defines the cluster by dense point of data set. Fu and Medico used the FLAME algorithm in biology (Fu, etal., 2007). They used the FLAME algorithm because of its simplicity, better performance and strength of system.

FLAME algorithm is divided into three major stages:

Stage 1: Extracting data structure from dataset:

A. build a neighborhood graph to connect each node to its K-Nearest Neighbors (KNN).

B. Estimate the density for each node based on its proximities to its KNN.

C. Node are divided into three types:

1. Cluster Supporting Object (CSO), sometime called Inner: node with higher density than other nodes.

2. Cluster Outliers: node with lower density than all its neighbors, and far away from all clusters.

3. The Rest : node that is located between CSO and Outlier.

Stage 2 : Initialization of fuzzy membership. Each CSO has full membership to itself and represent its cluster. The Rest are assigned with equal memberships to all clusters. The outlier assigned full membership to outlier cluster.

Stage 3 : bulid the network cluster by fuzzy membership.

Figure 3.1 shows an example of FLAME algorithm.

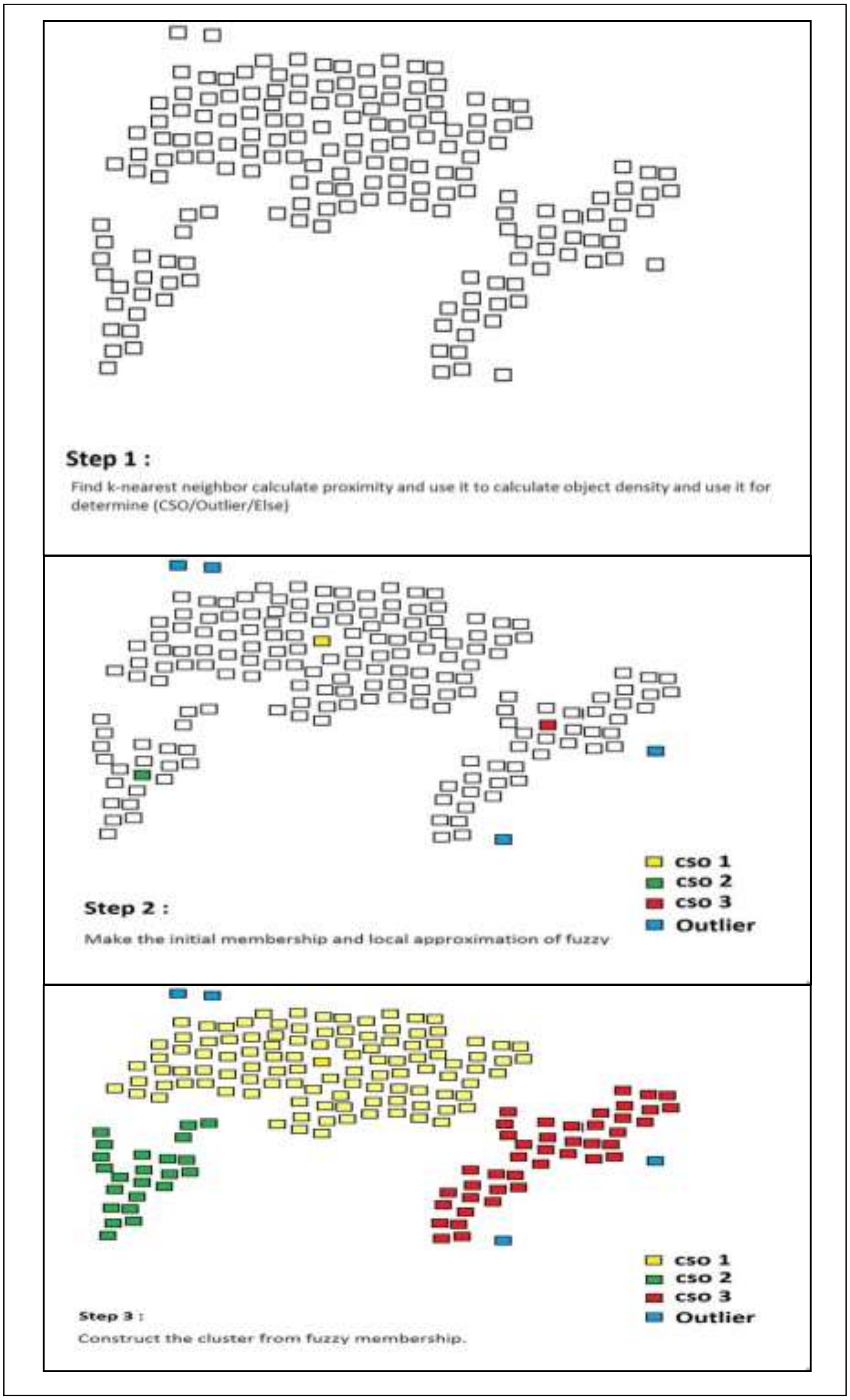

Step 1 :
Find k-nearest neighbor calculate proximity and use it to calculate object density and use it for determine (CSO/Outlier/Else)
🟨 cso 1
🟩 cso 2
🟥 cso 3
🟦 Outlier
Step 2 :
Make the initial membership and local approximation of fuzzy
🟨 cso 1
🟩 cso 2
🟥 cso 3
🟦 Outlier
Step 3 :
Construct the cluster from fuzzy membership.

Figure 3.1: FLAME algorithm example.

2.4 K-nearest algorithm neighbors:

K-nearest neighbor (KNN) was proposed by Cover in 1968 (Cover, 1968). The idea of this algorithm is to collect the same node sample and make it one cluster. These samples are divided by category which contains the most number of the same k-sample.

The closest definitions are interrelated to the k-nearest algorithm is reverse k-nearest neighbors (Cover, 1968). In this definition, if sample data A is k-nearest neighbors of sample data B then the sample data of B is reverse k-nearest neighbors of sample A. The K-nearest prove that it is effective if data set is large. On the other hand, the computation cost is high because the need of calculate the distance between all nodes.

2.5 Data set:

We used two type of dataset. The following is an explaination of these two datasets.

2.5.1 Iris Data set:

The Iris flower dataset or Fisher's Iris dataset has been introduced by the biologist Ronald Fisher (Fisher, 1936). Iris dataset sometimes called Anderson's Iris data set because Edgar Anderson collected the data from Iris flowers of three related species (Anderson, 1936). Iris dataset becomes a test case for classification techniques. These dataset include three clusters, first cluster called Setosa, the second is Versicolor, and the third one is Virginica. For each cluster, there are four numeric attributes: Sepal length, Sepal width, Petal length, and Petal width. Each numeric attributes include 50 instances.

The table (2-1) show the sample of Iris data set.

Table (2-1) Iris data set samples

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	s
4.9	3	1.4	0.2	s
4.7	3.2	1.3	0.2	s
4.6	3.1	1.5	0.2	s
5	3.6	1.4	0.2	s
5.4	3.9	1.7	0.4	s
4.6	3.4	1.4	0.3	s
5	3.4	1.5	0.2	s
4.4	2.9	1.4	0.2	s
4.9	3.1	1.5	0.1	s
5.4	3.7	1.5	0.2	s
4.8	3.4	1.6	0.2	s
4.8	3	1.4	0.1	s
4.3	3	1.1	0.1	s
5.8	4	1.2	0.2	s
5.7	4.4	1.5	0.4	s
5.4	3.9	1.3	0.4	s
5.1	3.5	1.4	0.3	s
5.7	3.8	1.7	0.3	s
5.1	3.8	1.5	0.3	s
5.4	3.4	1.7	0.2	s
5.1	3.7	1.5	0.4	s
4.6	3.6	1	0.2	s
5.1	3.3	1.7	0.5	s
4.8	3.4	1.9	0.2	s
5	3	1.6	0.2	s
5	3.4	1.6	0.4	s
5.2	3.5	1.5	0.2	s
5.2	3.4	1.4	0.2	s
4.7	3.2	1.6	0.2	s
4.8	3.1	1.6	0.2	s
5.4	3.4	1.5	0.4	s
5.2	4.1	1.5	0.1	s
5.5	4.2	1.4	0.2	s
4.9	3.1	1.5	0.2	s
5	3.2	1.2	0.2	s
5.5	3.5	1.3	0.2	s
4.9	3.6	1.4	0.1	s

