Clustering Based on Neural Networks in Wireless Sensor Networks

Duggirala Vijitendra, Harshit Agrawal, Sharad Tripath
Computer Science and Engineering,
National Institute of Technology, Tiruchirappalli

Abstract-
Wireless Sensor Networks (WSNs) ordinarily incorporate numerous constrained sensors to screen their environmental factors, gather information, and move it to distant workers for additional handling. Despite the fact that WSNs are viewed as very pliable ad-hoc networks, networking them have been a principal challenge given the sending size and the related quality concerns, for example, adaptability, dependability, etc .The topology of handling WSNs is viewed as a practical strategy to address these worries. Clustering is the most notable topology executed in WSNs, that gathers nodes to oversee them or potentially execute different tasks in an appropriate way. LEACH is the primary clustering directing convention which is demonstrated to be better compared with other such algorithms. Despite the fact that clustering methods are fundamentally known to improve energy utilization, there are different quality-driven goals that can be acknowledged through clustering. We have focused also on other clustering techniques using Unsupervised Machine Learning algorithms and compared energy efficiency with traditional algorithms. We review all the techniques based on energy metric derived from the distance and assigned weights based on the remaining energy of the nodes, giving exceptionally valuable experiences into the plan of clustering procedures in WSNs.

I. Introduction

A wireless sensor organization (WSN) is a wireless organization that comprises disseminated free nodes utilizing sensors to screen the physical or natural conditions, for example, temperature, sound, vibration, pressure, movement, at various areas. The improvement of wireless sensor networks was at first advanced by military applications like landmark observation. As of now, wireless sensor networks are utilized in numerous non-military personnel application regions, including climate, medical care applications, home computerization, and traffic signals. The WSNs can contain many to a few large numbers of sensors where every single hub is associated with one another Sensor hub is associated with a focal spot called a base station that gives an association with the world.

Wireless Sensor Networks (WSNs) consist of a few sensors that are called nodes and Centralized nodes called Base stations.

- **Sensor Node**
  A sensor is a tiny electronic device that is used to sense the information and process them by passing this information to the base station. It is used to measure things like the change in physical environmental parameters such as temperature, pressure, humidity, sound and the change in the health parameter of people such as heartbeat and blood pressure. A sensor node ideally must be small in size, consume very low energy, operate independently and has a very high volume thickness and is equipped for adjusting to the climate in which it is found and working without relying on anyone.

- **Base Station**
  Each distinctive wireless sensor network is associated with a base station. It comprises a chip, reception apparatus, radio board and USB interface board. For correspondence with wireless sensor nodes, the Base station is prearranged with low-power network organizing programming. All the sensor nodes handover their information to the base station in one way or the other which is imperative to the functioning of the network. Further, this information is investigated at the base station for handling and dynamic propagation. During the organization of the base station in the sensor network including the sensor nodes, energy preservation and unwavering quality issues have to be dealt with. For the most part base stations are expected to be static in nature yet in certain situations, they are thought to be portable to gather information from sensor nodes.

- **Clustering**
  Clustering is a basic mission in Wireless Sensor Networks for energy effectiveness and organization consistency. Clustering in wireless sensor networks is notable and being used for quite a while. Presently clustering is advancing for managing issues like organization lifetime and energy. Clustering in sensor nodes is significant to tackle numerous issues like adaptability, energy and lifetime issues of sensor networks. Clustering calculations wilderness the correspondence in a nearby space and send just important data to the remainder of the organization through the sending nodes (passage nodes). A gathering of nodes structure a cluster and the nearby connections between cluster individuals are controlled through a cluster head (CH). Cluster individuals for the most part speak with the cluster head and the gathered information is totalled and melded by the cluster head to ration energy.

