DEVELOPMENT OF WIRELESS SENSOR NETWORKS EFFICIENT CLUSTERIZATION METHOD

Nawar Mohammad [1]

[1] Al-Baath University, Homs, Syrian Arab Republic

DOI: 10.36724/2664-066X-2024-10-1-32-38

Received: 10.12.2023
Accepted: 20.01.2024

Citation: Nawar Mohammad, "Development of wireless sensor networks efficient clusterization method," *Synchroinfo Journal* **2024**, vol. 10, no. 1, pp. 32-38.

ABSTRACT

Wireless sensor networks are a critical element of the Internet of Things and are widely used in various fields. Effective clustering is key to energy efficiency and network stability in wireless sensor networks. Optimize the density of clusters, their number and size to increase the stability and service life of the network – It is one of the tasks of clustering wireless sensor networks. Another problem – effectively organize the network topology in order to balance the load and extend the network lifetime. Clustering has proven to be an effective approach for organizing a network into a cohesive hierarchy. However, there are several key issues that affect the practical application of clustering methods in sensor network applications. These issues include the rationale for developing different clustering approaches, classification of proposed approaches based on their goals and design principles, and problems associated with clustering. K-means algorithm is one of the unsupervised machine learning methods for clustering tasks and is widely used in wireless sensor networks. This article proposes a modified K-means clustering algorithm that eliminates the shortcomings of the basic K-means algorithm. The modified algorithm increases network lifetime, reduces power consumption and improves network stability.

KEYWORDS: *efficient clustering, K-means algorithm, energy efficiency, stability, lifetime, wireless sensor networks.*

1 Introduction

Wireless sensor networks (WSNs) are a collection of autonomous sensors distributed in space that monitor physical conditions or the state of the environment [1, 2].

The integration of WSN and the Internet of Things (IoT) has led to the creation of intelligent systems capable of remotely monitoring and managing various processes, as well as developing applications in various fields. These areas include environmental monitoring, home automation, detection of chemical/biological attacks, healthcare for monitoring patient vital signs and detecting abnormalities, smart agriculture for optimizing crop yields by monitoring humidity, temperature and soil moisture, smart transport for monitoring traffic flows and detecting accidents, energy for monitoring energy consumption and detecting power outages, and manufacturing for monitoring the production process and detecting defects [3-5].

The sensors cooperatively transmit their data through the network to the base station. There are various methods for organizing data transmission in a WSN, but clustering is predominantly used by selecting certain nodes as cluster head nodes (CHN) to perform data aggregation and communication tasks.

Grouping WSN nodes into clusters provides a number of advantages, such as improved scalability, reduced routing latency, and increased energy efficiency. Thus, Clustering reduces the number of paths for data transmission and reduces the load on the network and improves the energy efficiency, scalability and lifetime of the WSN [6].

One of the tasks of WSN clustering is to optimize the density of clusters, their number and size to increase the stability and service life of the network. Another problem is the need to effectively organize the network topology in order to balance the load and extend the network lifetime. Clustering has proven to be an effective approach for organizing a network into a cohesive hierarchy. However, there are several key issues that affect the practical application of clustering methods in sensor network applications. These issues include the rationale for the development of various clustering approaches, the classification of proposed approaches based on their goals and design principles, as well as problems associated with WSN clustering [7, 8].

The K-means algorithm is one of the unsupervised machine learning methods for performing clustering tasks and is widely used in WSNs [9-12].

2 Effective Clustering Method

The K-means algorithm is one of the clustering methods. However, it has two disadvantages:

the number of clusters (K) is determined manually before the start of modeling,

The algorithm initially initializes the cluster centers randomly, which leads to different results (cluster centers) each time the algorithm is run on the same data and to the instability of the algorithm.

The pseudocode of the basic K-means algorithm is described as follows:

Stage 1. Initialize K cluster centers randomly.

Stage 2. Determination of cluster centers and nodes belonging to each cluster:
where *m* – number of nodes; *K* – number of clusters; $\mu_1, \mu_2, \ldots, \mu_K \in R^n$ – cluster centers; $c^{(i)} \in C$ – a vector of nodes containing the cluster number to which each node belongs.

Repeat {

For i = 1 to m

$c^{(i)}$ = index (from 1 to K) of the cluster center, nearest к $x^{(i)}$,

For k = 1 to K

$\mu_k \mu_k$ = calculated as the averaged sum over the coordinates of vectors nodes included in cluster *k*

}

The clustering process ends when all cluster centers coordinates $\mu_1, \mu_2, \ldots, \mu_K$ do not change compared to their previous coordinates.