4.4	3	1.3	0.2	s
5.1	3.4	1.5	0.2	s
5	3.5	1.3	0.3	s
4.5	2.3	1.3	0.3	s
4.4	3.2	1.3	0.2	s
5	3.5	1.6	0.6	s
5.1	3.8	1.9	0.4	s
4.8	3	1.4	0.3	s
5.1	3.8	1.6	0.2	s
4.6	3.2	1.4	0.2	s
5.3	3.7	1.5	0.2	s
5	3.3	1.4	0.2	s
7	3.2	4.7	1.4	v
6.4	3.2	4.5	1.5	v
6.9	3.1	4.9	1.5	v
5.5	2.3	4	1.3	v
6.5	2.8	4.6	1.5	v
5.7	2.8	4.5	1.3	v
6.3	3.3	4.7	1.6	v
4.9	2.4	3.3	1	v
6.6	2.9	4.6	1.3	v
5.2	2.7	3.9	1.4	v
5	2	3.5	1	v
5.9	3	4.2	1.5	v
6	2.2	4	1	v
6.1	2.9	4.7	1.4	v
5.6	2.9	3.6	1.3	v
6.7	3.1	4.4	1.4	v
5.6	3	4.5	1.5	v
5.8	2.7	4.1	1	v
6.2	2.2	4.5	1.5	v
5.6	2.5	3.9	1.1	v
5.9	3.2	4.8	1.8	v
6.1	2.8	4	1.3	v
6.3	2.5	4.9	1.5	v
6.1	2.8	4.7	1.2	v
6.4	2.9	4.3	1.3	v
6.6	3	4.4	1.4	v
6.8	2.8	4.8	1.4	v
6.7	3	5	1.7	v
6	2.9	4.5	1.5	v

5.7	2.6	3.5	1	v
5.5	2.4	3.8	1.1	v
5.5	2.4	3.7	1	v
5.8	2.7	3.9	1.2	v
6	2.7	5.1	1.6	v
5.4	3	4.5	1.5	v
6	3.4	4.5	1.6	v
6.7	3.1	4.7	1.5	v
6.3	2.3	4.4	1.3	v
5.6	3	4.1	1.3	v
5.5	2.5	4	1.3	v
5.5	2.6	4.4	1.2	v
6.1	3	4.6	1.4	v
5.8	2.6	4	1.2	v
5	2.3	3.3	1	v
5.6	2.7	4.2	1.3	v
5.7	3	4.2	1.2	v
5.7	2.9	4.2	1.3	v
6.2	2.9	4.3	1.3	v
5.1	2.5	3	1.1	v
5.7	2.8	4.1	1.3	v
6.3	3.3	6	2.5	i
5.8	2.7	5.1	1.9	i
7.1	3	5.9	2.1	i
6.3	2.9	5.6	1.8	i
6.5	3	5.8	2.2	i
7.6	3	6.6	2.1	i
4.9	2.5	4.5	1.7	i
7.3	2.9	6.3	1.8	i
6.7	2.5	5.8	1.8	i
7.2	3.6	6.1	2.5	i
6.5	3.2	5.1	2	i
6.4	2.7	5.3	1.9	i
6.8	3	5.5	2.1	i
5.7	2.5	5	2	i
5.8	2.8	5.1	2.4	i
6.4	3.2	5.3	2.3	i
6.5	3	5.5	1.8	i
7.7	3.8	6.7	2.2	i
7.7	2.6	6.9	2.3	i
6	2.2	5	1.5	i

6.9	3.2	5.7	2.3	i
5.6	2.8	4.9	2	i
7.7	2.8	6.7	2	i
6.3	2.7	4.9	1.8	i
6.7	3.3	5.7	2.1	i
7.2	3.2	6	1.8	i
6.2	2.8	4.8	1.8	i
6.1	3	4.9	1.8	i
6.4	2.8	5.6	2.1	i
7.2	3	5.8	1.6	i
7.4	2.8	6.1	1.9	i
7.9	3.8	6.4	2	i
6.4	2.8	5.6	2.2	i
6.3	2.8	5.1	1.5	i
6.1	2.6	5.6	1.4	i
7.7	3	6.1	2.3	i
6.3	3.4	5.6	2.4	i
6.4	3.1	5.5	1.8	i
6	3	4.8	1.8	i
6.9	3.1	5.4	2.1	i
6.7	3.1	5.6	2.4	i
6.9	3.1	5.1	2.3	i
5.8	2.7	5.1	1.9	i
6.8	3.2	5.9	2.3	i
6.7	3.3	5.7	2.5	i
6.7	3	5.2	2.3	i
6.3	2.5	5	1.9	i
6.5	3	5.2	2	i
6.2	3.4	5.4	2.3	i
5.9	3	5.1	1.8	i

2.5.2 Dummy Data set:

The Dummy dataset is dataset that has a random generated one time data by relying on distance between node, these data could be taken randomly like Facebook dataset (Kaur, 2015). We generated randomly data to use it in our Methodology.

Chapter 3

Literature Review

3.1 Introduction:

Many researchers have used Fuzzy method as an artificial intelligence technique, accompanied with k-nearest algorithm to make clustering efficient in environments like WSN networks.

3.2 FLAME Clustering

Limin Fu (Limin Fu, et al., 2007) performed a Fuzzy clustering by Local Approximation of Memberships (FLAME) on DNA microarray data. They divided the approach into two main categories; firstly defining the (gene or cluster) neighbor and identifying of objects with "archetypal" called CSO (Cluster Supporting Objects) and the constructing clusters, secondly, assigning each object by its membership of its neighboring objects. The membership spreads from the CSO to all neighbor. The results showed that the FLAME has a good performance in large data sets.

3.3 Fuzzy Clustering

Wang (Wang, et al., 2006) have improved the fuzzy c-means by removing initial conditions. It is hard to make the c-means clustering fast without the initial condition so they solved this problem by applying fast global fuzzy c-means. The result showed that the fast fuzzy c-means are faster and better than global fuzzy c-means.

Torra (Torra, et al., 2005) have presented an algorithm that permits the construction of a fuzzy hierarchy by using less parameters and obtains a fuzzy partition without manipulating the fuzzy memberships of the cluster techniques. The results confirmed that the fuzzy partition of fuzzy sets needs a large computational power to compute.

Li (Li, et al., 2010) have improved the fuzzy k-means by using the k-center algorithm, where the k-center function is used to determine the initial value of k means and makes the outside distance as large as possible while the distance is kept as small as possible. At the same time, they built a binary tree for k means clusters for reducing the number of difference calculation between data points to their means. The results showed that this technique has a higher accuracy while reducing number of calculations of training data points of k-means.

Likas (Likas, et al., 2011) have presented the method of reducing computational load without changing or affecting solution quality. The proposed modification is called fast global k-means (FGK). The presented modification define a fast computed bound on the clustering error that is used in local searches of clusters. Also, FGK provide same quality solutions for reducing computational effort. The results show that the FGK has less error rates than k-means.

Hoang (Hoang, et al., 2010) have presented a centralized cluster-based protocol. Fuzzy C-means FCM clustering. This protocol creates a cluster structure that reduces the distance between sensor node to make a better cluster combination. The results show that the FCM protocol has reduced the consumption power and extended the lifetime of the network compared with k-means.

Zhou (Zhou, et al., 2015) have proposed a novel distributed K-means clustering algorithm. This algorithm makes every sensor node performs a distribution clustering by collaborating with the neighbor sensor to improve the clustering result. The results showed that the distributed K-means clustering algorithm has a good performance compared with the centralized methods.

Yadav (Yadav, et al., 2016) have improved the working efficiency of the existing efficient k-means algorithm. They used the previous iteration data in the current and the next iteration for clustering the node in the network. This step reduced the computational complexity of the k-means. The results showed that the improved k-means clustering has less time computation and better accuracy than the Enhanced k-means (EKM).

