K-means Clustering and PSO Algorithm for Wireless Sensor Networks Optimization

Goran W. Hamaali[†], Khalid A. Abduljabbar[‡], and Diary. R. Sulaiman[†]

[†]Department of Electrical Engineering, College of Engineering, Salahaddin University-Erbil, Kurdistan region, Iraq, goran.hamaali@su.edu.krd. [‡]Department of Electrical Engineering, College of Engineering, Salahaddin University-Erbil, Kurdistan region, Iraq, khalid.abduljabbar@su.edu.krd.

[†] Department of Electrical Engineering, College of Engineering, Salahaddin University-Erbil, Kurdistan region, Iraq, diary.sulaiman@su.edu.krd. Abstract

K means algorithm is one of the powerful and popular unsupervised machine learning algorithms, typically used in wireless sensor networks WSNs to separate the network into subnetwork or smaller networks known as clusters K.

The purpose of clustering is to reduce the amount of energy consumed in the network which results into improve the lifetime of the network, Determining the optimal number of K is the most challenging in WSN research area. In the first part of this paper the two most powerful methods El-bow, and Silhouette approaches are used to form clustering and implemented on three different types of real-world applications depending on the size of network. Extensive simulations show that Elbow method is more appropriate in small and medium sized networks compared to silhouette method which shows its robustness in large network due to a minimum amount of time is used to form subgroups. In the second part of this study low energy adaptive clustering hierarchy (LEACH) and Particle Swarm Optimization PSO based LEACH protocol is utilized on small sized network to validate the efficacy of the clustering used in the first part, The results show that PSO based LEACH protocol outperforms LEACH protocol in terms of energy consumption and number of packets sent when the nodes are communicating. PSO based LEACH protocol sends more packets and has fewer dead nodes, resulting in lower energy consumption.

Keywords: WSN, Clustering, Optimal K-means, ELBOW method, SILHOUETTE method, PSO.

1. Introduction

Clustering is the process of arranging a set of patterns into disjoint clusters. This is done so that patterns belonging to two different clusters are different. Clustering has been a widely studied problem in a variety of application domains including neural networks, AI, and statistics [1]. The performance of a clustering algorithm is significantly impacted by the selected value for K. Instead of using a single predefined K, a set of values are possibly adopted. It is crucial that a sizable number of values are considered to reflect the specific characteristics of the data sets [2, 3]. In general, clustering methods may be divided into partition methods, hierarchical methods with several levels of grouping, density-based methods, and grid-based methods to handle spatial data. Partition methods use a mean or medoid to represent the cluster center with one level of grouping. The average mean partitioning technique for small and medium data is the most effective and efficient, based on features, ease of implementation, and computing efficiency [4].

Several research have been conducted on determining the centroid value, including the principle of heightening each cluster variation using min-max kmeans [1]. Selecting the optimal number of clusters using the elbow approach and calculating the centroid using mean and median data. The study's results demonstrate that choosing the initial centroid based on mean data reduces the number of iterations required by 22.58 percent, while choosing the optimal number of clusters using the elbow technique reduces the number of iterations by 25% [5]. Shi Na et al proposed an modified K means cluster algorithm. Data structure is used to store

the distances between the nodes to avoid the unwanted repeated distance calculations which improves the efficiency of the network. Their results show that the accuracy of clustering [6]. Amine et al suggest a multihop and dynamic k-medoids idea and elbow based algorithms for energy saving, The method uses better communications between the nodes. Extensive simulation proves that an improvement of 30% in terms of first node FN compared to normal K-means and improving of 108% in terms of the lifetime of the applied WSN network [7]. For a wide range of applications, the problem of data aggregation among sensor nodes is crucial. This issue may be effectively solved by clustering the sensor nodes. The authors proposed KPSO Algorithm, as the initial proposal, which is a hybrid clustering algorithm built on K-means clustering and particle swarm optimization which was able to significantly lengthen the lifetime of nodes [8]. Alaa and Basma investigated two hybrid K-means clustering methods. They are K-Means/ Genetic Algorithm called (KGA) and K-Means / Particle Swarm Optimization called (KPSO). Every algorithm has two stages: Phase 1 is to divide the network into a predetermined number of clusters. Phase 2 is to choose the optimum Cluster Head (CH) for each phase. Compared to the KGA algorithm, KPSO demonstrated superior outcomes in terms of transmission rate, longer network life consequently. The amount of data samples obtained from the region of interest should increase as the network lifetime increases[9]. El-Awady et al proposed a combined Evolutionary Algorithm (EA) and Kmeans clustering algorithm, in order to shorten the communication range in a sensor network. The recommended approach is an effective way to solve the clus-

