ISSN 2348 – 8034
Impact Factor- 3.155

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

A SIMPLEST AND EFFICIENT K-MEANS CLUSTERING METHOD FOR WSN

Ritu Mehta[*1] and Mr. Vinod Saroha(guide)[2]
[*1,2]Computer Science and Engineering(Network Security), B.P.S M.V., Khanpur kalan Haryana, India

ABSTRACT

A Wireless Sensor Network (WSN) is a network of small sensor nodes which are energy constraint devices and have limited data transmission and computational power. Clustering is an important mechanism in large multi-hop wireless sensor networks for obtaining scalability, reducing energy consumption and achieving better network performance. Most of the research in this area has focused on energy-efficient solutions, but has not thoroughly analyzed the network performance, e.g. in terms of data collection rate and time. In this paper we are presenting the clustering of wireless sensor network by using k-means approach, over a large dynamic network. As it is the oldest and simplest method of clustering. This method requires only local communication and synchronization. Due to growing in area of peer to peer and mobile sensor networks, data analysis in large, dynamic network in large garner importance in the near future. Our algorithm shows best result for the large dynamic network. We tested our algorithm in a simulated environment up to 100 nodes in a dynamic environment and analyze its behavior with good accuracy.

Keywords- *Wireless Sensor Network, K-means clustering method, Dynamic Network, distance functions, clustering, Euclidean distance .*

I. INTRODUCTION

Clustering is a method that divides the raw data reasonably and searches the hidden patterns that may be present in datasets [2]. It is a process of grouping data objects into disjointed clusters; these obtained clusters should reflect some mechanism at work in the domain from which instances or data points are drawn [3], a mechanism that causes some instances to bear a stronger resemblance to one another than they do to the remaining instances . The Greater the similarity (or homogeneity) within a group and greater the difference between groups, the better is the clustering . There are various techniques available for clustering like k-means clustering technique, hierarchical clustering technique, density based clustering techniques etc. but k-means clustering algorithm is most widely used algorithm because it is simple, efficient and easy to implement [6]. So the author studies the simple k-means algorithm and effect of various distance functions on it because distance function plays major role in finding relationship between various objects in a dataset. K-means is a numerical, non-deterministic, iterative method [2]. So for many practical applications, this method is proved to be very effective to obtain good clustering results. In this paper we first gives the introduction then discussion about k-means clustering algorithm and various distance functions is done in second and third section. In fourth section experimental results are shown and in last section conclusion is given.

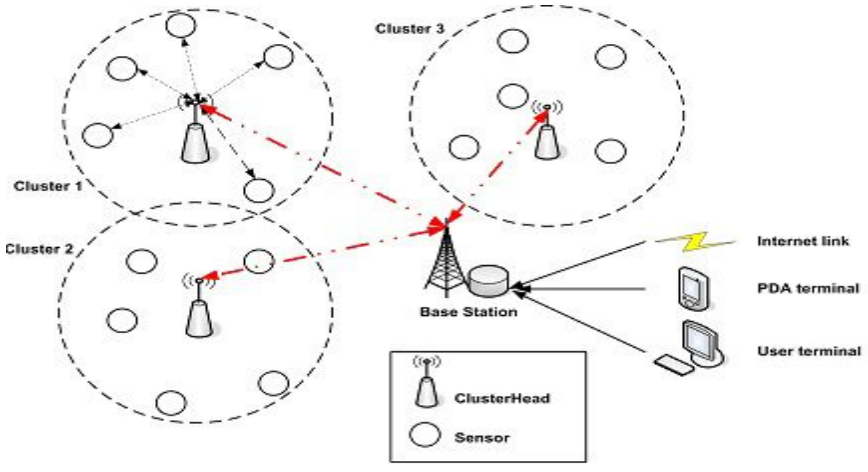


Figure 1: Clustered Sensor Network

ISSN 2348 – 8034
Impact Factor- 3.155

II. THE K-MEANS CLUSTERING METHOD

K-means clustering algorithm is first proposed by Macqueen in 1967 which was uncomplicated, non-supervised learning clustering algorithm [9]. K-means is a partitioning clustering algorithm, this technique is used to classify the given data objects into k different clusters through the iterative method, which tends to converge to a local minimum. So the outcomes of generated clusters are dense and independent of each other [5].
The algorithm consists of two separate Steps.
(i). In the first phase user selects k centres randomly, where the value k is fixed in advance. To take each data object to the nearest centre. Several distance functions are considered to determine the distance between each data object and the cluster centres. When all the data objects are included in some clusters, the first step is completed and an early grouping is done.
(ii). Then the second phase is to recalculate the average of the early formed clusters. This iterative process continues repeatedly until the criterion function becomes the minimum.