Raval (Raval, et al., 2016) have improved the k-means technique for determining the initial centroid and assigning the center data to its nearest cluster. Also, this step produces more accuracy with less time complexity than the traditional k-means.

3.4 K-nearest Clustering

Chao (Chao, et al., 2013) have used the concept of clique in the k-nearest. The clustering in this algorithm is based on gathering nodes into a cluster. If the cluster meets the requirement then it will be considered. Otherwise, the clustering will be repeated again until it meets the algorithm requirements. The result showed that the algorithm gets lower Error Rate than the k-means algorithms, which shows the advantage of the k-nearest neighbor clique clustering (KNNCC) algorithm.

3.5 link-state algorithm

Levchenko (Levchenko, et al., 2008) have proposed a new link-state routing algorithm called Approximate Link state (XL). This algorithm was designed to minimize network communication, and reduce the frequency of routing updates in the network. The results showed that the (XL) had slightly better convergence times than one second.

Ortakci (Ortakci , 2017) has presented the Parallel Particle Swarm Optimization in Data (PPSO), This method is presented for data clustering. The idea of this method is to take random initial position called swarm. The Parallel Particle Swarm Optimization is a development of the research method Particle Swarm Optimization. The results show that the Parallel Particle Swarm Optimization has a better performance in terms of execution time and the error rate than the Particle Swarm Optimization.

Despite the importance of the aforementioned studies; this study presents a different perspective. As mentioned earlier, the previous studies combined the same type of node and choose one head cluster to each cluster. However, this study will perform the clustering by using FLAME. Then the clustering will be conducted on each cluster. Finally, multiple head cluster of well-known data will be elected.

Chapter 4

Proposed Methodology

4.1 Introduction

This chapter contains a detailed description of the proposed algorithm for this research. Several algorithms that group nodes will be investigated and the one with the less error rate and best accuracy will be indicated. The idea of our proposed technique is using a FLAME algorithm in a two level hierarchy clustering. Several phases will be considered, starting with data analysis and defining a grouping node that has the same characteristics. The sections below presents a description for each stage;

4.2 System Stages

The proposed work consists of two main stages with multi step, as shown in Figure 4-1. The following section will clarify this stage and their steps.

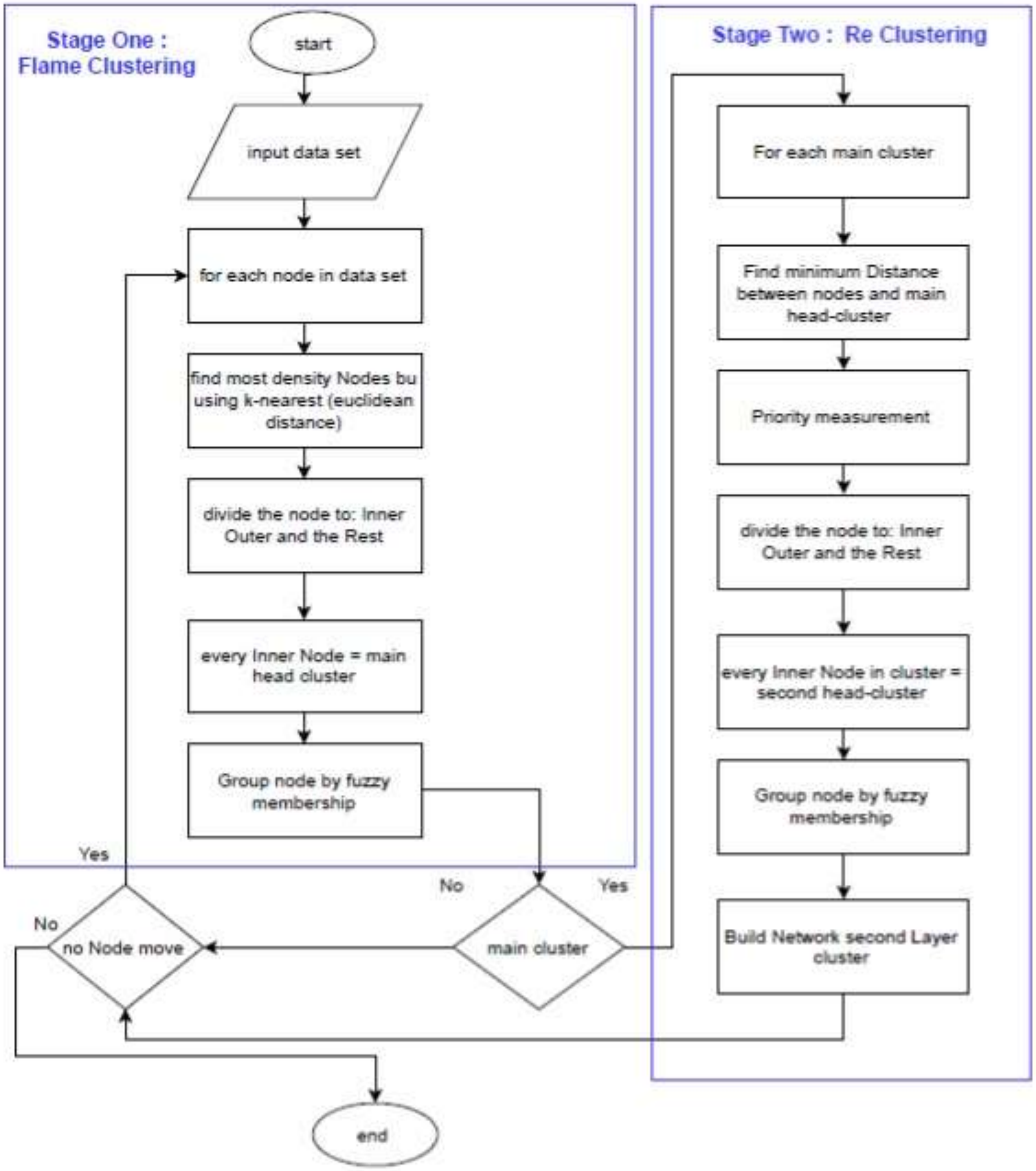

Stage One :
Flame Clustering
start
input data set
for each node in data set
find most density Nodes bu using k-nearest (euclidean distance)
divide the node to: Inner Outer and the Rest
every Inner Node = main head cluster
Group node by fuzzy membership
Yes
No
Yes
No
no Node move
main cluster
end
Stage Two : Re Clustering
For each main cluster
Find minimum Distance between nodes and main head-cluster
Priority measurement
divide the node to: Inner Outer and the Rest
every Inner Node in cluster = second head-cluster
Group node by fuzzy membership
Build Network second Layer cluster

Figure (4-1): System Stages

4.2.1 Stage One: FLAME Clustering

The FLAME clustering algorithm defines the elements of the data set by following three steps:

4.2.1.1 Step One: Extraction the data structure from the dataset

In this step, we build a neighborhood graph to connect each node to its K-Nearest Neighbors (KNN). Then estimates a density for each node based on its proximities to its KNN. The cluster is divided into three groups: Inner, Outer, and Rest. The Inner group or Cluster Supporting Object (CSO) contains the nodes that have the highest density among neighbors. The Outer group contains the nodes that are far away from the high density nodes or they have the lowest density than a predefined threshold. The remaining nodes will be under a new group called the Rest.

4.2.1.2 Step Two: Initialization fuzzy membership

The next step is to determine where does each node belongs. Each CSO has full membership to itself and represents its cluster. The rest are assigned with equal memberships to all clusters, and the outlier assigned full membership to outlier cluster. Then, each CSO will represent as a Head Cluster

4.2.1.3 Step Three: Building the cluster network by fuzzy membership

Building the cluster from fuzzy memberships will be conducted in two ways:

1. One-to-one node assignment. To assign each node to the cluster that it has the highest membership;
2. One-to-multiple node assignment. To assign each node to the cluster that has the highest membership compared to the threshold.

4.2.2 Stage Two: Re-clustering

The second stage will be as follows;

4.2.2.1 Step One: Re-extraction of the data structure from every cluster

In this step, we have to build a priority graph to connect each node to its priority value. Then calculate the distance between each node and the main head cluster.