K-means Clustering and PSO Algorithm for Wireless Sensor Networks Optimization

tering problem quickly and simply according to experimental data. This is accomplished by using the K-mean technique to divide the n population into k known clusters, and the evolutionary algorithm to determine the minimum fitness of each cluster [3]. Reham and Mustapha created a hierarchical clustering technique based on PSO for Wireless Sensor Networks (WSN), By establishing an upper bound on the number of CHs, the protocol improves WSN energy efficiency. It also improves the network's scalability by adopting a two-tier cluster topology [10].

Elbow and Silhouette procedures, two of the most effective methods for clustering, are employed in the first section of this study and are applied to three distinct kinds of practical applications based on the network size. Numerous simulations demonstrate that the Elbow technique is better suitable for small and medium-sized networks than the silhouette method, which demonstrates its robustness in big networks by requiring less time to generate subgroups. To verify the effectiveness of the clustering used in the first section, the LEACH and PSO-based LEACH protocol is applied on a small network in the second portion of the study. According to the findings, PSO-based LEACH protocol performs better than LEACH protocol in terms of energy.

Qabouchi Ali et al. presented a novel approach to address the issue of energy efficiency in Wireless Sensor Networks (WSN). The authors propose a Particle Swarm Optimization (PSO)-based routing process that takes into account the distance of transmission and the residual energy of Cluster Heads (CHs) to optimize the overall routes to the Base Station (BS). The proposed approach optimizes the routing problem, which results in less traveled distance and reliability of links while avoiding CHs death during the process by relying on residual energies as a factor to select routes. The study demonstrates the effectiveness of the proposed technique by comparing it with existing techniques such as MHT-LEACH, DEEC, and LEACH. The results show that the proposed approach outperforms the existing techniques, achieving an improvement of 19.92% compared to MHT-LEACH, 64.46% compared



Fig.1Poor Clustering (large sum of distances)

The k-

means approach randomly introduces as many centroids as the specified number of k into the data the data points closest to a particular centroid will then be aggregated to form a cluster once distances between each data point and each centroid have been to DEEC, and more than 81% compared to LEACH. Moreover, the proposed technique stabilizes the network, reduces energy consumption, and improves the number of packets received [11].

Savita et al presented a study on energy-saving in sensor networks, which is a significant issue in various applications. They discuss how data aggregation and flooding can affect the efficiency of sensor nodes and propose clustering as a possible solution. However, to optimize this approach, there must be an optimal distribution of sensors and CHs to minimize energy consumption and improve the network's lifespan. The authors introduce a hybrid clustering technique that combines K-Means clustering and Particle Swarm Optimization (PSO) to enhance the performance of wireless sensor networks. They evaluate the proposed KPSO algorithm against traditional clustering techniques, including Mod-LEACH and K-means clustering, to demonstrate its effectiveness [12].

2. K-Means Clustering

For many real-world applications, the k-means approach has been emerged as an important tool in giving efficient clustering results [1]. The K-means algorithm is a popular data-clustering algorithm, but one of its drawbacks is the requirement for the number of clusters K, to be specified before the algorithm is applied [2]. The K-Means algorithm is widely used in industrial and scientific applications, but there are still weaknesses in the grouping algorithm. Weaknesses of the algorithm include determining the number of clusters based on assumptions and relying heavily on the initial selection of centroids to overcome this weakness [5]. The goal of K-means is to discover a grouping where each cluster has little variance within it, and to utilize the centroid of each cluster as a sample.