To eliminate these shortcomings, a modified K-means algorithm developed by the author is used, which will be described below taking into account radio visibility zones and GPS coordinates of nodes, which allows optimizing the number of clusters and avoiding "crowding" of the CHN in a small area of the coverage area.

Effectiveness of clustering ultimately depends on finding the optimal number of network clusters. To achieve this goal, the author modified basic K-means algorithm as follows.

Stage 1. The initial number of clusters is determined based on the shape of the sensing field (square field), radio communication model (free space) and number of nodes distributed in sensing area [10].

$$K_{opt} = \sqrt{\frac{3*N}{\pi}} = 0.977*\sqrt{N} \quad (1)$$

where N – number of nodes.

Stage 2. The disadvantage of cluster centers (RCI) random initialization is eliminated by finding the initial virtual points of cluster centers along a circle in accordance with formula (2):

$$\theta = \frac{2*\pi}{K_{opt}},$$

$$X_{CNH} = R*\cos(\theta*k); k = 1,....,K_{opt}; R = \frac{L}{4} \quad (2)$$

$$Y_{CNH} = R*\sin(\theta*k)$$

where K_{opt} – optimal number of clusters; L – sensor field length; θ – the angle defining distance between CHNs on the circle; R – radius of a circle equal to a quarter of the probing zone length. This approach avoids CHN crowding in a small area of the coverage area. After determining the cluster center point for each cluster, the node closest to it is found, which is considered as CHN. The nodes within each cluster are then checked to see if they fall within the CHN node range.

Stage 3. Determine the cluster centers and nodes belonging to each cluster as in step 2 of the basic K-means algorithm. Nodes that are not within radio visibility of any CHN node send their data directly to the base station.

Stage 4. An algorithm is performed to find the closest node to the center of cluster and assign it as a CHN using the Euclidean distance according to formula (3).

$$p(x,y) = \| x-y \| = \sqrt{\sum_{p=1}^{n}(x_p - y_p)^2}; x,y \in R^n$$

$$\text{nodes } (x^{(1)}, x^{(2)},...,x^{(i)},...,x^{(m)}); x^{(i)} \in R^n$$

$$\mathrm{p(x,y)} = \|\mathrm{x\text{-}y}\| = \sqrt{\sum_{\mathrm{p}=1}^{\mathrm{n}}\left(\mathrm{x_p\text{-}y_p}\right)^2}, \mathrm{x,y} \in \mathrm{R^n} \quad (3)$$

Figure 1 shows the block diagram of modified K-means algorithm. Initial parameters are set at the input, then the optimal number of clusters Kopt is determined using formula (1) and testing experiments are performed to confirm the values used, then the initial coordinates of cluster centers μk on the circle are initialized using formula (2). Inside the cycle, Euclidean distance p(x,y) between each node and cluster center closest to it is calculated using formula (3), and a vector of nodes C is found, containing the numbers of clusters to which each node belongs.

Then the coordinates of cluster centers μк in each cluster are recalculated by finding averaged sum over the vectors nodes coordinates included in each cluster. Then it is checked whether the coordinates of current cluster centers have changed compared to previous ones. If they have changed, then the cycle is repeated again to find new coordinates of the cluster centers. If the coordinates have not changed, then the cycle stops, and in each cluster there is a node closest to the center of this cluster, which is designated as a CHN node.