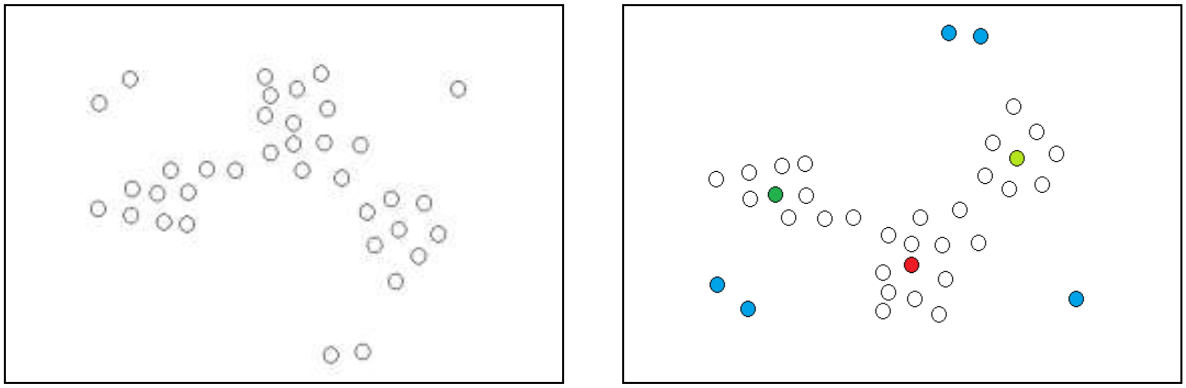
4.2.2.2 Step Two: Second initialization level by fuzzy membership

The next step, is to determine where each node belongs. each CSO represents a second head cluster in cluster. The rest are assigned with equal priority value to each cluster, and the outlier node will be assigned a full membership to outlier cluster in each cluster.

4.2.2.3 Step Three: Build the second cluster level with fuzzy membership

This step is similar to step three in stage one.

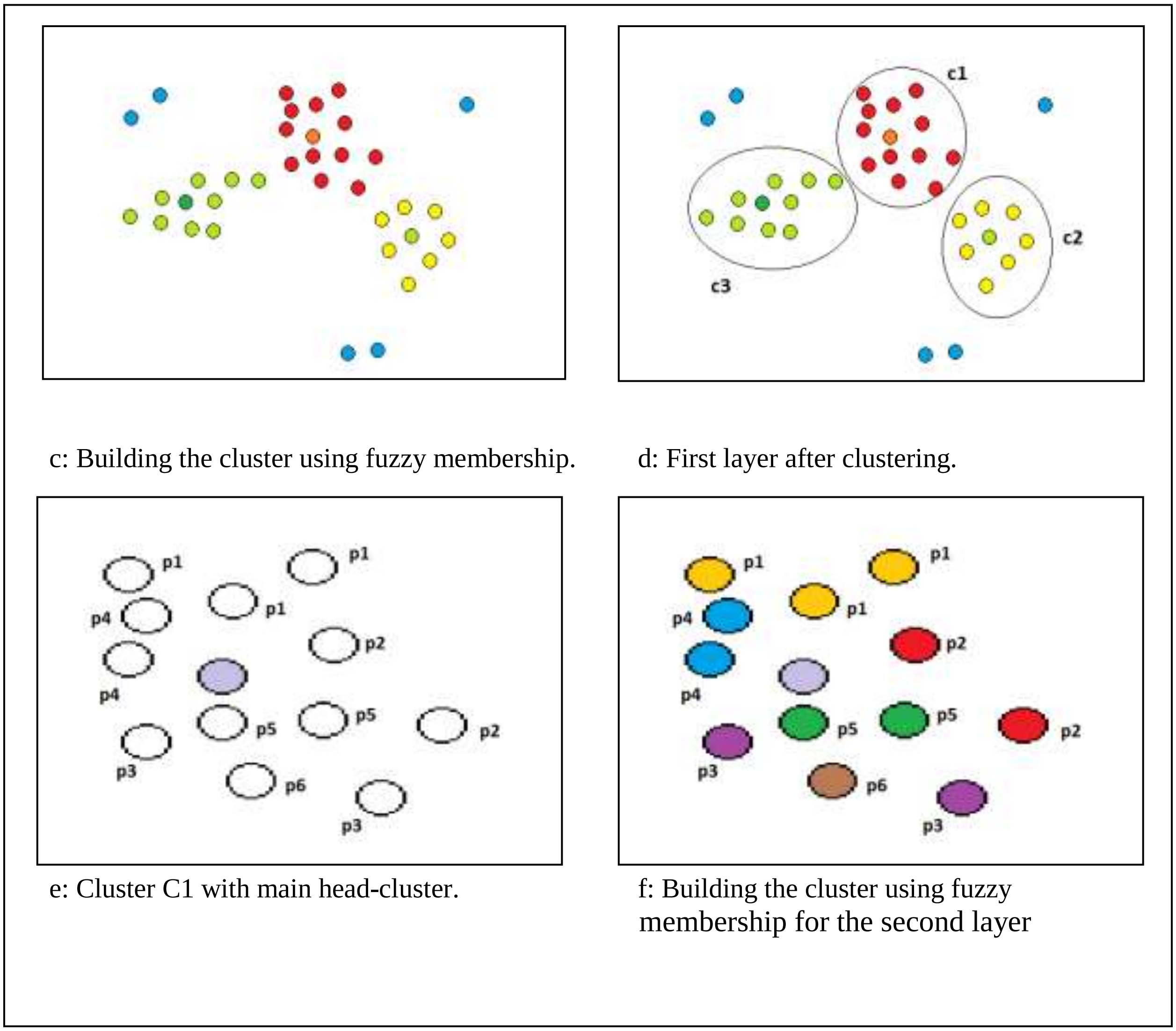
Figure 4.2 (a and b) shows an example of our proposed algorithm.



a : Some of nodes in WSN

b : Classifing the nodes to Inner, Outer, and the rest. inner node is a head cluster.

Figure 4-2 a: example of our research method



c: Building the cluster using fuzzy membership.

d: First layer after clustering.

e: Cluster C1 with main head-cluster.

f: Building the cluster using fuzzy membership for the second layer

Figure 4-2 b Example of our research method

Pseudo code :

Step 1: Extracting data structure from dataset.

Stage 1: *Start first layer clustering:*

A. Build a neighborhood graph to connect each node to its K-Nearest Neighbors (KNN).

B. Estimate the density for each node based on its proximities to its KNN.

C. Node are divided into three types:

1. Cluster Inner: node with higher density than other nodes.

2. Cluster Outliers: node with lower density than all its neighbors, and far away from all clusters.

3. The Rest: node that is located between inner and Outlier.

Stage 2: *Initialization of fuzzy membership. Each inner has full membership to itself and represent its cluster by elect it main head-cluster. The Rest are assigned with equal memberships to all clusters. The outlier assigned full membership to outlier cluster.*

Stage 3: *Build the network cluster by fuzzy membership.*

Step 2: Re-clustering.

stage 1: *Re-extraction of the data structure from every cluster*

1. build a priority graph (Matrix) to connect each node to its priority value.

2. calculate the distance between each node and the main head cluster and divide the node into (Inner, Outer, Rest).

Stage 2: *Second Initialization level of fuzzy membership*

1. determine where each node belongs:

A. Each inner represents a second head cluster in cluster.

B. The rest are assigned with equal priority value to each cluster.

C. outlier node will be assigned a full membership to outlier cluster in each cluster.

Stage 3: *Build second cluster level by fuzzy membership*

Chapter 5
Experimental Results

5.1 Introduction:

This chapter presents the result of the research method, where all experiments are tested on an Asus desktop computer with Intel core i5, 6 GB of RAM, and 2TB of Hard Disk, running on Microsoft Windows 10 and Microsoft visual studio 2017 to develop the C++ code.

5.2 Clustering Measurements:

To analyze the performance of our research method, we applied two kinds of data sets; dummy data set, and Iris data sets (UCI, 2017). The measurementes used in our experiment are Error Rate, Accuracy, Precision, Euclidean Distance and Manhattan Distance between nodes (Eltibi, et al.,2011) (Singh, et al.,2013).