Objective: Create k groups for a given k such that the total of the (squared) distances between the mean of the groups and their components is as little as possible as shown in Figure 1 and Figure 2 [4, 13].



Fig.2 Optimal Clustering (minimal sum of distances) [13]

calculated after all the data has been merged with each cluster the centroids values will change depending on where each clusters data center is located the distance between each piece of data and the most recent centroids will then be calculated these steps will be repeated until the centroid values difference is too small it doesn't change after being updated or until a certain number of iterations **[14]**.

3. Optimum K Selection

Cluster cohesion and separation, the two key values required to determine the internal measure, are one approach to evaluate if a cluster was generated with the appropriate values in clustering validation. Sum square error (SSE) can be used to determine cluster cohesion and cluster separation. Cluster separation could be determined using "between cluster sum squares, (BSS)" whereas cluster cohesion can be assessed using "within cluster sum of square" (WSS) [14].

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} ||x - m_i||^2$$
(1)

Where, x is a data point in cluster *Ci*, *mi* is the center for cluster *Ci*.

Cluster Cohesion: Within cluster sum of squares (WSS=SSE):

$$WSS = \sum_{i=1}^{K} \sum_{x \in C_i} ||x - m_i||^2$$
(2)

Cluster Separation: Between cluster sum of squares (BSS):

$$BSS = \sum_{i=1}^{n} |C_i| \, ||m_i - m||^2 \tag{3}$$

Where Ci is the size of cluster i and m is the centroid of the data space.

3.1 Elbow Method

The elbow method could be used to examine the cohesiveness of a cluster of related data. The value of cohesiveness will ultimately reach a limit close to zero as the number of clusters is increased. The elbow method suggests that the graph has a large bend when K is at its best value [14]. The elbow technique determines the squared variation between several k values, the average distortion degree decreases and moves closer to the center of gravity as the k value increases, hence the elbow approach the elbow method is appropriate for relatively small k values [15].

3.2 Silhouette Method

Utilizing a silhouette coefficient (s) that blends cohesion and separation as given in Eq. (4), the silhouette method incorporates both. If the cohesiveness measure is greater than the separation, the silhouette coefficient is calculated by dividing the separation measure by the cohesion measure, subtracting the result by one, and then choosing the K with the greatest silhouette coefficient value [14].

$$s = 1 - \left(\frac{cohesion \ measure}{separation \ measure}\right)$$

If cohesion < separation

$$s = \left(\frac{\text{separation measure}}{\text{cohesion measure}}\right) - 1 \tag{5}$$

If cohesion > separation

4. Low Energy Adaptive Clustering Hierarchy (LEACH)

LEACH is a routing protocol for adaptive clustering. There are numerous rounds in the LEACH implementation process. The setup phase and the steady data transmission phase for each round. The cluster head nodes are selected at random from all the sensor nodes during the setup phase, and many clusters are built dynamically. Each cluster's member nodes transmit data to its respective cluster heads during the steady-state data transmission phase. The cluster head then compresses the member nodes' data and transmits it to the sink node. According to a round time, the LEACH protocol elects the cluster head nodes and re-establishes the clusters on a regular basis, which ensures energy dissipation of each node in the network is relatively evenly [16] [17]. Each node chooses a random integer between 0 and 1 at the start of the setup phase, and then determines an edge condition [18]. If the picked subjective number is less than the threshold number T(n), the node becomes a fortunate CH for such round. Threshold number T(n) is shown in Eq. (6)[19].

$$T(n) = \begin{cases} \frac{P}{1 - P \times (r \mod \frac{1}{P})} & \text{if } n \in G\\ 0 & \text{otherwise} \end{cases}$$
(

where P is the CHs desired percentage, r is the current round, and G is the competing nodes that was not chosen as CHs in the last 1/P rounds. Node gets to be CH for the current round if the number is not as much as limit T(n). When node is chosen as a CH, then it cannot get to be CH again until every one of the nodes of the group has gotten to be CH once. This is valuable for adjusting the energy utilization. Steady state is the second phase; non-CHs get the CH requests and after that send join demand to the CH advising that they are individuals from the group under that CH. During the steady-state phase, each sensor node aggregates and transmits data to its CH in perspective of the TDMA schedule. LEACH protocol employs TDMA/CDMA MAC to thwart inter-cluster and intracluster collision. Before sending the data to the BS, the CHs collect all the data. A certain amount of time later, the framework begins a new round by returning to the setup and reconfiguring the enduring state once more [20, 21].