The process of k-means algorithm as follow:

- Arbitrarily generate k points (cluster centers),k being the number of clusters desired.
  Initial K points are chosen randomly.
- Calculate the distance between each of the data points to each of the centers, and assign each point to the closest center. “Closeness” is measured by Euclidean Distance, cosine
  Similarity, correlation, etc.
- Calculate the new cluster center by calculating the mean value of all data points in the respective cluster.
- With the new centers, repeat step 2. If the assignment of cluster for the
- data points changes, repeat step 3 else stop the process.

The distance between the data points is calculated using Euclidean distance as fol-lows. The Euclidean distance between two points or tuples, X1 = (x11; x12 : : : x1n), X2 = (x21; x22 : : : x2n).

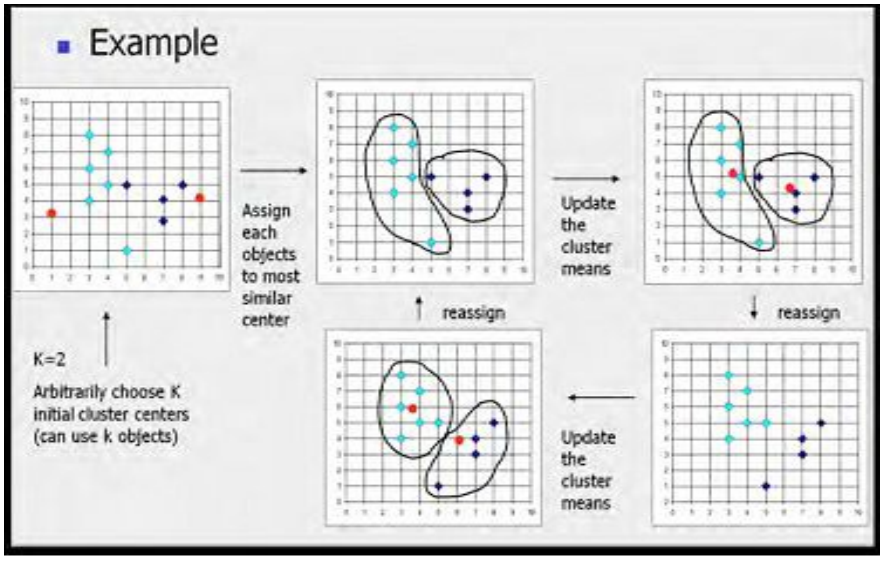


Figure 2: Original K-means Algorithm

ISSN 2348 – 8034
Impact Factor- 3.155

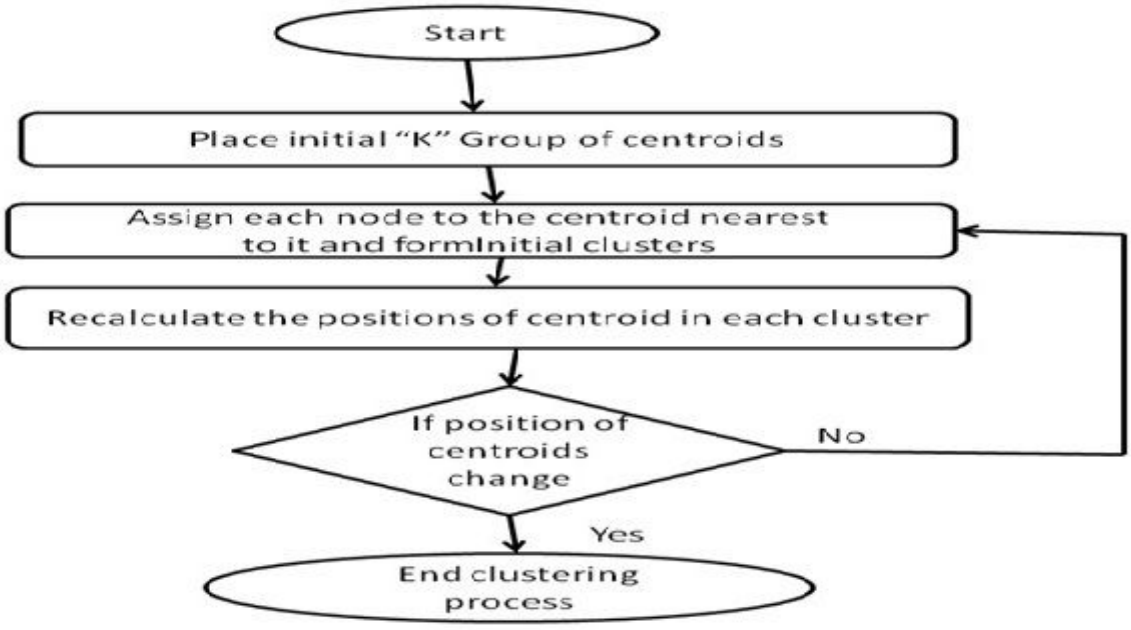
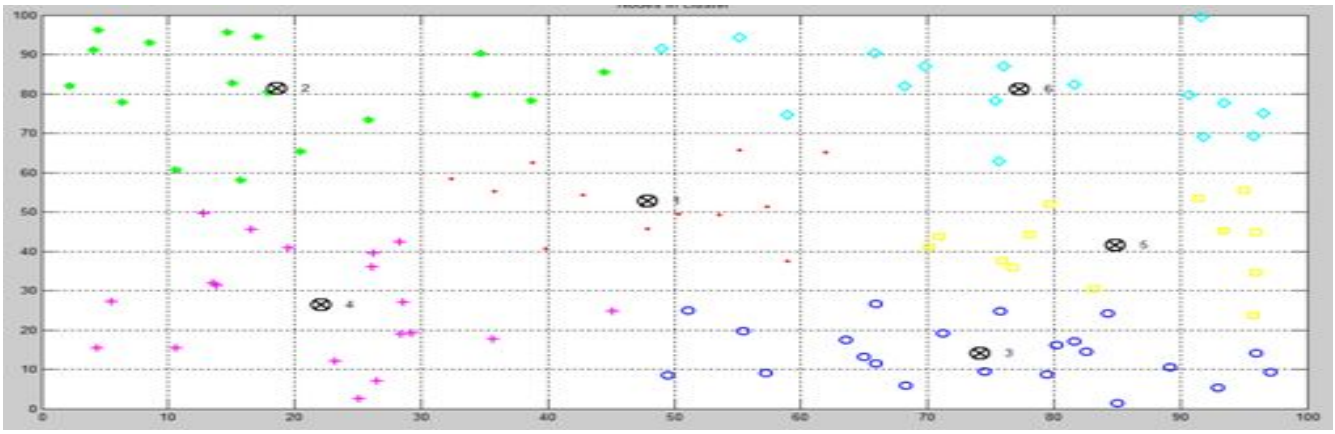


Figure 3: The Flow Chart for the k-means algorithm

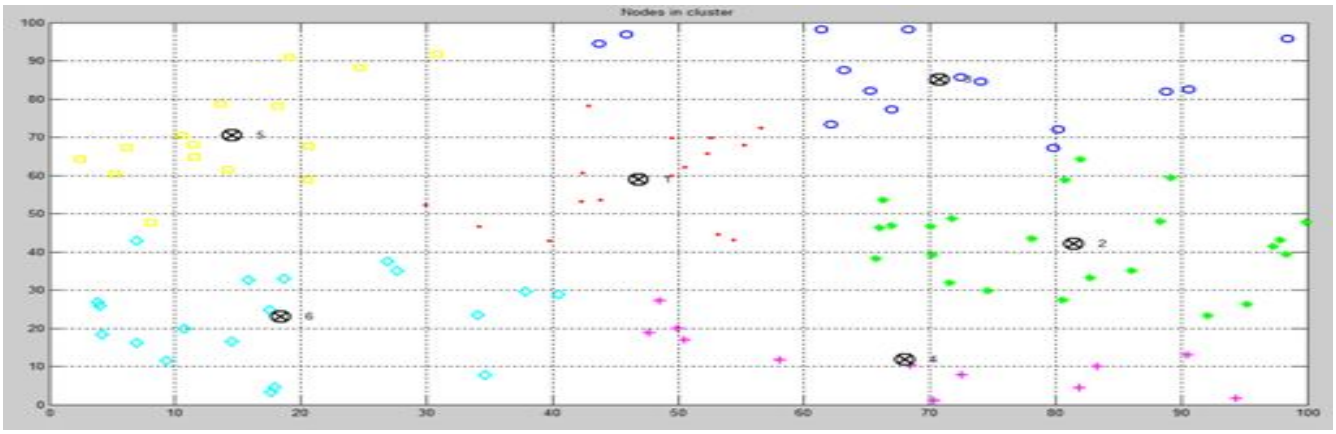
III. SIMULATION OUTPUT FOR K-MEANS METHOD IN MATLAB