5.2.1 Error Rate:

The calculation of Error rate depends on the number of misclassified node and the total number of tested node in dataset (Eltibi, et al.,2011),or false positive divided by True Positive and False Negative It is defined in the Table (5-1), as equation qualified in (1).

$$\text{Error Rate} = \frac{\text{False positive}}{\text{True Positive} + \text{False Negative}} * 100 \qquad (5\text{-}1)$$

5.2.2 Accuracy Rate:

The Accuracy was defined as the proportion of correct predictions of the size of the actual data set. To compute Accuracy, two methods were utilized (Elhamahmy, et al., 2010). The first used the number of true node positions classified and the total number of tested node in the data set (Ariel Linden, et al., 2006), as equation qualified in (2).

$$\text{Accuracy} = \frac{number\ of\ true\ node\ position}{total\ number\ of\ node} * 100 \qquad (5\text{-}2)$$

The second method was used to measure the True Positive and False Negative which is defined in the Table (5.1) (Mulak, et al., 2013). The accuracy is the percentage of test nodes that are correctly classified by the classifier, as in the equation qualified in (3).

$$Accuracy = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \qquad (5\text{-}3)$$

5.2.3 Precision:

The precision is defined as the proportion of positive nodes that is correctly classified (Elhamahmy, et al.,2010), it is defined in the Table (5-1). Precision is calculated by:

$$\text{Precision} = \frac{true\ positve}{true\ postive + false\ postive} *100 \qquad (5\text{-}4)$$

Table (5-1) Measurement Parameters

Parameter	**Definition**
True Positive (TP)	Node in same class in same cluster
False Positive (FP)	Node in different classes in same cluster
FalseNegative(FN)	Node in different classes in different cluster

5.2.4 Distance Measurements:

It includes two Distance Measurements:

5.2.4.1 Manhattan Distance:

Manhattan distance calculates the absolute differences between the coordinates of two nodes (Singh, et al., 2013), as the equation qualified in (5).

$$\text{Manhattan distance} = |\ (\ X_1 - X\ ) + (\ Y_1 - Y\ )\ | \qquad (5\text{-}5)$$

5.2.5.2 Euclidean Distance:

Euclidean distance calculates the root of square difference between the coordinates of two nodes (Singh, et al., 2013), Euclidean is calculated by:

$$\text{Euclidean distance} = \sqrt{(x_1 - x\ )^2 + (y_1 - y\ )^2} \qquad (5\text{-}6)$$

5.3 Dataset:

The research method was tested on Iris's flower data set or Fisher's Iris data set from the UCI datasets (UCI, 2017). It was introduced by the biologist Ronald Fisher in his paper, which applied multiple measurements in taxonomic problems (Fisher, 1936), Iris's data set became a test case for classification techniques, which included three sets of 150 instances. The three sets are called setosa, versicolor, and virginica. Each set has four numeric attributes, Sepal length, Sepal width, Petal length, and Petal width.

5.4 Experimental Result:

The experiment was divided into Four experiments. The first one was carried out for the internal result. The second was for Accuracy, Error rate, and precision between this research and other research using iris data set. The third experiment was done for distance measurement with other methods, and the fourth was for Accuracy, Error rate, and precision between this research approach and other methods by using dummy data set.

5.4.1 The First Experiment: Internal Result

A distinct number of instances of the Iris data set has been applied in the experiment. A 95 instance of data set has been applied to this approach. Table (5.2) shows the result for different instances applied to our algorithm 95, 98, 99, 109, 117, and 122 node, where the best value for the Accuracy, precision, Error rate was received when this research method was tested on 95 instances. the reason that the table has 122 as a maximum number of nodes that we use three attributes of iris data set of x, and y coordinates and one of p priority.

Table (5-2) Error rate, Accuracy, precision in iris data set applying in this research.

		Inner			Outer		
dataset	Number of tested node	Number of inner cluster	Error rate	Accuracy	Number of outer cluster	Error rate	Accuracy
Iris Data set	95	3	9.3%	90.7%	2	93%	7 %
	98	3	21.5%	78.5%	5	92.6%	7.4 %
	99	3	20.6%	79.4%	4	95.2%	4.8 %
	109	2	41.6%	59.4%	6	98.9%	1.1%
	117	4	33%	67%	8	83.4%	16.6 %
	122	3	33%	67%	14	75%	25 %

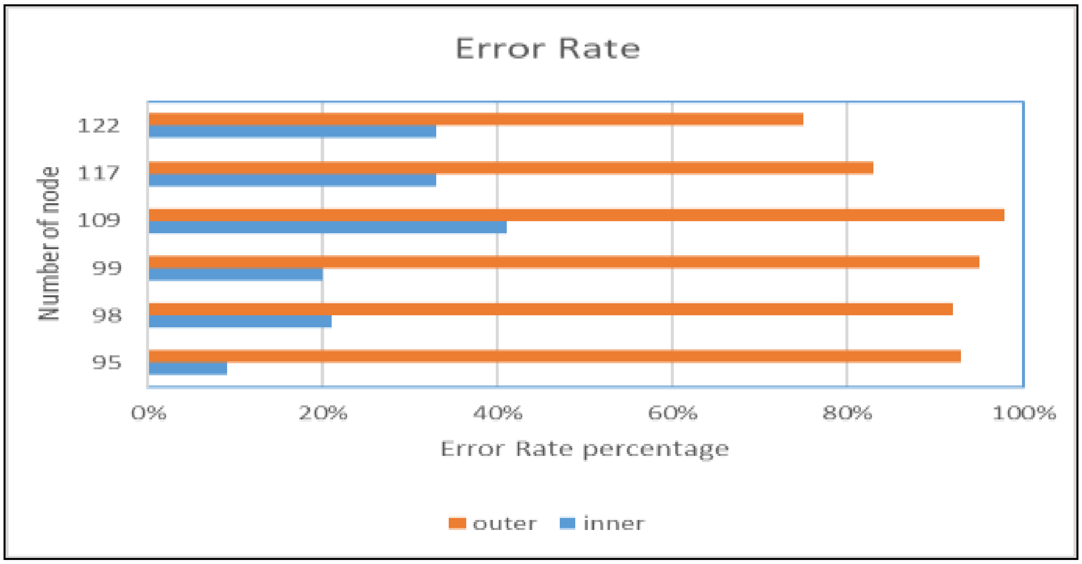


Figure (5-1): Error rate comparison for internal result

Figure (5-1) shows the Error rate comparison for internal result. For more improvements, there are multiple values for weight used in this work. The optimal weight that is given by the inner cluster; Here strength is fond of FLAME By utilizing the inner cluster. The best value of inner cluster is for 95 instances; it had 9% Error rate; the outer cluster has poor value of 93% Error rates.

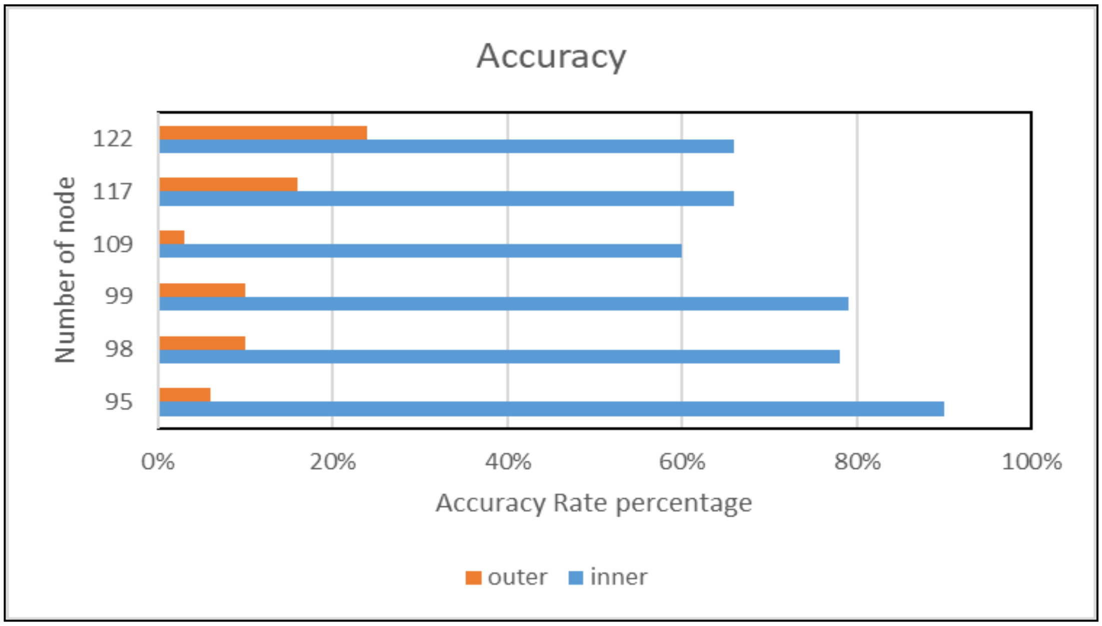


Figure (5-2): Accuracy comparison for internal result

As presented in the tables (5-2), Figure (5-2) shows the Accuracy comparison for internal result. As we mentioned before the 95 instances is the best value for Clustering, the inner cluster had 90% accuracy, and also the outer cluster had a poor value of 6% accuracy.

