ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 10, No. 3, June 2022, pp.276-285

K-Means clustering of optimized wireless network sensor using genetic algorithm

Azhar M. Kadim[1], Farah Saad Al-Mukhtar[2], Najwan Abed Hasan[3], Aseel B. Alnajjar[4], Mohammed Sahib Mahdi Altaei[5]
[1,2,3,4,5] Department of Computer science, College of Science, Al-Nahrain University, Baghdad, Iraq

ABSTRACT

Wireless sensor network is one of the main technology trends that used in several different applications for collecting, processing, and distributing a vast range of data. It becomes an essential core technology for many applications related to sense surrounding environment. In this paper, a two-dimensional WSN scheme was utilized for obtaining various WSN models that intended to be optimized by genetic algorithm for achieving optimized WSN models. Such optimized WSN models might contain two cluster heads that are close to each other, in which the distance between them included in the sensing range, and this demonstrates the presence of a redundant number of cluster heads. This problem exceeded by reapplying the clustering of all sensors found in the WSN model. The distance measure was used to detect handled problem, while K-means clustering was used to redistributing sensors around the alternative cluster head. The result was extremely encouraging in rearranging the dispersion of sensors in the detecting region with a conservative method of modest number of cluster heads that acknowledge the association for all sensors nearby.

Keywords: WSN, GA, K-means, clustering

Corresponding Author:

Azhar M. Kadim
Department of Computer science
Al-Nahrain University
Baghdad, Iraq
Email: azhar.mawlodkadim@nahrainuniv.edu.iq

1. Introduction

Wireless sensor networks (WSNs) depend on a co-operation of many small sensors stand with three other components to consist of a WSN system, these three components are: processor, transceiver, and a power source (i.e. battery). The function of such WSN system is to sense the surroundings and collect information about it. The information of surrounding area is obtained by sensors and then transferred toward the processor, which converts the analog information into digital form. Then, this converted information are transmitted by the transceiver into the gateway (i.e. base station) immediately or using intermediate sensors. The device of WSN system should be a less cost, less power and a system of embedded multi-functions. Recently, such embedded systems field is significantly grown, where an active WSNs studies are produced. The applications of WSN for monitoring the environmental surroundings are used in a wide variety fields, such as: medical, security, military and agricultural industries. It also used in monitoring the air pollution for detecting the concentration of dangerous gases, and so on the more and more applications especially that related to modern concept such as internet of things (IOT) [1]. Figure 1 shows many sensors are transmitting messages to their cluster head (CH) directly, such network design of direct transmission is very straightforward, but it is power-consuming due to the long distances from sensors to CH. Sensors that contributed in such model are classified into two types: high sensor range (HSR) and low sensor range (LSR). The search for alternative designs of WSNs based on the clustering is newly created to shorten the communication links and minimize the power consuming that leads to extend lifetime of the network [2]. Clustering has known to be successful in obtaining or obtaining important

© The Author 2022. This work is licensed under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/) that allows others to share and adapt the material for any purpose (even commercially), in any medium with an acknowledgement of the work's authorship and initial publication in this journal.

information more quickly. Because of its ability and effectiveness in clustering data, the K-means technique is indisputably popular among the numerous clustering algorithms that already exist [3],[4]. K-means is a non-hierarchical data clustering method that aims to partition existing data into one or much more clusters [5], boosting clustering centroids to obtain optimal clustering centers[6],[7]. Research results showed that the algorithm of K-means had no weakness except the dependency on the initial data, acceptable initial data leads to fast convergence with qualitative clustering [8]. In addition, heuristic methods likes GA had been applied with success to many combinatorial optimization problems that easily adapted to implement a routing algorithm for WSNs [9]. Where, WSNs optimization can be largely classified to single or multi-objective improvement issue. In single objective optimization the leading objective of the optimizer is to lminimize or maximize one goal under different requirements. while the multi objective optimization incorporates various targets are at the same time improved [10]. In many successful applications, GA is used with single or multi-objective modes of sensor network design that prompted to the improvement of a sundry other GA-based application explicit methodologies in WSN design [11]