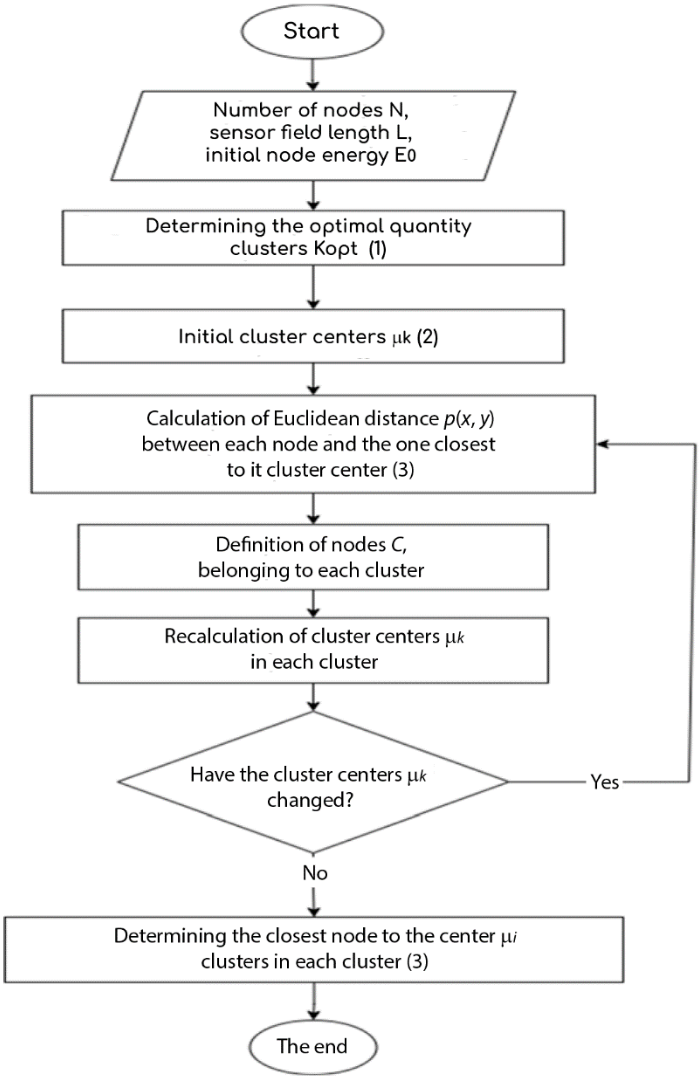


Figure 1. Block diagram of the modified K-means algorithm

As a result, an optimal sequence of clusters was formed in the sensor field and all the head cluster nodes were found.

3 Experiments

Determining the optimal number of clusters is one of the important tasks of clustering, which allows reducing energy consumption, extending the network lifetime and grouping nodes into clusters. In this regard, several experiments will be carried out to find the optimal number of clusters (K) in a modified K-means algorithm with a different number of nodes (100, 250, 500, 1000) in a model square on a plane, with the gateway/base station located in the center of the sensor fields by choosing a value of K at which the network lifetime will be as high as possible.

Table 1

Network lifetime for different numbers of clusters, when the BS is in the center of sensor field and number of nodes is N = 100

	Number of nodes N = 100, BS is located in the center of sensor field						
K	7	8	9	10	11	12	13
Time (round)	4838	4894	4917	4992	4951	4838	4877

From Table 1 it is clear that with a value of K = 10, the network lifetime is longer than with other values of K. Therefore, we can assume that in the case when the base station is located in the center of the sensor field, and the number of nodes is 100, the optimal number of clusters is K equals 10.

In the case when number of nodes is 250, from Table 2 shows that the optimal number of clusters K is 15.

Table 2

Network lifetime for different numbers of clusters, when the BS is in the center of sensor field and number of nodes is N = 250

	Number of nodes N = 250, BS is located in the center of sensor field						
K	12	13	14	15	16	17	18
Time (round)	4528	4605	4583	4702	4670	4644	4648

Из таблицы 3 также видно, что оптимальное число кластеров K равно 22 при числе распределенных узлов 500 и расположении базовой станции в центре сенсорного поля.

Table 3

Network lifetime for different numbers of clusters, when the BS is in the center of sensor field and number of nodes is N = 500

	Number of nodes N = 500, BS is located in the center of sensor field						
K	18	19	20	21	22	23	24
Time (round)	4575	4574	4547	4565	4587	4557	4540

Table 4 shows that when the number of nodes is 1000 and the BS is located in the center of sensor field, optimal number of clusters is 31.