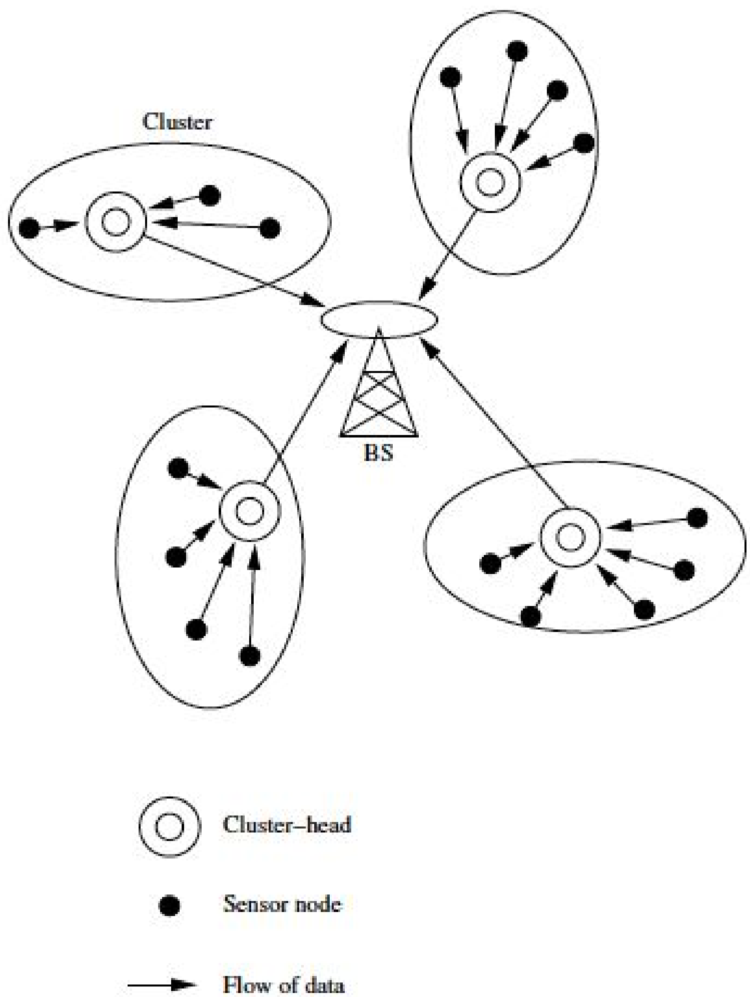
The cluster heads can likewise shape another layer of clusters among themselves prior to arriving at the sink.

- **Cluster Head**

  In the wireless organization, each cluster has a head node, which is known as the cluster head (CH) and ordinarily plays out the uncommon errands and collection, and a few regular sensor nodes (SN) as individuals. A significant test in WSNs is to choose the cluster heads:-

  1. Residual Energy
  2. Number of Neighbours

A pictorial representation of how a WSN looks like



II. Related Work

There have been many clustering algorithms proposed for WSNs.
They can be divided into four categories :

1. *Heuristic* Clustering
   They try to form clusters based on a specific metric in a tree fashion until some cut off is provided.
   Example: LCA (life cycle assessment), MAX-MIN D Cluster Algorithm, etc.

2. *Weighted* Schemes
   Like the Water Cycle Algorithm (WCA), which is a relatively new algorithm that tries to find a long-lasting architecture and reconfigures only when absolutely necessary.

3. *Hierarchical* Algorithms
   This is when clusters are merged iteratively until all the elements belong to one single cluster. It is the most straightforward and popular among all the other categories.
   Example: LEACH protocol (discussed below), EECS, TL-LEACH(two-levelwo level LEACH), HEED (hybrid energy-efficient distributed clustering), etc.

4. *Grid Schemes*
   Like PEGASIS which stands for Power-Efficient Gathering in Sensor Information Systems. These algorithms try to form cluster grids based on the cluster heads for more efficient queries and data transfer.

We discuss one of the popular hierarchical algorithms i.e LEACH protocol. LEACH is an abbreviation for Low Energy Adaptive Clustering Hierarchy. It is a TDMA (Time Division Multiple Access) based MAC (Media Access Control) protocol. Leach Protocol is application-specific and tries to reduce the energy dissipation among all nodes, thus improving the network lifetime significantly.
Assumptions made: Each node has a radio that is powerful enough to directly arrive at either a cluster head or the base station. Also assumes that the energy utilization of each node with respect to the overall energy of the network is homogenous, which is seldom the case in real-life situations.
A simple inference is that using the radio for longer durations of time for larger distances will drain the node's energy rapidly. LEACH proposes to randomly select nodes for cluster heads after every iteration. It is a probability-based model, where the selection is semi-random and also based on the percentage of maximum cluster heads and the last

time a particular node was made cluster head. By doing so, the energy dissipation is spread across the network instead of draining only a few cluster heads.

Algorithm:
Each round is split into two phases

1. *Setup Phase*
   - Every sensor node selects a randomly generated number between 0 and 1.
   - If this is lower than the threshold for node n, the sensor node becomes a cluster-head.
   - The threshold is calculated as given below, where P is the maximum percentage of nodes that can be cluster-heads (decided beforehand), r is the current iteration/round#, and G is the set of nodes that have not yet been cluster-heads in the past 1/P iterations.

$$T(n) = \begin{cases} \dfrac{P}{1 - P \times \left(r \ mod \ \frac{1}{P}\right)} & if \ (n \epsilon \ G) \\ 0 & else \end{cases}$$

2. *Steady Phase*
   - In this phase, the actual data transmission between the nodes and the cluster heads and between the cluster heads and the Base Station takes place.
   - This transfer is based on the TDMA schedule.
   - The cluster heads that receive information from the nodes perform data fusion/aggregation using it's local processing resources. and only send the relevant compressed information further to the Base Station.