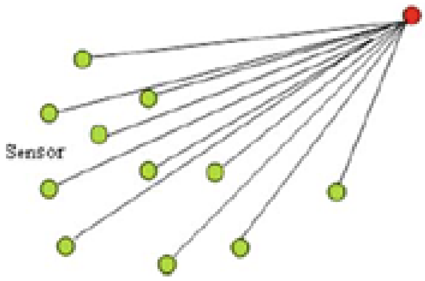

Figure 1. WSN clustering [2]

In [12], GA was utilized to tackle the problem of ad hoc WSNs distance optimization, in which the corresponding value over a conveyed sensor network was minimized in order to reduce the overall broadcaster's energy consumption. Over the fault tolerant network, GA was utilized to describe the placement of CH and network senses at the shortest possible communication distance. GA was used in [13] as a multi-objective optimization for self-organizing, modular wireless sensor network topology and energy management, factoring application explicit goals, communication limits, and conservation and efficiency attributes.. GA's optimized WSN satisfies all information standards, bypasses prevailing connectivity barriers, and combines heat features. Secondly, energy kept was upgraded to assure the network's maximum life expectancy while maintaining the organizational necessary elements by the application. In contrast,[14] developed a GA for the WSN problems considered as a multi optimization. Covariates, needed results, priorities, and parameters are indeed part of the GA-based WSN problem. A collection of requirements is also included to provide a more overview of the numerous needs that are taken into account when developing WSN enhancement issue. This article audit and investigate different helpful targets to show whether they clash with one another, support one another or they are design dependent. The results showed acceptable distribution for assumed WSN design.

2. Problem statement and contribution

The problem we address in this paper is a notice that has been observed in the WSN models resulting from the genetic algorithm, which may contain two heads that are close to each other and that the distance between them is smaller or equal to the amount of the sensing range, and this indicates the presence of an additional number of cluster heads. Therefore, our contribution is to replace these two cluster heads together with one cluster head in which all the sensors of the two cluster heads are connected, so that the sensing area becomes containing a least number of the cluster heads is enough to credit the connections of all sensors with the closest cluster head. Thus, the novelty of this work stands in the development the sensors clustering around CH by using k-means that applied on the optimal WSNs resulted from GA approach. This enables to overcome the problem of acceptability the initial centroids of k-means approach, and of course leads to improve the performance of the WSN as application-specific requirements.

3. Ga for WSN optimization

The implementation of GA includes successive stages within as displayed in Figure 2, they are: selection, crossover, and mutation until reach termination conditions and get the best value for the fitness function (F) in the last generation. First, each WSN model is encoded using predefined coding scheme as string to represents individual chromosome (Cij: where i is index refers to the chromosome in the population, and j is index refers to the generation), which is expressing a solution of WSN distribution problem. The reproduction process begins with the selection stage that based on F of each Cij, the Cij of higher fitness value has bigger probability of contributing offsprings to the next generation. Then, the crossover is process of mating two individuals from the last population at irregular in the intermarriage pool so a couple of offsprings is created later mutation is an irregular modification of chromosome position[15]. The evaluation of each offspring in the current generation is computed according to the fitness function. The procedure of regeneration is repeated for a particular number of generations or until no more improvement is observed. The best chromosome produced in last generation of the optimal resulted chromosome, which representing the solution for the problem of GA optimization [16]. Such that, the last generation shows final nodes locations and specific number of CH in the considered area of WSN[10].