5. Particle Swarm Optimization

Particle swarm optimization is one of the most effective approaches for swarm intelligence to tackle optimization problems. PSO's technique begins with a set of random particles and then iteratively updates generations to search for optima. Each particle is flown (4) through the search space, with its location updated based on its distance from both its personal best position and the swarm's best particle[22]. Each particle's

K-means Clustering and PSO Algorithm for Wireless Sensor Networks Optimization

performance, or how close it is to the global optimum, is evaluated using a fitness function that is dependent on the optimization issue. The PSO technique was used to identify near-optimal thresholds by minimizing the cross entropy between the each nodes [23]. The operating principles of the PSO algorithm are described in the equations below.

$$V_{ij^{k+1}} = WV_{ij^k} + c_1r_1(pbest_{ij^k} - X_{ij^k} + c_2r_2(gbest_{j^k} - X_{ij^{k+1}} - X_{ij^{k+1}} - X_{ij^k} + V_{ij^k}$$
(8)

Where V_{ij} is the particle velocity and X_{ij} is the position of a specific particle that is updated every iteration k. Each particle has a memory that stores its previous position lowest cost (*pbest*_{ijk}) and the best particle in the population (*gbest*_{ijk}), W representing inertia. c1 and c2 are assigned weights to the local and global best solutions, respectively. r1 and r2 are random values having a uniform distribution in the range [0,1] [23].

6. Proposed Approach

The Flowchart for the proposed algorithm is shown in Figure 3. At first two optimum and powerful methods

are used to choose the optimum K in the selected environment, then Leach and PSO algorithms have been applied differently to find their effectiveness. In the proposed approach, PSO aids in new cluster head selection, while normal nodes are assigned to cluster heads using the LEACH protocol without changing the cluster formation.

$$(7 X_{i,k})$$

Initially, the algorithm begins by identifying the optimal value of k through the execution and comparison of two established algorithms: Elbow and Silhouette. Subsequently, based on the algorithm that requires the least amount of time to complete, the preferred approach is selected. Following this, cluster head selection is carried out, with the process being conducted twice. The first iteration is conducted without utilizing Particle Swarm Optimization (PSO), while the second iteration involves the use of PSO with the initialization of both velocity and position and the evaluation of functions until the maximum iteration is reached. This process culminates in the generation of new cluster heads and a new data population



Fig.3 Flowchart of Proposed Approach

7. Results and Discussion

The simulation is run for three types of wireless sensor network distributed across 100m-by-100m area for different types of applications. Table 1 shows WSN types used in this study.

Table 1 WSN types is used in this study

Types of WSN	No. of Nodes	Applications
Small Sized Network	50 Nodes	eHealth Monitoring, Intelligent Home Ap- pliances
Medium Sized Network	500 Nodes	Smart Building Monitoring, Agriculture Production System
Enterprise Network	2500 Nodes	Advanced Industry, Intelligent transporta- tion system

The nodes are initially distributed across a 100m-by-100m area according to the networks in Table 1, as shown in Figures 4,5, and 6.



Fig.4 Small Sized Network Distribution Distributed Sensor Nodes Before Clustering



Fig.5 Medium Sized Network Distribution

Distributed Sensor Nodes Before Clustering



Fig.6 Enterprise Sized Network Distribution

The simulation result of small sized network to find optimal K by applying both Elbow and Silhouette Methods are in Figure 7(a) and Figure 7(b).