Generic Analysis:

	Classification	Mobility	Power usage	Scalability
LCA	Heuristics	CH mobile	Max	Good
WCA	Weighted	CH fixed	Limited	Limited
LEACH	Hierarchical	BS fixed	Max	Good
PEGASIS	Grid	BS fixed	Max	Good

Source - A Survey on Clustering Techniques for Wireless Sensor Network, International Journal of Research in Computer Science

III. Overall Inference and Motivation

Need for Clustering

The number of nodes in a sensor organization can be significant degrees bigger than the number of nodes in an ad hoc network. Sensor nodes are more inclined to failure and energy loss, and their battery sources are generally not replaceable or rechargeable. Sensor nodes might not have exceptional global identifiers, so novel addressing isn't generally possible in sensor organizations. Most routing rules utilized in ad hoc networks like AODV, DSR and DSDV can't be straightforwardly ported to sensor networks on account of constraints in memory, force, and processing capacities in the sensor nodes and the non-adaptable nature of the conventions. A significant component of sensor networks is information combination/accumulation, whereby the sensor nodes total the nearby data prior to handing-off. The primary objectives of the information combination are to diminish transmission capacity utilization, media access deferral, and force utilization for correspondence.

Need for Machine Learning Algorithms in WSN

Having discussed the need for clustering, we leverage the numerous clustering algorithms in the realm of machine learning and neural networks, given their superior advantage in automation, minimal human intervention, diversified scope and continuously evolving attribute.

IV. Existing Work

Machine Learning is classified into two categories:

1. *Supervised* Learning

   This is when both the input and the expected output is fed to the algorithm in the form of a dataset, and the model trains itself to fit them with some form of error as a performance metric.
2. *Unsupervised* Learning

   The dataset in an unsupervised learning algorithm doesn't involve an output variable. This means that the model will itself have to recognise and form patterns given the data.

   Clustering, Segmentation, etc are examples of unsupervised algorithms.

Our main goal is to cluster the sensor nodes of a Wireless Sensor Network. We discuss some of the famous clustering algorithms in machine learning by considering the nodes to be "points" in an N-dimensional space with the Euclidean distance as the distance metric between points.

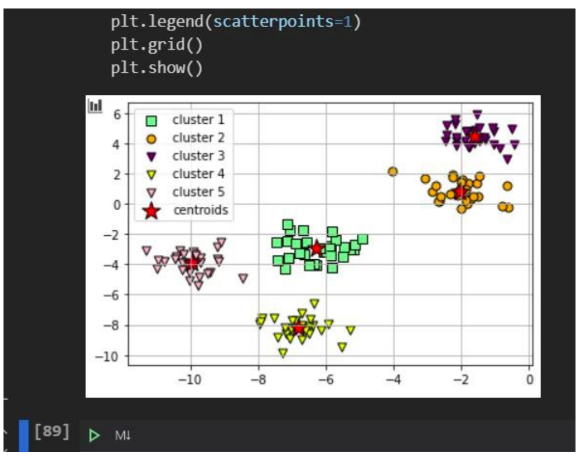
- **K - MEANS Clustering**

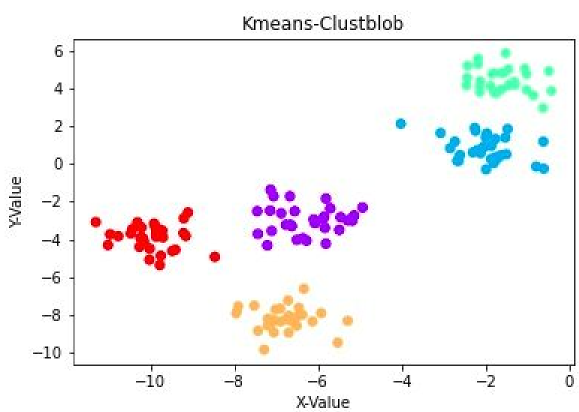
  It is a repetitive algorithm that tries to segment the dataset into K pre-determined non-overlapping clusters that are unique to each other. Each point is assigned only to one of the K clusters. The algorithm tries to ensure that within a cluster, the points are as similar to each other as possible. The algorithm also tries to maximise the inter-cluster distance

  Algorithm:
  1. Define the number of clusters K
  2. Rearrange the dataset and choose K distinct points arbitrarily.
  3. These chosen points are the cluster centroids for the initial iteration.
  4. Continue the iterations until it converges i.e until there is no change in the cluster centroids.
  5. In each iteration, calculate the distance between each point and all cluster heads. Once assigned to the minimum cluster head, recompute the cluster head by taking the new centroid.