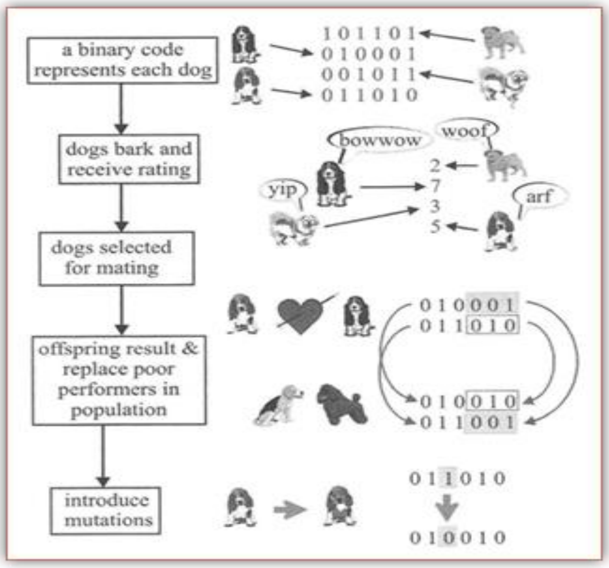

Figure 2. GA steps [17]

The location of a business is significant. WSN provides an estimate of sensor node gps devices, in which GA modifies the direction of sensor nodes with restoration, with the main objective of maximising availability with the fewest number of nodes and lowering overlapping area around vertices [18]. making use of GA The populous matrix is constructed once nodes are distributed haphazardly well within aim region, with each row denoting a project plan. The number of destination nodes specifies the amount of columns for every row. These words describe pseudorandom data transmission ranks: LSR or HSR, all with a different sensing field. As a results, the audience has chosen the optimization algorithm. The GA-based boost was impacted by diminishing energy-related aspects and increasing the accuracy of sensing areas [10]. Operating output (OE) and interaction energy (CE) are the two most fascinating resources characteristics, whereas sensor nodes per cluster nose error (SCE) and sensor systems out of range error are the two most interesting monitoring point equality parameters (SORE). The abbreviation OE denotes between how much energy a sensor uses during a specified operating life. It is mostly dictated by the device's mode, that is, whether it is a CH, HSR, or LSR gauge. [13] gives the OE waste quantity:

$$OE = \left(20 \times \frac{N_{CH}}{N}\right) + \left(2 \times \frac{N_{HSR}}{N}\right) + \left(\frac{N_{LSR}}{N}\right) \quad (1)$$

The number of CH, HSR, and LSR cameras in the connection is represented by NCH, NHSR, and NLSR, consecutively, while N is the sum of NCH, NHSR, & NLSR.

CE indicates to the energy utilization caused by contact between sensing devices and clusterheads in steady state conditions. It is relied somewhat on the intervals between these senses and their related anchor nodes, shown in [14]:

$$CE = \sum_{i=1}^{N_{CH}} \sum_{j=1}^{n_i} \mu . d_{ji}^{k} \quad (2)$$

Where ni is the number of systems throughout the boundary region, dji denotes the Centroids between sensor j and thus its clusterhead (ith cluster), and and k describe the design and utilization site of the WSN, appropriately. These variables are =1 and k=3 for a specific application precision of open continuous monitoring.

In standard procedure mode, SCE is used to insure that each cluster head does not have even a maximum set number of components, it's as follows. [19]:

$$SCE = \begin{cases} \frac{\sum_{i=1}^{N_{full}} n_i}{N_{full}} & if\ N_{full} > 0 \\ 0 & Otherwise \end{cases} \tag{3}$$

Where, Nfull is the number of cluster heads (or clusters) that have more than 15 active sensors in their clusters and ni is the number of sensors in the ith cluster (the limit of 15 active sensors is assumed as default value). Thus, SCE describe the actual communication capacities of the sensors as well as their data administration abilities as far as amount of information that can be handled by a cluster head sensor. Whereas, the SORE is utilized to guarantee that every sensor can communicate with its cluster head which essentially relies upon the signal range ability of the sensor. By assuming the HSR sensors cover a round region with radius of 10 length units and LSR sensors cover a roundabout region with radius of 5 length units, then the quantity of active sensors that can't communicate with their cluster head (Nout) comparative with N demonstrates the all out number of active sensors in the network as follows [14]:

$$SORE = \frac{N_{out}}{N} \tag{4}$$

The employment of relations given in Eq.s (1-4) to form suitable objective function throughout the numeric evaluation of the fitness function of GA enable to give high quality to every conceivable solution of the enhancement issue[13],[14].