5.4.2 The Second Experiment: Accuracy, Error Rate, precision using IRIS dataset

The results were compared with Kohonen SOM, K-means (Kaur, et al., 2015), KMS, NJW, Self-Tuning, Transitive Distance Clustering with K-Means Duality (Yu, et al., 2014), Decision Stump, Multilayer Perceptron, Naïve Bayes, Multi Class Classifier (Patel, et al.,2014), and EKM (Priya, et al.,2012). First, it was passed through the initial stage, which applied the FLAME algorithm by dividing the node into three groups; the inner, outer, and rest. Then elects a head cluster for each cluster from the inner group. This clustering is called layer one. Moving to the second stage for re-clustering; each cluster from layer one by the measurement priority, then the node was divided into three groups the inner, outer, and rest. After that, elects a head cluster for each cluster from the minimum distance between the cluster and master head cluster in layer one from the inner group in layer two.

5.4.2.1 Error Rate

The results show that this research method has 0.09 (9%) Error Rate, it is equal to the NJW method and has the better value than the other methods as listed in table (5-3).

Table (5-3) Error Rate using Iris data set

Studies	Error rate
Kohonen SOM (Kaur , 2015)	33%
K-means (Kaur , 2015)	35%
KMS (Yu, 2014)	11%
NJW (Yu, 2014)	9%
Self-Tuning (Yu, 2014)	15%
Transitive Distance Clustering with K-Means Duality (Yu, 2014)	90%
Fuzzy C-Means Algorithm (Kumar, etal, 2017)	12%
Improved Artificial Bee Colony and Fuzzy C-Means Algorithm (Kumar, etal, 2017)	10%
Our research method	9%

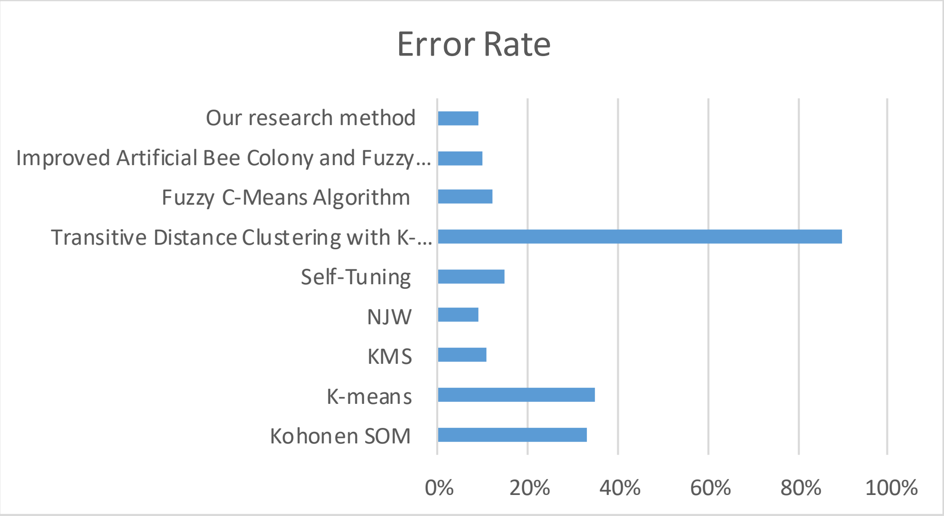


Figure (5-3): Error Rate comparison

As presented in the table (5-3) and figure (5-3), the result show that the comparison of seven methods. This research with NJW have the least value of Error rate (0.9) than the other methods.

5.4.2.2 Precision

We increased number of instances to 117 and 122. The result show in table (5-4) that this research has 0.84 (84%) Precision, it has better results than Decision Stump, and Grid-based.

Table (5-4) Precision using Iris data set

Studies	precision
Decision Stump (Patel,2014)	50 %
Multilayer Perceptron (Patel,2014)	97 %
Naïve Bayes (Patel,2014)	96 %
Multi Class Classifier (Patel,2014)	96 %
Grid-based (Kim, etal, 2017)	49 %
Our approach	84 %

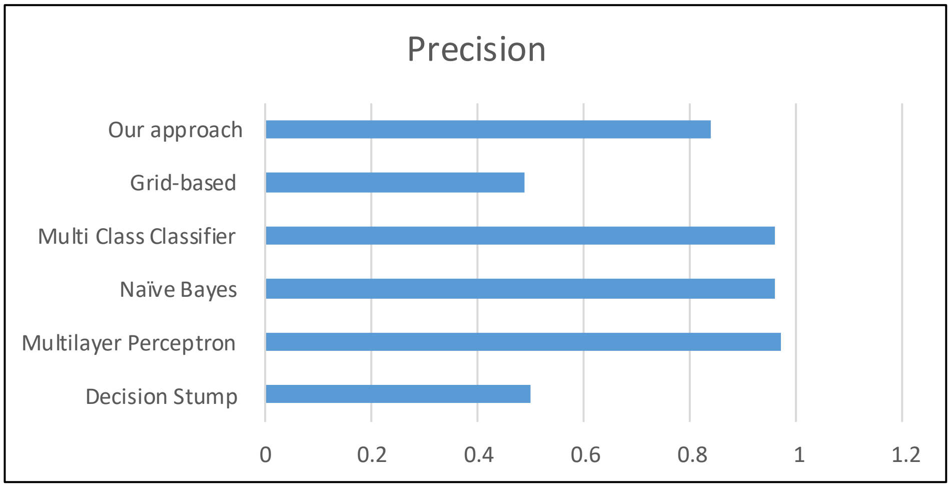


Figure (5-4): Precision comparison

Figure (5-4) shows the result for the precision; it shows that this research has 0.84 which is better than Decision Stump (0.50), Grid-based (0.49). Moreover, it shows that the Multilayer Perceptron has the best value (0.97), our method had the fourth place in the

Figure (5-4). Because the multilayer perceptron, naïve bayes, and multi class classifier has used the fuzzy clustring more than ones to filtering thier result.

5.4.2.3 Accuracy

The result, in the second experiment used 95 instances. The result is shown in table (5-5) . The Accuracy of this research is better than the other method that is shown in table (5-5). Figure (5-4) shows the différences between the three methods.