  To find the optimal K value, we make use of the elbow method.

  Simulation :



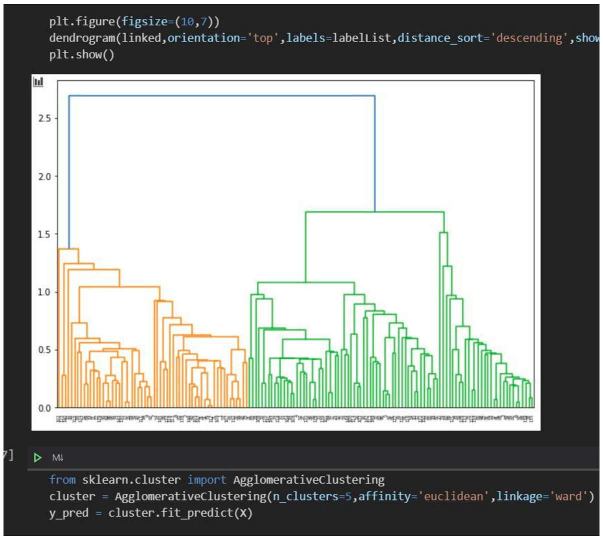


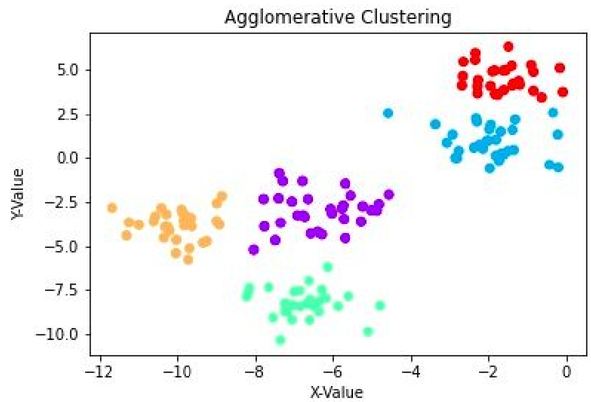
- **Agglomerative Hierarchical Clustering**

  In this algorithm, each individual data point is assumed to be a single cluster. At each iteration, similar clusters are merged until there is only one cluster.

  Algorithm:
  1. Calculate the proximity matrix
  2. Merge the two closest clusters and recalculate the proximity matrix
  3. Do Step-2, until there is only one cluster remaining

Simulation:





The entire process can be visualised using the dendrogram which is a tree-like structure that records the progression of merges. Using the dendrogram, one can determine where to split.

V. Proposed Work

We propose two methodologies to cluster sensor nodes.
One is a *Kohonen Neural Network*, which is significantly better than traditional clustering algorithms and is relatively new. Secondly, we propose to introduce a *weighted Fuzzy C Means clustering algorithm* in the setup phase of the LEACH protocol.

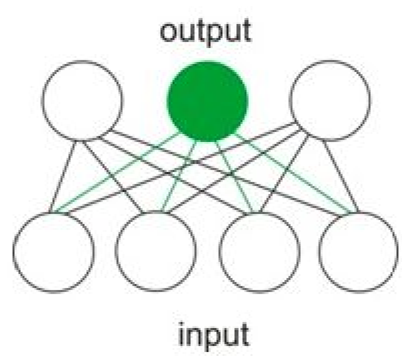
Kohonen Neural Network:
It is a *self-organizing map* (SOM) which is an unsupervised algorithm for analysis and visualisation of high dimensional data. The network has two layers, input and output. The number of nodes in the output layer is equal to the number of clusters. The number of nodes in the input layer is the dimensionality of the dataset.

The advantages of a Kohonen Neural Network are:

- Easy interpretation of data mapping
- Can handle large and complex datasets

Few drawbacks are also listed below:

- Computationally expensive
- Results cannot be explained intuitively



Algorithm :

1. Initialise each node's weight to a random value
2. Compute the distance between the input vector and the weight vector
3. Track the node that produces the smallest distance.
4. Find the overall Best Matching Unit (BMU), i.e. the node with the smallest distance from all calculated ones.
5. Update the weight vector of the BMU node by adding a fraction of the difference between the input vector x and the weight w of the neuron.
6. Repeat this whole iteration until reaching the chosen iteration limit

The base paper proposes to optimise the selection of cluster heads in the setup phase of LEACH protocol using Machine Learning. We implement a Fuzzy C Means Clustering to calculate the new cluster heads using WEIGHTED Clustering where the weight factor is the residual energy of each node and the distance factor is the spatial location.