4. Proposed method

The proposed method incorporates three stages: WSN modeling, WSN modeling optimization, and WSN sensors clustering. The WSN modeling is based on the assuming a specific form of the WSN model that can be used theoretically in later computer processing. While the WSN optimization depends on the application of GA on the initial WSN models in order to produce the optimal WSN models that represent the last developed generation. Later, the clustering process is applied to the optimal models that produced by GA in order to reduce the number of cluster heads and achieving WSN models containing a least number of cluster heads in which the levels of expended energy are as low as possible. The following subsections present more details about the stages of the proposed method:

5. WSN modeling

The used WSN model is a two-dimensional square system of 30×30 length unit containing a certain number of cluster heads (NCH) that equal to 10% of the model area (i.e. 90 cluster heads). While, the number of HSR sensors is twice the number of cluster heads (i.e. 180 sensors), and the number of LSR sensors is four times the number of cluster heads (i.e. 360 sensors). All of these sensors are located in the sensing area. This number of cluster heads is randomly distributed in the sensing area and away from the edges, as well as the two types of sensors (HSR and LSR) are also randomly distributed and connected to the nearest cluster head. The uniformity of distributing these sensors along the sensing area is calculated through an objective function that depends on the sum of four terms, each term representing one function given in equations (1, 2, 3 and 4). Figure 3 represents one WSN model computed based on the above assumed conditions.

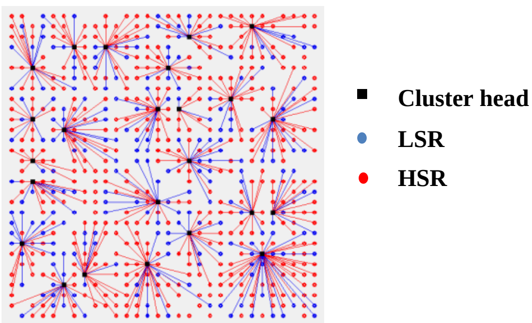

Figure 3.WSN model assumed in the present work

Table 1 shows the computed values of the four objective functions adopted in estimating the quality of WSN design displayed in Figure 3. From observing the values of these four objective functions, it becomes clear that their scales are different from each other, and to credit that the results are not influenced to any function greater than others, it is necessary to calibrate their outputs within the range (0-1) by dividing their outcomes on the achieved maximum value that the function can obtain. The maximum values of these functions are computed and given in the third column, while the fourth column presents the achieved normalized values of these function.

Table 1. Objective functions values of WSN model given in Figure 3

Objective Function	Computed Value	Highest Value	Normalized value
OE	5.837	6.713	0.869
CE	9.529	26.472	0.359
SCE	0.016	0.729	0.021
SORE	87	108	0.805

4.1 WSN models optimization

The used optimization technique depends on GN, which manages the first generation of WSN population that addressing the population of the initial generation. In GA, there are 50 pairs of population individuals are randomly chosen for crossover with one another to produce 100 new individuals from the population of the next generation. The crossover process is a single substitution of one half parent with the other half of the second parent, as displayed in Figure 4. Also, the process of genetic mutation happens sometimes within 0.001 occurrence probability, in which a small number of no more than 0.05 of the sensors are changed to another type not their type as shown in Figure 5. The selection of individuals for generating next generations is based on the fitness function (F) given in eq.(*), the candidate individual are that of lowest F. The process of producing new generations continues until reaching the termination condition which consists of two parts, they are: there is no noticeable change in the fitness function with more generations is happen, or the reaching to 10,000 successive generations.