Table 4

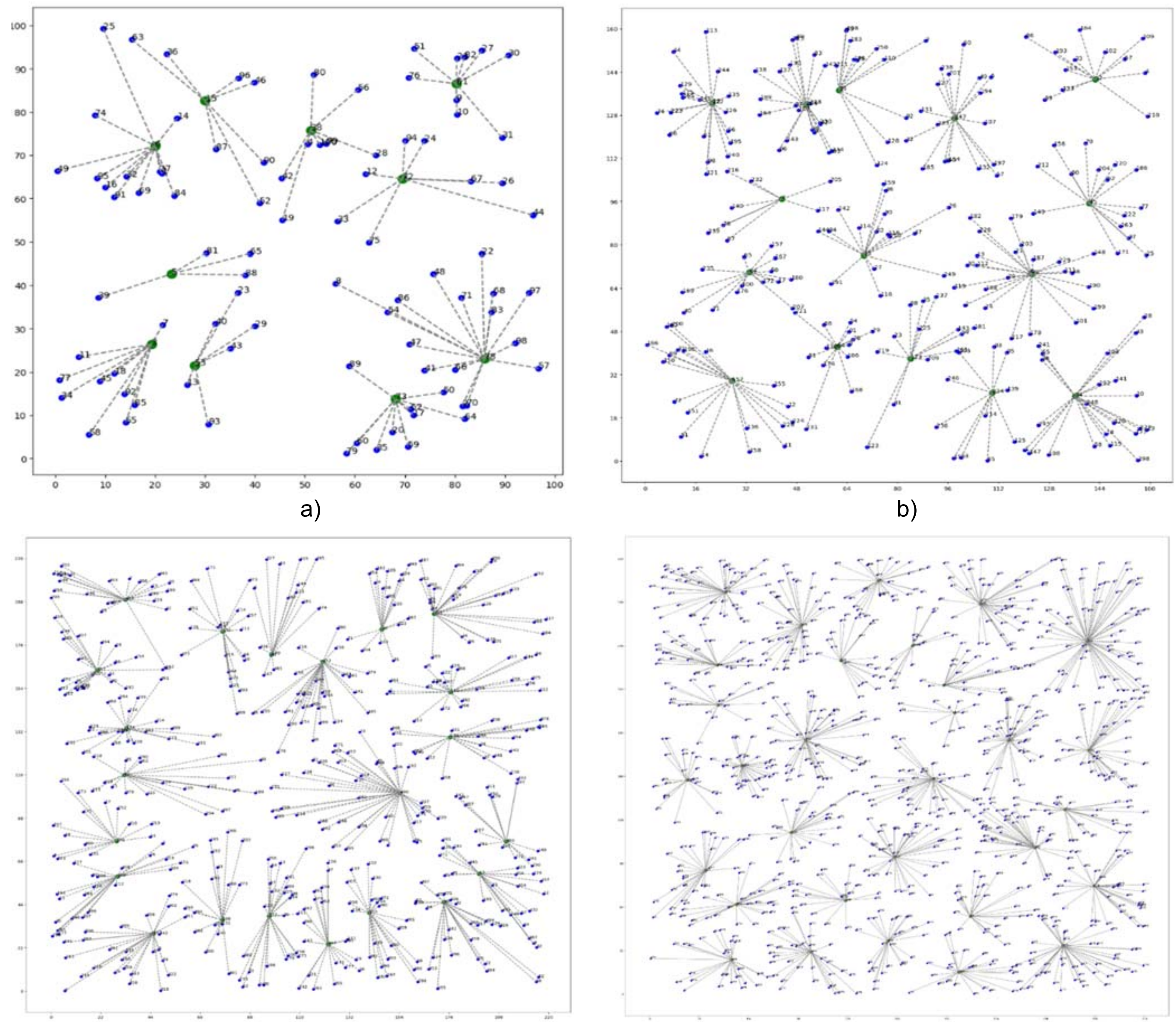
Network lifetime for different numbers of clusters, when the BS is in the center of sensor field and number of nodes is N = 1000

	Number of nodes N = 1000, BS is located in the center of sensor field					
K	28	29	30	31	32	33
Time (round)	4548	4555	4560	4580	4577	4559

Based on the previous tables (1-4), we can conclude that the results obtained confirm effectiveness of formula (1) in calculating optimal number of clusters in the modified

K-means algorithm. Accordingly, it can be considered as a criterion for determining the optimal number of clusters.

Figure 2 shows a visualization of the resulting distribution, showing the connection of CNH with nodes included in cluster.



a) b) c) d)

Figure 2. Visualization of the distribution of different nodes in clusters with different area sizes: a) 100 nodes per area (100*100): b) 250 nodes per area (160*160): c) 500 nodes per area (220*220): d) 1 0 nodes on the area (320*320)

Table 5

Results of experiments on effective clustering when gateway/BS is located inside the sensor field

No.	Number of nodes (N)	Square touch field length, m	PC operating time (t), sec	Number of clusters (K)
1	100	100	0,03	10
2	250	160	0,06	15
3	500	220	0,09	22
4	1000	320	0,28	31

Table 5 shows the results of experiments with different numbers of nodes in the sensor field. For each coverage area, the optimal number of clusters was found, and the nodes belonging to each cluster were determined.

Modeling efficient clustering of WSNs. The simulation is carried out with coordinates *x*, *y* and *z* in two-dimensional space, initially assuming that $z = 0$.