Based on both methods the clusters are formed as shown in Figure 8(a) and Figure 8(b)





Fig.8(a) Clusters formation for small sized network based on Elbow



Fig.8(b) Clusters formation for small sized network based on Silhouette





Fig.9(a) Optimum K using Elbow Method

Fig.9(c) Clusters formation based on Elbow



Fig.9(b) Optimum K using Silhouette Method



Fig. 9 Optimal k and cluster formation of medium sized network



Fig.10(a) Optimum K using Elbow Method



Fig.10(c) Clusters formation based on Elbow & Silhouette



Fig.10(b) Optimum K using Silhouette Method

Fig. 10 Optimal k and cluster formation of enterprise sized network

As demonstrated in Table 2, the Elbow technique for optimal k selection performs better in small and medium-sized networks while the silhouette method performs better in large networks with more than 2000 nodes based on the average length of time to build a cluster per k.

The average time to form K: it is the taken time for the algorithm to converge can depend on several factors, such as the complexity and distribution of the data, the number of data points, the initial placement of the centroids, and the chosen value of "k". In general, the algorithm can converge in a few iterations for small or well-clustered data sets, while for larger or more complex data sets, it may take more iterations.

Table 2 Average time to form a cluster per K

	Optimum K		Average time per K (Seconds)	
Network Size	Elbow	Silhouette	Elbow	Silhouette
Small	7	3	0.023858333	0.025054193
Medium	5	3	0.031215	0.038924593
Enterprise	4	4	0.081418318	0.067219344

To validate the effectiveness of the optimum methods and implementation of the cluster formation in real world scenarios, LEACH and PSO based LEACH protocols applied to small sized network using MATLAB environment, the simulation parameters are shown in Table 3.

Table 3 Simulation Parameters D				
no	Parameters	Specifications		
1	Size of the network	100m * 100m		
2	No. of sensor	50		
3	Initial Energy	0.5 J		
4	Location of the BS	center		
5	Packet data size	350 bytes		
6	No. of rounds	1500-2000		
7	Initial Energy	0.0001 J		
8	Data Aggregation Energy	5 nJ/bit/signal		

K-means Clustering and PSO Algorithm for
Wireless Sensor Networks Optimization

9	Power Transmission	50 nJ/bit
10	The Probability of CH	12%

The network lifetime parameters such as (First dead node FDN, Tenth dead Node TDN and All dead node AND), (total number of packets send to CH & sink



Fig.11(a) FND, TND and AND vs. no. of rounds



Fig.11(b) Total number of packets send to SINK node





node, life time of nodes) and (Total residual energy in the network and number of the dead nodes over the time) is determined as shown in Figure 11.



Fig.11(d) Alive nodes in each round



Fig.11(e) Total remaining energy in each round



Fig.11(f) Lifetime of the nodes in each round

Fig.11 LEACH and PSO based LEACH protocols applied to small sized network

The simulation results in Figure 11 reveals that the PSO based LEACH protocol sent more packets either to sink

node or CH compared to LEACH protocol. Also, there is a great number of dead nodes in LEACH protocol this

is due to less energy consumption of PSO based LEACH protocol. The combination of PSO and LEACH can be used to optimize the clustering performance and the energy efficiency of the wireless sensor network. Specifically, PSO can be used to optimize the parameters of the LEACH protocol, such as the cluster size, the threshold distance for cluster formation, and the number of clusters. This combination of PSO and LEACH can provide a more efficient and effective way to optimize the performance of wireless sensor networks.

Conclusion

In this paper optimal K selection for cluster formation to improve WSN network coverage is studied. Three separate network types are taken into consideration, each based on a set of real-world scenarios and applications. K means algorithm is widely used in image segmentation to divide the WSN network and form clustering for maximizing the lifetime of the network. As results calculating optimal number of K is the major