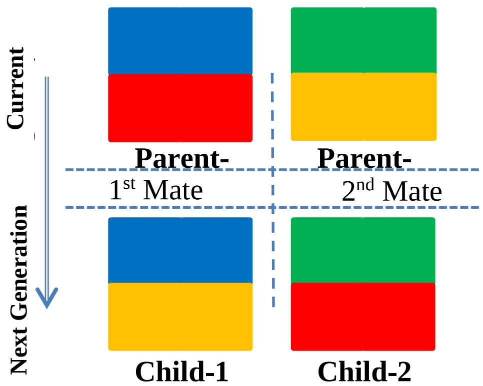

Figure 4. Crossover of individual pairs to produce four children in the next generation

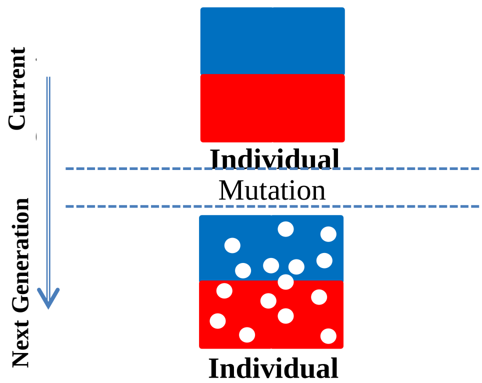

Figure 5. Mutation of one individual to produce new one in the next generation

4.2 Optimized WSN clustering

The use of the clustering process for achieving best WSN models from the optimal ones is useful in decreasing the number of cluster heads in WSN model as little as possible. This of course prompts a decrease in the amount of operating energy and consumed energy, which making the fitness function for each WSN model as little as could be expected. The clustering process is applied to each optimal WSN model independently, as it depends on the hypothesis of compensating each two cluster heads that are near one another with a distance equivalent to or less than the sensitivity range of the LSR sensor supposedly with one cluster head located in the middle of the distance between them to connect the sensors close to it and within allowed ranges. This requires looking for any two cluster heads in the optimal WSN to which the above situation applies to replace them with one cluster head. The number of cluster heads and their positions in the optimal WSN model is already known, so it is simple to find these cases and handle them according to the adopted hypothesis. Figure 6 shows the case of two cluster heads close to each other with a distance d is less than the sensing range, and that the first cluster head (CH1) relates to the sensors (L11, L12, H11 and H12), while the second cluster head (CH2) is connected to the sensors (L21, L22, H21, H22 and H23). When these two cluster heads are replaced by one, they are deleted from the list of the cluster heads of that WSN model and an alternative cluster head is created instead of them (i.e. CH), which is located at the halfway between them. Then, the sensors (L11, L12, H11 and H12) that were associated with the first deleted cluster head are located at the allowed sensing region of the CH, and therefore are associated with it. Also, the sensors (L21, L22, H21, H22 and H23) that are connected to the second deleted cluster head are connected to the new created cluster head CH because they are within the sensing region of it. Whereas the sensor H21 cannot be connected to the new cluster head because the distance between them has become larger. Such that the sensor H21 connected to other cluster head is close to it. through the sensing area.

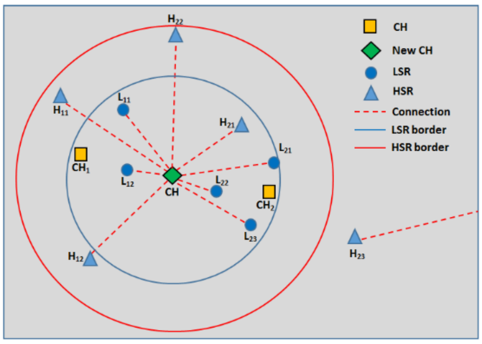

Figure 6. Clustering approach used in the proposed method

6. Results and discussions

The two-dimensional WSN model was used to obtain several WSN models, each containing 90 cluster heads located in random places, as well as the model contains the assumed number of HSR and LSR sensors located randomly and connected to the nearest cluster head within the sensing area. Figure 7 shows two models of such obtained models, and it appears that the cluster head occupied positions of different cases, they are convergent at one time and divergent at another, but it is certain that all of them are far from the edges of the distribution area.