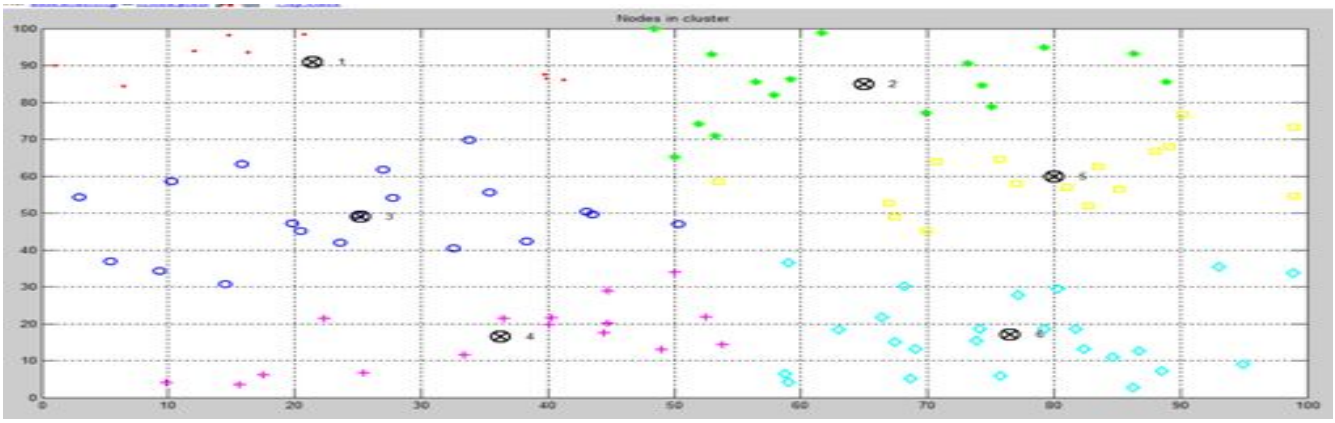
Here we are taking the 6 clusters with 100 nodes in dynamic network and use the MATLAB simulator that shows how nodes or dataset is distributed by using k-means in various iteration.

At iteration 1,2 and 3 respectively



ISSN 2348 – 8034
Impact Factor- 3.155





At Various iteration we see that placement of nodes with their cluster position changes dynamically and automatically gain their optimum position. This type of clustering is gaining importance in various application and simulated environment to achieve the desired goal.

IV. SOME IMPORTANT APPLICATIONS OF K-MEANS CLUSTERING

1. **Image Segmentation**

The *k*-means clustering algorithm is commonly used in computer vision as a form of image segmentation. The results of the segmentation are used to aid border detection and object recognition.

2. Clustering can be applied to detect abnormality in wind data (abnormal vibration)
3. Monitor Wind Turbine Conditions.
4. Beneficial to preventative maintenance.
5. K-means can be more powerful and applicable after appropriate modifications.

V. ADVANTAGES & DISADVANTAGES OF K-MEANS CLUSTERING

- **Advantage:**

1. If variables are Huge, then K-Means most of the times computationally faster than hierarchical clustering, if we keep k smalls.
2. K-means produce tighter clusters than hierarchical clustering, especially if the cluster is globular.
3. Transmit aggregated data to the data sink.
4. Reducing number of nodes taking part in transmission.
5. Useful Energy consumption.
6. Scalability for large number of nodes.