Fuzzy C Means:
Fuzzy clustering is a form of clustering where each data point can belong to more than one cluster. This is called Soft Clustering.

Algorithm :

1. Choose a number of clusters.
2. Assign coefficients randomly to each data point for being in the clusters.
3. Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than ε)
4. Compute the centroid for each cluster (shown below).
5. For each data point, compute its coefficients of being in the clusters.

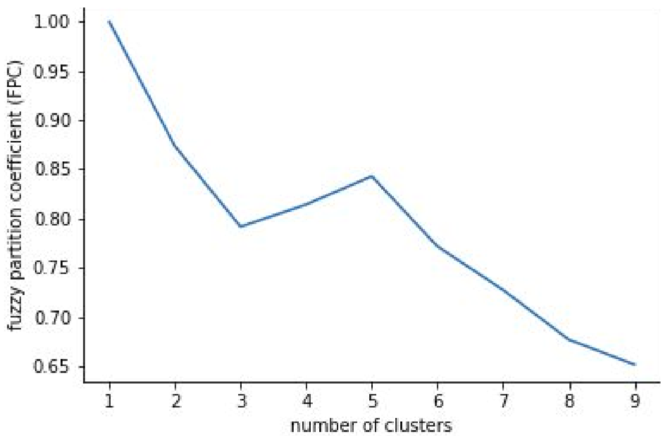
Source: https://en.wikipedia.org/wiki/Fuzzy_clustering

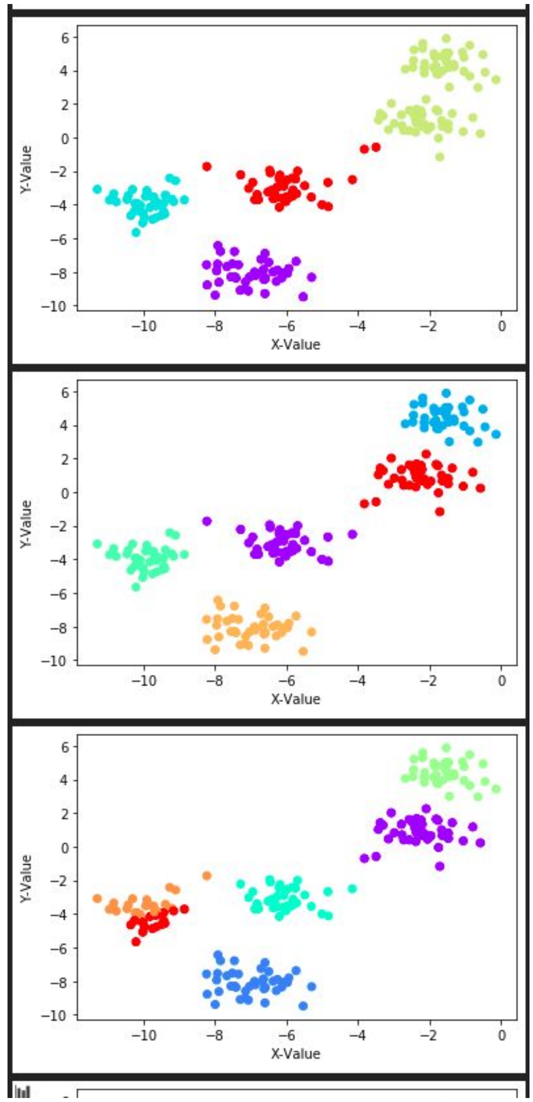
$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)}$$