Table (5-5) : Accuracy

Studies	Accuracy
Improved K-means (Raval,2016)	90%
Traditional K- Means (Raval,2016)	84%
Enhanced k-means (Yadav,2016)	88%
Hierarchical (Yadav,2016)	66%
Naive Sharding (Mayo, 2017)	88%
Our approach	90%

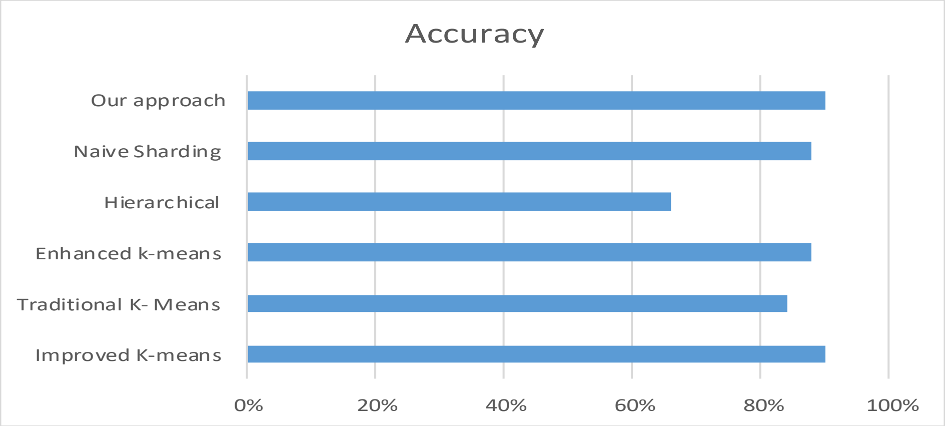


Figure (5-5): Accuracy comparison

Figure (5-5) shows that the result of the accuracy that was applied in six methods. As a result, our method has the best value (90%) shared with improved k-means by Raval (Raval, et al,2016). The enhanced k-means has a close result of (88%) shared with Naive

Sharding then Traditional K- Means has (84%). The other method has far and poor results compared with the first five method.

5.4.3 The Third Experiment: Distance Measurement

In the third experiment, the dummy data set was used. The data is generated one time randomly. The distance measurement is used to calculate the distance between the principal node, the sub main node and the other nodes in each cluster. Table (5-6) shows the result for 25, 50, 75 nodes that are applied in two algorithms, k-means and this research method.

Table (5-6) Distance Measurement

Data set	Clustering method	Number of node	Manhattan distance	Euclidean distance
Dummy Data set	k-means	25	Cluster 1 = 24	Cluster 1 = 17
			Cluster 2 = 15	Cluster 2 = 11
			Cluster 3 = 11	Cluster 3 = 9
			Sum = 50	Sum = 37
		50	Cluster 1 = 47	Cluster 1 = 30
			Cluster 2 =34	Cluster 2 =24
			Cluster 3 = 17	Cluster 3 = 13
			Cluster 4 = 20	Cluster 4 = 15
			Sum = 118	Sum = 82
		75	Cluster 1 = 71	Cluster 1 = 47
			Cluster 2 = 49	Cluster 2 = 35
			Cluster 3 = 28	Cluster 3 = 22
			Cluster 4 = 20	Cluster 4 = 15
			Sum = 168	Sum = 119
Dummy Data set	Our approach	25	Cluster 1 = 24	Cluster 1 = 16
			Cluster 2 = 12	Cluster 2 = 8
			Cluster 3 = 15	Cluster 3 = 12
			Sum = 51	Sum = 36
		50	Cluster 1 = 35	Cluster 1 = 26
			Cluster 2 =35	Cluster 2 =22
			Cluster 3 = 17	Cluster 3 = 11
			Cluster 4 = 27	Cluster 4 = 18
			Sum = 114	Sum = 77
		75	Cluster 1 = 59	Cluster 1 = 42
			Cluster 2 = 47	Cluster 2 = 30
			Cluster 3 = 32	Cluster 3 = 23
			Cluster 4 = 27	Cluster 4 = 18
			Sum = 165	Sum = 113

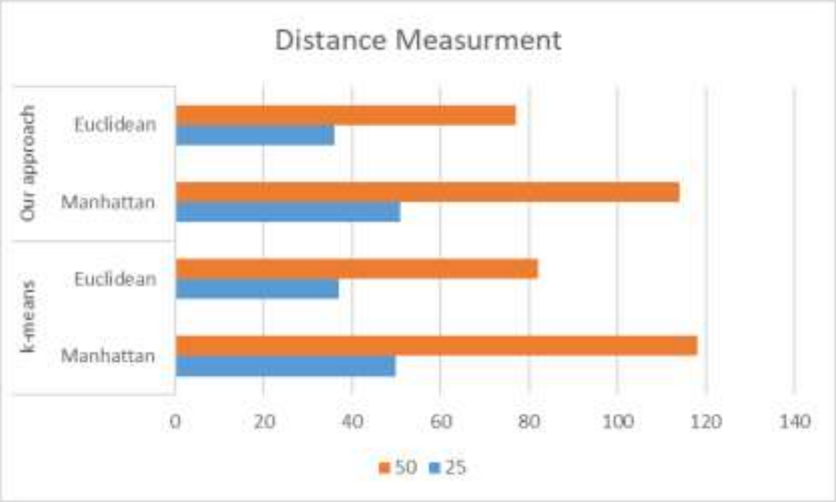


Figure (5-6): Distance Measurement comparison for 50 and 25 node

As presented in table (5-6) and Figure (5-6), the result for the distance measurements between all nodes and head-clusters. The Figure (5-6) shows the distance between 25 nodes and its head clusters applied on two different measurements; Manhattan and Euclidean distance, then the number of nodes is increased to 50. As a result, this research has the optimum value on Euclidean on both 25 and 50 nodes than the value on k-means. Also the best value of Manhattan in 50 nodes, but, in 25 nodes the value of k-means is better than the value in this research.

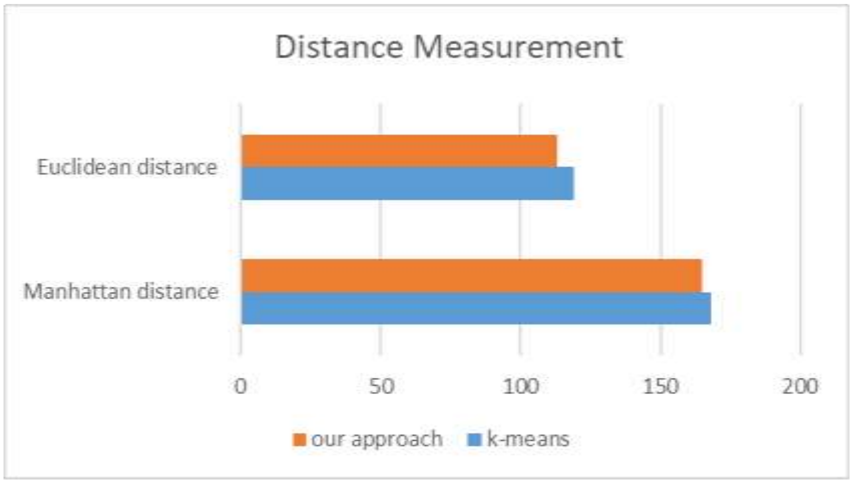


Figure (5-7): Distance Measurement comparison for 70 node

As presented in table (5-6), Figure (5-7) shows the result for distance between 70 node, the result show that our research method distance result using both measurement is better than distance result for the k-means.

5.4.4 The Fourth Experiment: Using Random Data

In the final experiment, the dummy data set was used, and data is generated one time randomly. These data were used to compare approach with others that used random data set.

The table (5-7) shows the comparison of the research method Error rate and Accuracy with kohonen som and k-means methods Error rate and Accuracy.

Table (5-7) Error rate and Accuracy comparison using random data

Studies	Data set	Error rate	Accuracy
Kohonen SOM (Kaur , 2015)	Random data	33%	33%
K-means (Kaur , 2015)	Random data	35%	35%
Our approach	Random data	9%	91%

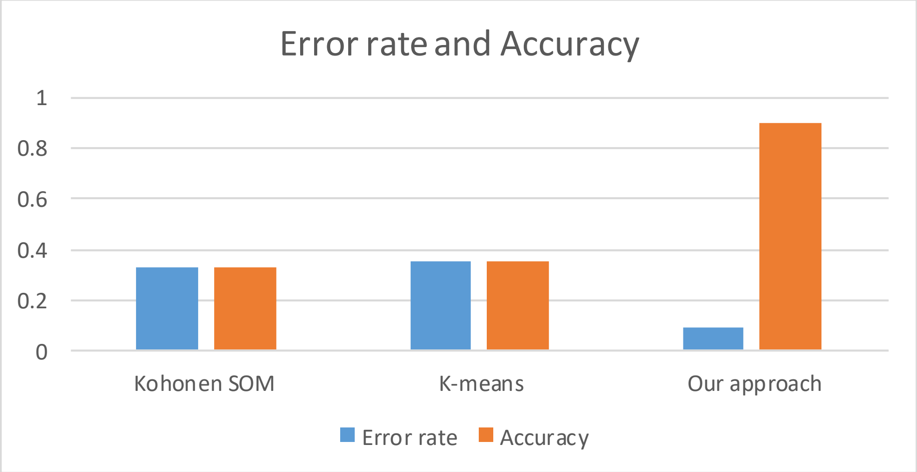


Figure (5-8) :Error rate and Accuracy comparison table (5-7) and Figure (5-8) shows that our approach has the lower value of Error rate (0.9) and best value of Accuracy.