ISSN 2348 – 8034
Impact Factor- 3.155

7. Reduces communication overhead.
8. E_cient use of resources in WSNs.

- **Disadvantages:**

1. Difficult to predict K-value.
2. With global Cluster, it didn't work well.
3. Different initial partitions can result in different final clusters.
4. It does not work well with clusters (in the original data) of different size and different density.

VI. CONCLUSION

K-Means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem because of its fast execution and easy implementation. Finally, the objective function of this algorithm aims at minimizing the total distortion (squared error). Distortion is also known as sum of distances of points from their cluster centroids.

REFERENES

[1] P. Sasikumar, Sibaram Khara "k-means clustering in wireless sensor networks" 2012 Fourth International Conference on Computational Intelligence and Communication Network

[2] H. G. Wilson, B. Boots, and A. A. Millward 2002, "A Comparison of Hierarchical and Partitional Clustering Techniques for Multispectral Image Classification", 0-7803-7536-X/$17.00 (C) 2002 IEEE .

[3] Glenn Fung 2001, "A Comprehensive Overview of Basic Clustering Algorithms" June 22, 2001

[4] Michael Steinbach , Levent Ertöz and Vipin Kumar, "The Challenges of Clustering High Dimensional Data", Access to computing facilities was provided by AHPCRC and the Minnesota Supercomputing Institute.

[5] Tung-Shou Chen, Tzu-Hsin Tsai, Yi-Tzu Chen, Chin-Chiang Lin, Rong-Chang Chen, Shuan-Yow Li and Hsin-Yi Chen 2005, "A Combined K-Means And Hierarchical Clustering Method For Improving The Clustering Efficiency Of Microarray", Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems December 13-16, 2005 Hong Kong, 0-7803-9266-3/05/$20.00 ©2005 IEEE

[6] Wang, Xuan, "Clustering in the Cloud: Clustering Algorithms to Hadoop Map/Reduce Framework" http://ecommons.txstate.edu/cscitrep/19 Paper 19, 2010.

[7] Eisuke Shimoda, Hiroshi Ohtake, Shinichi Iwamoto and Masayuki Namba Reliability Evaluation of Multiple Facility Faults for Sub-Transmission Systems Using Tabu Search and Cluster Analysis, Power Engineering Society General Meeting IEEE, 2006.

[8] Seema Bandyopadhyay and Edward J. Coyle An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks, 22nd Joint Conference of the IEEE Computer and Communications. IEEE, Vol.3, pp. 1713-1723, 2003.

[9] Jamal N. Al-Karaki and Ahmed E. Kamal, Routing techniques in wireless sensor networks: A survey, IEEE Wireless Communications", Vol. 11, Issue. 6, pp. 6-28, December 2004.

[10] Wanming Chen, Max Q.-H. Meng, Shuai Li, Tao Mei, Huawei Liang and Yangming Li, Energy Efficient Head Node Selection Algorithm in Wireless Sensor Networks, IEEE International Conference on Robotics and Biomimetics, pp.1366-1371, Dec. 2007.

[11] Wanming Chen, Max Q.-H. Meng, Shuai Li, Tao Mei, Huawei Liang and Yangming Li, Energy Efficient Head Node Selection Algorithm in Wireless Sensor Networks, IEEE International Conference on Robotics and Biomimetics, pp.1366-1371, Dec. 2007.

[12] Mohammad Karimi, Hamid Reza Naji, and Shahrzad Golestani, Optimizing Cluster-Head selection in Wireless Sensor Networks using Genetic Algorithm and Harmony Search Algorithm, 20th Iranian conference on Electrical Engineering (ICEE) , pp. 706-710, 2012.

[13] Buddha Singh and Daya Krishan Lobiyal, A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks, Springer, July 2012.

[14]Pham, D. T., Dimov, S. S., and Nguyen, C. D. Incremental K-means algorithm. Proc. Instn Mech.Engrs, Part C: J. Mechanical Engineering Science, 2003, 218, 783–795.

[15] Al-Daoud, M. B., Venkateswarlu, N. B., andRoberts, S. A. Fast K-means clustering algorithms.Report 95.18, School of Computer Studies.