Four experiments were carried out with different numbers of nodes (100, 250, 500, 1000) in a model square on a plane, with varying area sizes (100*100, 160*160, 220*220, 320*320) using Python in the Anaconda environment.

Thus, the experiments have shown that the use of a modified K-means algorithm for clustering different numbers of WSN nodes gives adequate results, showing that with an increase in the number of nodes over an increasing coverage area, with a generally unchanged network topology, the dimensions clusters do not differ too much in the number of incoming nodes, which leads to a natural gradual increase in the number of clusters.

5 Conclusion

The research described in this paper proposes a modified K-means algorithm for efficient clustering in WSNs based on the basic K-means algorithm to improve its performance in wireless sensor networks. The modified K-means algorithm eliminates the limitations of the basic K-means algorithm by determining the optimal number of clusters and placing cluster centers around a circle to avoid crowding of the GCU in a small area of the coverage area. The results show that the modified K-means algorithm can find the optimal number of clusters and effectively determine cluster centers, reducing energy consumption, extending the lifetime of the WSN, and increasing its stability.

REFERENCES

[1] M. Ahmad, B. Shah, A. Ullah, F. Moreira, O. Alfandi, G. Ali, and A. Hameed. Optimal clustering in wireless sensor networks for the Internet of things based on memetic algorithm: memeWSN. *Wireless Communications and Mobile Computing*, 2021, pp.1-14.

[2] I. R. Aseydulin. Study of low-energy adaptive clustering based on hierarchical routing of wireless sensor networks. *Trends in the development of science and education*. 2020. No. 63-3, pp. 25-28.

[3] M. Majid, S. Habib, A. R. Javed, M. Rizwan, G. Srivastava, T. R. Gadekallu, J. C.-W. Lin. Applications of Wireless Sensor Networks and Internet of Things Frameworks in the Industry Revolution 4.0: A Systematic Literature Review. *Sensors* 2022, 22.2087. https://doi.org/10.3390/s22062087.

[4] A. H. Najim, S. Kurnaz. Study of Integration of Wireless Sensor Network and Internet of Things (IoT). *Wireless Pers Commun*. 2023. https://doi.org/10.1007/s11277-023-10556-4.

[5] A. Khanna, S. Kaur. Internet of Things (IoT), Applications and Challenges: A Comprehensive Review. *Wireless Pers Commun.* 114, pp. 1687-1762. 2020. https://doi.org/10.1007/s11277-020-07446-4.

[6] Q. Tang, F. Nie. Clustering routing algorithm of wireless sensor network based on swarm intelligence. *Wireless Networks*, 2023, pp. 1-12.

[7] M. Gheisari et al. A Survey on Clustering Algorithms in Wireless Sensor Networks: Challenges, Research, and Trends. *2020 International Computer Symposium (ICS)*, Tainan, Taiwan, 2020, pp. 294-299.

[8] A. M. Jubair, R. Hassan, A. H. M. Aman, H. Sallehudin, Z. G. Al-Mekhlafi, B. A. Mohammed, M. S. Alsaffar. Optimization of Clustering in Wireless Sensor Networks: Techniques and Protocols. *Applied Sciences*. 2021, no. 11(23), p. 11448.

[9] L. I. Voronova, V. I. Voronov, N. Mohammad. Modeling the Clustering of Wireless Sensor Networks Using the K-means Method. *Proceedings of the 2021 IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies", T&QM&IS*, 06-10. Yaroslavl, 2021, pp. 740-745. DOI 10.1109/ITQMIS53292.2021.9642747.

10. N. Mohammad, L. I. Voronova Modeling of clustering of a wireless sensor network using a neural network constructive method. *Systems of synchronization, signal generation and processing*. 2021. Vol. 12. No. 3, pp. 4-19.

11. N. Mohammad, L. I. Voronova, V. I. Voronov. Modeling routing in a clustered swarm of UAVs using a genetic algorithm. *First Mile*. 2023. DOI: 10.22184/2070-8963.2023.114.6.46.52.

12. N. Mohammad, L. I. Voronova, V. I. Voronov, A.I. Larin. Program for modeling dynamic clustering of a swarm of UAVs using the K-means algorithm. Certificate of state registration of the computer program No. 2023618048, 2023.

13. A. Navid, V. Alireza, X. Wenyao, G. Mario, S. Majid. Cluster size optimization in sensor networks with decentralized cluster-based protocols. *Journal of Computer Communications*, 2012, no. 35(2), pp. 207-220.