References

- 1. Tzortzis, G. and A. Likas, *The MinMax k-Means clustering algorithm*. Pattern recognition, 2014. **47**(7), 2505.
- Pham, D.T., S.S. Dimov, and C.D. Nguyen, Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2005. 219(1), 103.
- El-Awady, R., R. Abdel-kader, and M. Tobal, *Optimization Algorithm for Wireless Sensor Networks using K-means and Evolutionary Algorithm.* Port-Said Engineering Research Journal, 2013. 17, 146.
- 4. Wilks, D.S., *Statistical Methods in the Atmospheric Sciences*. 2006. **2nd Edition**.
- 5. Umargono, E., J.E. Suseno, and S.V. Gunawan. *K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula.* in *The 2nd International Seminar on Science and Technology (ISSTEC 2019).* 2020. Atlantis Press.
- 6. Shi, N., X. Liu, and Y. Guan, *Research on k*means Clustering Algorithm: An Improved kmeans Clustering Algorithm. 2010. 63.
- 7. Faid, A., M. Sadik, and E. Sabir, *IHEE: An Improved Hybrid Energy Efficient Algorithm for WSN*. 2021., 283.
- 8. Solaiman, B. and A. Sheta, *Energy* optimization in wireless sensor networks using a hybrid K-means PSO clustering algorithm. Turkish Journal of Electrical Engineering & Computer Sciences, 2015. **24**.
- 9. Sheta, A. and B. Solaiman, *Evolving a Hybrid K-Means Clustering Algorithm for Wireless*

goals of this paper to extend network lifetime and reduce energy consumption. Due to this, two of the optimal techniques are used in this study for their effectiveness and robustness in selecting number of clusters, Python environment is used to find optimal K and cluster formation. Simulation results shows that Elbow method is more applicable in small and medium sized networks compared to silhouette method, while silhouette results show its effectiveness in large network due to less amount of time required to form clusters. In the study's second part, LEACH and PSO based LEACH protocol is implemented on small sized network to validate the effectiveness of the cluster formation in real world scenario, The results show that PSO based LEACH protocol consumes less energy and the nodes transmit more packets during each round compared to LEACH protocol which confirms the theoretical operation of PSO protocol as one of the optimizations algorithms can be used accurately and efficiently. PSO and LEACH can be used to optimize the clustering performance and energy efficiency of wireless sensor networks.

> Sensor Network Using PSO and GAs. IJCSI International Journal of Computer Science Issues, 2015. **12**, 23.

- 10. Elhabyan, R.S. and M.C.E. Yagoub. *PSO-HC: Particle swarm optimization protocol for hierarchical clustering in Wireless Sensor Networks.* in 10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing. 2014.
- 11. Qabouche, H., et al. Energy Efficient PSObased routing protocol for large WSN. in 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET). 2022.
- 12. Jadhav, S. and S. Jadhav. KPSO: K-Mean and PSO based clustering algorithm for Wireless Sensor Network. in 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA. 2022.
- Lloyd, S., *Least squares quantization in PCM*. IEEE Transactions on Information Theory, 1982. 28(2), 129.
- 14. SAPUTRA, D.M., D. SAPUTRA, and L.D. OSWARI. Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method. in Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019). 2020. Atlantis Press.
- Cui, M., Introduction to the k-means clustering algorithm based on the elbow method. Accounting, Auditing and Finance, 2020. 1(1):.
- 16. Heinzelman, W.R., A. Chandrakasan, and H. Balakrishnan. *Energy-efficient communication* protocol for wireless microsensor networks. in

Proceedings of the 33rd Annual Hawaii International Conference on System Sciences. 2000.

- 17. Tandel, R.I. Leach Protocol in Wireless Sensor Network : A Survey. 2016.
- 18. Ramesh, K. and K. Somasundaram, A comparative study of clusterhead selection algorithms in wireless sensor networks. CoRR, 2012. abs/1205.1673.
- Al-Baz, A. and A. El-Sayed, A new algorithm for cluster head selection in LEACH protocol for wireless sensor networks. International Journal of Communication Systems, 2018. 31(1), e3407.
- 20. M.Nawar, N., et al., Enhancement of Mobility Model for Cluster Hierarchical Routing Protocol in Wireless Sensor Networks. International Journal of Computer Applications, 2014. **94**, 12
- 21. Li, Y., et al., Enhancing the performance of LEACH protocol in wireless sensor networks. 2011. 223.
- 22. Abualigah, L., et al., A Novel Evolutionary Arithmetic Optimization Algorithm for Multilevel Thresholding Segmentation of COVID-19 CT Images. Processes, 2021. 9(7): 1155.
- 23. DIVYA, B.N.G., Otsu's Method of Image Segmentation using Particle Swarm Optimization Technique. ijsetr, april-2015. 04(10), 1805.