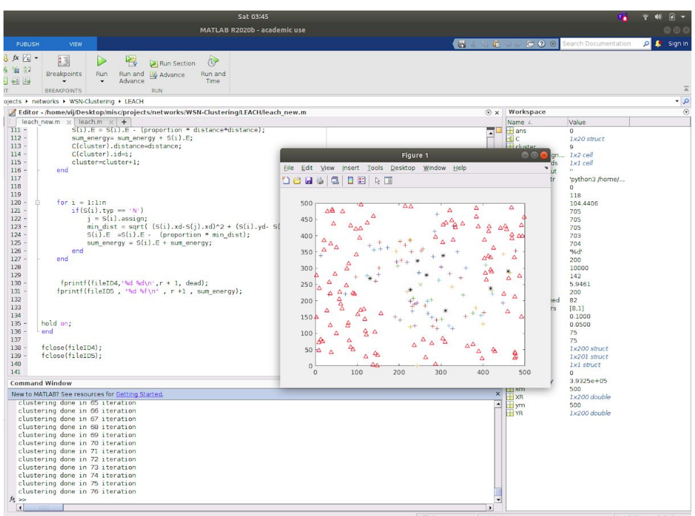
$$v_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i) / (\sum_{i=1}^{n} (\mu_{ij})^m), \forall j = 1, 2, ...c$$

The updations to the membership function u, and the cluster centres are done through these formulae.

We use the weighted clustering because the standard clustering algorithm would fail to take into account the fact that some nodes have higher energies than others, which implies that they are more suitable to become cluster heads so as to distribute the total energy dissipation among all nodes.







By including the weight factor i.e the residual energy factor, we take the best of both worlds, LEACH and EECS (energy-efficient clustering schemes).

In our proposed solution simulation, we initially assign the same energy to all the nodes. We modify the LEACH protocol, and in the setup phase, we feed our weighted clustering ML algorithm 2 parameters i.e the spatial location of each node and each node’s residual energy.

The ML algorithm takes into account the weight i.e energy and returns the cluster assignments and cluster heads, which are then used in the data transfer phase and accordingly energy drain of each node is recomputed.

VI. Implementation/Simulation Setup

We make use of Jupyter notebooks in Python for simulating the various clustering algorithms. Python has several inbuilt libraries which make algorithms scalable and hides the complex math from the user. We have used matplotlib which is a plotting library that provides object-oriented API for embedding applications into applications.

Visual Studio Code was used as the code editor for working with different code files. VSCode is an Open Source Code Editor mainly handled by Microsoft ™.

Python libraries used by us are:

- Numpy
- Pandas
- Scikit-Learn
- Matplotlib
- Random

We have primarily used MATLAB for the simulation of the LEACH protocol. Being a multi-paradigm programming language, modelling and simulation of networks is suitable in this platform.

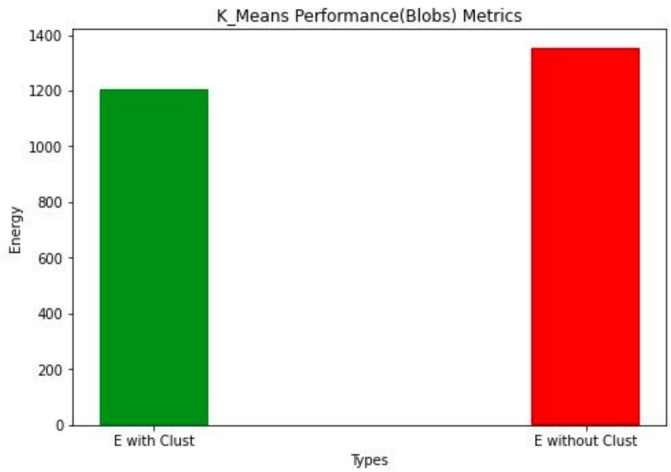
Using file handling techniques, we switch between MATLAB code and the python code, for seamless transfer of data and information via the text files(.txt)

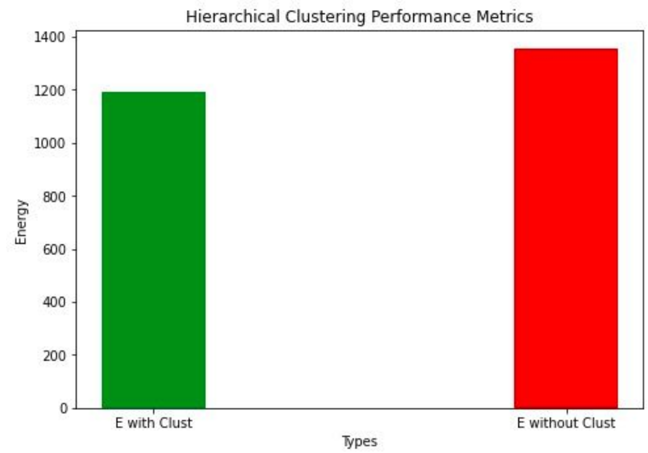
VII. Result Analysis and Comparison

We took the following initial state for our model:

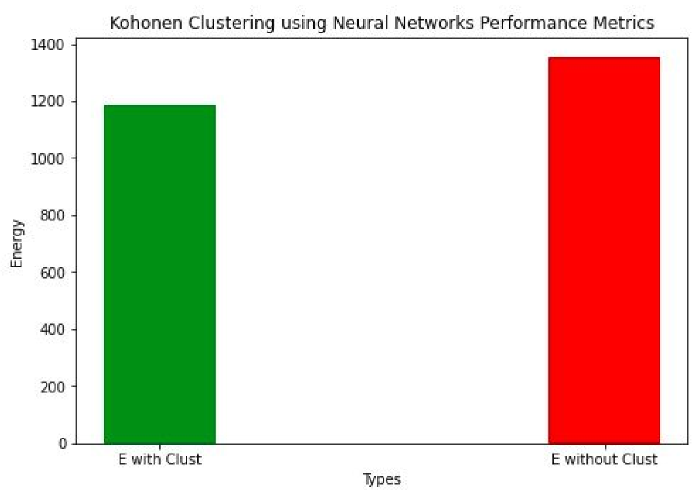
- Number of Active Nodes initially: *200*
- Base Station: *Centre of the network*
- Percentage of nodes that are cluster heads: *0.1*
- Number of total rounds: *75*
- Energy dissipation of cluster head - *proportional to distance squared to BS*
- Energy dissipation of nodes other than cluster heads - *proportional to the distance to CH*

Comparison of Energy before and after clustering:



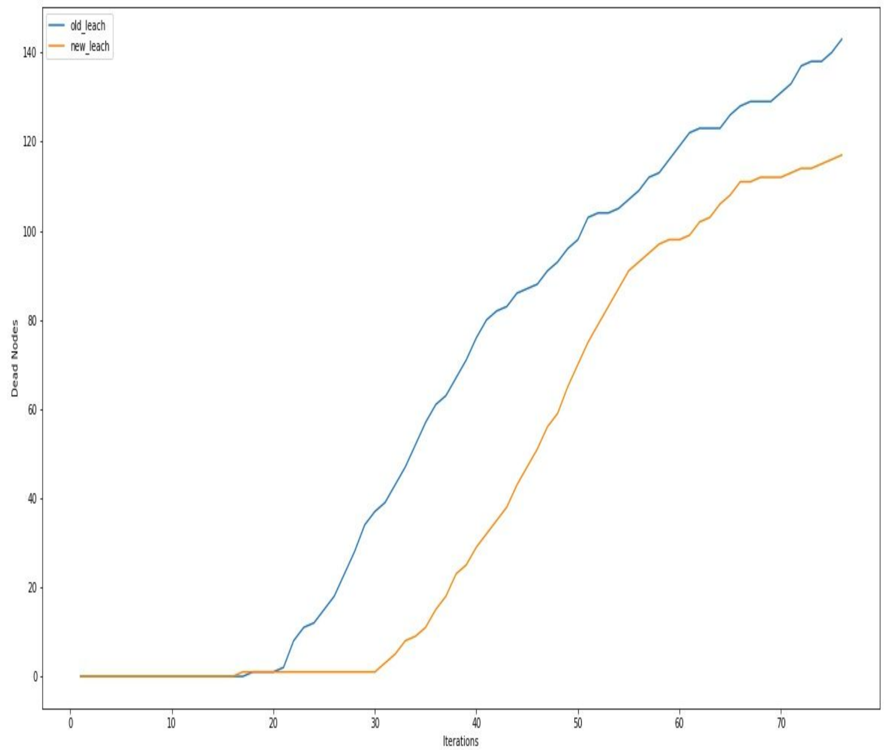


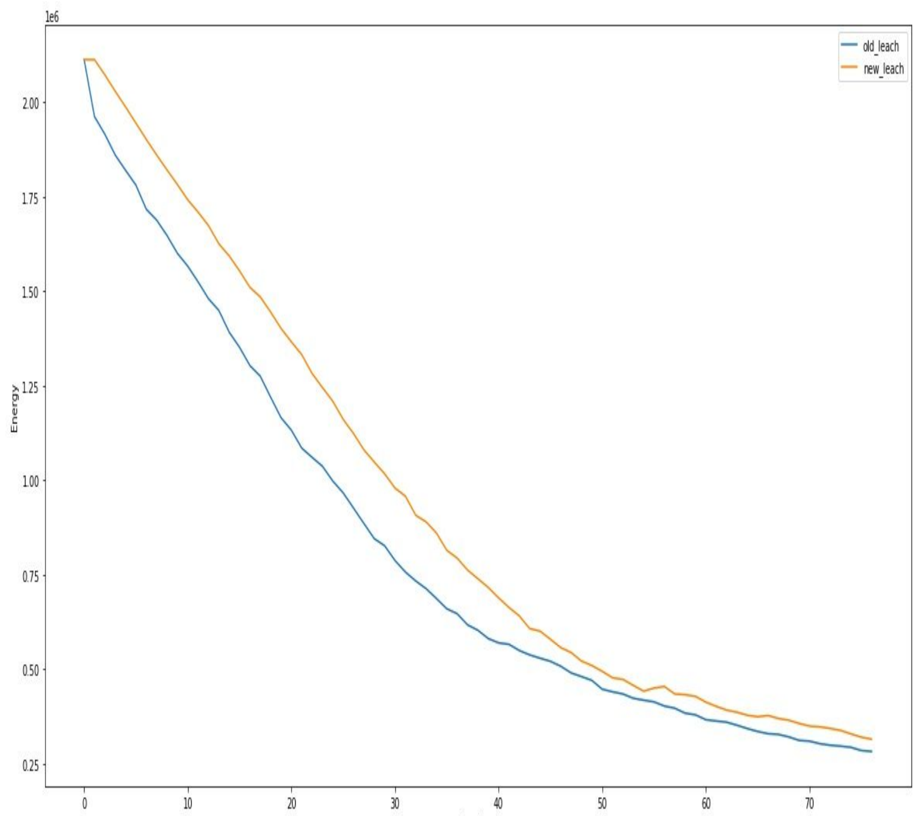
We compared the distances and applied a data aggregation factor to achieve the values of Energy dissipated in a network without clustering and a clustered network with Cluster Heads decided dynamically depending upon the distance.



Analysing the performance of all the three clustering techniques, we realised that clustering using a Kohonen Neural Network gives a result slightly better than Hierarchical clustering and K-Means clustering. Kohonen Neural networks use the concept of winner node which might have lead to the selection of better clusters which are a little different from the traditional clusters.

Comparison of LEACH and Modified LEACH protocol with Clustering Step:





Traditional LEACH protocol uses random cluster heads given according to the threshold function whereas the modified one made by us uses cluster head decided using a weighted clustering technique (Machine Learning Algorithm).

For the comparison between the two protocols, we have used the following performance metrics:

- The number of Dead Nodes after every iteration (Dead Nodes are the nodes that have drained their energy completely and can no longer be a part of the network)

- The sum of Residual Energies of all nodes after every iteration (Residual Energy is the remaining energy of every node after data transfer takes place in the transmission stage)

We can see from the above graphs plotted over 75 iterations that the modified protocol performs better than the original protocol over both the metrics.

VIII. Conclusion and Future Work

Traditional WSNs are evolving and with every new clustering algorithm, the performance is increasing. Not just with respect to residual energy and energy dissipation, but also with other parameters like longevity, quality of service, transmission rate, scalability, etc.

We have initially realised the need for clustering and subsequently implemented several clustering algorithms by simulating it on Jupyter notebooks to analyse their performance.
Namely, we have simulated K-Means clustering, Agglomerative Hierarchical Clustering, Kohonen Neural Networks as well as the LEACH protocol.
We then include a weighted clustering algorithm to optimize the selection of cluster heads in the setup phase of LEACH protocol, and as seen from the results, they are efficient in energy conservation and longevity.

A limitation of this algorithm is that at each iteration, a centralised server must run the ML model to cluster the nodes. Though the computation is not a roadblock, the broadcasting of the clustered results to all the nodes is definitely a challenge.
Further, to get a full picture, this algorithm must be analysed over many more performance metrics like average energy, homogeneity, quality of service, etc. The simulation makes assumptions which may not be true in a real-life scenario where it is more unpredictable and non-ideal. Future work may involve analysis along these lines.

IX. Acknowledgement

We would like to offer our thanks to the National Institute of Technology, Tiruchirappalli for the opportunity to do this research. We would also love to express our sincere gratitude to Dr B.Nithya from The Department of Computer Science Engineering at NIT Trichy. We would also like to express our special thanks to Ms Siron Anita Susan T for guiding us and providing us with valuable feedback.

X. Code

Github Link : https://github.com/HarshitCodex/WSN-Clustering

XI. References

1. https://www.researchgate.net/publication/241677960_A_Survey_on_Clustering_Techniques_for_Wireless_Sensor_Network
2. https://dl.acm.org/doi/10.1145/3167486.3167505
3. https://en.wikipedia.org/wiki/Fuzzy_clustering
4. https://www.edn.com/wireless-sensor-networks-the-basics-part-i/