Chapter 6
Conclusion and Future Work

6.1 Conclusion:

Two Levels clustering hierarchies using fuzzy clustering by Local Approximation of Memberships has been presented.

The main outcomes of this study are:

1) Choosing the most density nodes in the network based on its proximates to its k-nearest neighbors, the nodes are divided into three types: Inner, Outer, and the Rest, then Initialization of fuzzy membership by the location of the node.

2) Each Inner node in the network is set to be a main head-cluster subsequently build the network cluster by fuzzy membership. The network is built and clustered, in each cluster build a priority graph to connect each node to its priority value, after that calculate the distance between each node and leading head cluster in the cluster then divided the node to Inner, Outer, and the Rest by the priority value.

3) The second inner nodes in each cluster are set to be a second-head cluster, and its function is managing the network communication between its cluster and other clusters in the same main cluster.

We seek to guarantee the scalability, and adaptively, that means when the number of the node increases the network will not be overloaded or has a poor performance.

The experiment result shows that our research method can give more consistency and accuracy in the distribution of the node with the lowest possible percentage of error rate than the other methods in terms of used measures that included in this research and achieves best ranking among most methods. Our result proves that is our research methods have been better clustering partition than the k-means, fuzzy c-means and some other method, moreover; these result leads us to the conclusion that the Two Level clustering hierarchies using fuzzy clustering by Local Approximation of Memberships is a robust and good technique for data mining (clustering).

6.2 Future Work:

Future research is to evaluating the changes affecting the network. When a malfunction or damage in the main head cluster is recorded, the nearest and most effective second head cluster will be assigned to act as the main head cluster. Also, changes to main head cluster node can be performed based on energy consumption.

References

1. Anoop Kumar Jain and Satyam Maheswari, Survey of Recent Clustering Techniques in Data Mining, International Archive of Applied Sciences and Technology, Vol 3 June 2012: 68 – 75, 2012.

2. Archana Singh, Avantika Yadav, Ajay Rana, K-means with Three different Distance Metrics, International Journal of Computer Applications (0975 – 8887) Volume 67– No.10, April 2013.

3. Ariel Linden, Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis, Journal of Evaluation in Clinical Practice, 2006.

4. Aristidis Likasa, Nikos Vlassisb, JakobJ. Verbeekb, The global k-means clustering algorithm, Pattern Recognition 36 (2003) 451 – 461, 2003.
5. Asmita Yadav and Sandeep Kumar Singh, An Improved K-Means Clustering Algorithm, International Journal of Computing Academic Research (IJCAR) ISSN 2305-9184, Volume 5, Number 2 (April 2016), pp.88-103. 2016
6. Chao Qu, Ruifen Yuan, Xiaorui Wei, KNNCC: An algorithm for k-nearest neighbor clique clustering, Machine Learning and Cybernetics (ICMLC), 2013 International Conference, 14-17 July 2013.
7. Cover. Estimation by the nearest neighbor rule. IEEE Transactions on Information Theory,14:21–27, January 1968

8. Dajin Wang, An Energy-efficient Clusterhead Assignment Scheme for Hierarchical Wireless Sensor Networks, Int J Wireless Inf Networks, Published online: 15:61–71, 2008.
9. Dunn, C, A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Cybernetics and Systems, 3: 3, 32 — 57, 1973.

10. Edgar Anderson, The species problem in Iris, Annals of the Missouri Botanical Garden. 23 (3): 457–509, 1936.

11. Hoang, DC and Kumar, R and Panda , SK , "A Robust Harmony Search Algorithm based Clustering Protocol for Wireless Sensor Networks", IEEE International Conference on Communications Workshops (ICC), 2010.

12. Ida Moghimipour1 & Malihe Ebrahimpour, Comparing Decision Tree Method Over Three Data Mining Software, International Journal of Statistics and Probability; Vol. 3, No. 3; 2014.2014

13. Iris data set available at:
https://archive.ics.uci.edu/ml/datasets/iris

14. JAMES C. BEZDEK, ROBERT EHRLICH and WILLIAM FULL, FCM: THE FUZZY c-MEANS CLUSTERING ALGORITHM, Journal of Parallel and Distributed Computing, Computers & Geosciences Vol. 10, No. 2-3, pp. 191-203, 1984.

15. Jaskaranjit Kaur and Gurpreet Singh, Review of Error Rate and Computation Time of Clustering Algorithms on Social Networking Sites, International Journal of Computer Applications (0975 – 8887)Volume 113 – No. 8, March 2015.

16. J. B. Mac Queen, Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297, 1967.

17. Jin Zhou, Philip, Chen Long Chen, A distributed K-means clustering algorithm in wireless sensor networks, i nformative and Cybernetics for Computational Social Systems (ICCSS), 2015 International Conference, 13-15 Aug. 2015

18. Kirill Levchenko, Geoffrey Voelker, Ramamohan Paturi, and Stefan Savage, An Efficient Network Routing Algorithm, CCR October 2008. 2008

19. Lotfi .Zadeh, Fuzzy Sets, national science fundation, no 64-44, Gp-2413, 1965.

20. Limin Fu and Enzo Medico, FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data, BMC Bioinformatics 2007 8:3, 2007.

21. Matthew Mayo, KDnuggets. Toward Increased k-means Clustering Efficiency with the Naive Sharding Centroid Initialization Method available at: https://www.kdnuggets.com/2017/03/naive-sharding-centroid-initialization-method.html

22. Mohammad Elhamahmy, Hesham Elmahdy and Imane Saroit, A New Approach for Evaluating Intrusion Detection System, CiiT International Journal of Artificial Intelligent Systems and Machine Learning, Vol 2, No 11, November 2010.

23. Mohammad Eltibi, Wesam Ashour, Initializing K-Means Clustering Algorithm using Statistical Information, International Journal of Computer Applications (0975 – 8887)Volume 29– No.7, September 2011.

24. Ronald Fisher, "The use of multiple measurements in taxonomic problems, Annals of Eugenics. 7 (2): 179–188. doi:10.1111/j.1469-1809.1936.

25. Shijin Kumar, Dharun VS2, Combination of fuzzy c-means clustering and texture pattern matrix for brain MRI segmentation, Biomedical Research 2017; 28 (5): 2046-2049. 2017

26. Subhagata Chattopadhyay, A COMPARATIVE STUDY OF FUZZY C-MEANS ALGORITHM AND ENTROPY-BASED FUZZY CLUSTERING ALGORITHMS, Computing and Informatics, Vol. 30, 2011, 701–720,2011.

27. Taoying Li, Yan Chen, Xiangwei Mu, Ming Yang, An improved fuzzy k-means clustering with k-center initialization, Advanced Computational Intelligence (IWACI), 2010 Third International Workshop, 2010

28. Unnati Raval, Chaita Jani, Implementing & Improvisation of K-means Clustering Algorithm, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016, pg. 191-203. 2016

29. Vicenç Torra, Fuzzy c-means for Fuzzy Hierarchical Clustering, The 14th IEEE International Conference, DOI: 10.1109/FUZZY.2005.1452470, 2005.

30. Yasin ORTAKCI, Parallel Particle Swarm Optimization in Data Clustering, Big Data (Big Data), IEEE International Conference, 29 Oct.-1 Nov. 2017.

31. Zhiding Yu, Chunjing Xu, Deyu Meng, Zhuo Hui, Fanyi Xiao, Wenbo Liu, Jianzhuang Liu, Transitive Distance Clustering with K-Means Duality, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference 23-28 June 2014. 2014