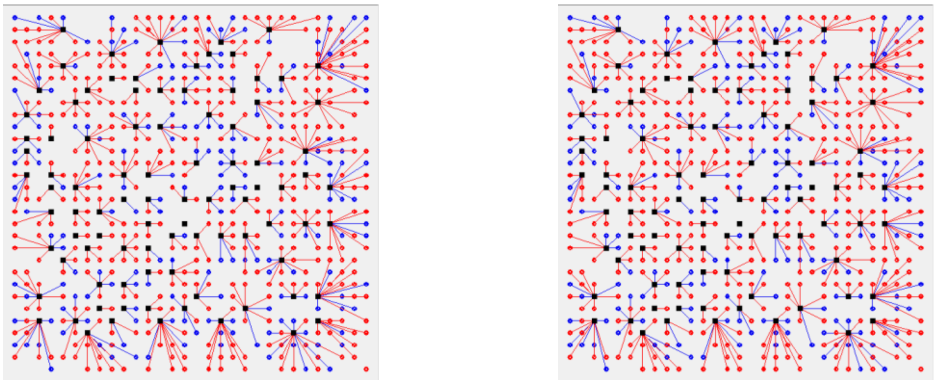

Figure 7. Two-dimensional WSN models obtained from the random generation method

The GA was applied to a population of created WSN models like that shown in Figure 7. The number of populations was 100 individuals for each generation, and half of them were randomly chosen to perform the mating by substitution of genes of each half between the parent pair. Each pair of the first generations produce two individuals from the subsequent generations, and therefore the number of individuals in successive generations remains constant. The genetic mutation rarely occurs and leads to a change of the type of gene (sensor) to another type. Both the mating and genetic mutation stages lead to a change the number of genes for each type, and therefore generations are indirectly going towards improving the genes when choosing the best individuals for mating. Figure 8 shows two models of WSN belonging to the last generation after applying the

GA. It is noticeable that the grouping of sensors around the cluster head in Figure 8 became better than they were in Figure 7.

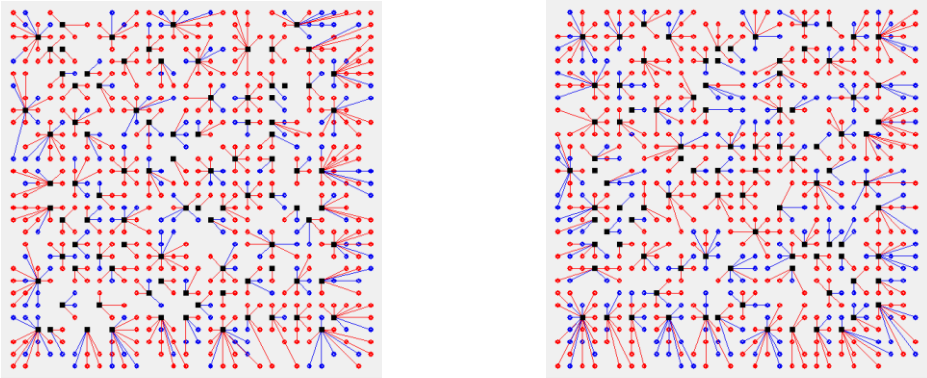

Figure 8. Optimized 2D WSN models obtained from GA

It is noticeable that the number of cluster heads in the optimized WSN models has changed slightly, as its number was 90 when it was generated, while it became little less or more after applying the GA. The change in the number of cluster heads is due to the two processes of single crossover and mutation that replace genes with others. It was found that the number of cluster heads in the optimal WSN models in the last generation is less than that of the first generations, in which the sensor was uniformly distributed as shown in Figure 9. This confirms the improvement happen in the WSN models due to the GA.

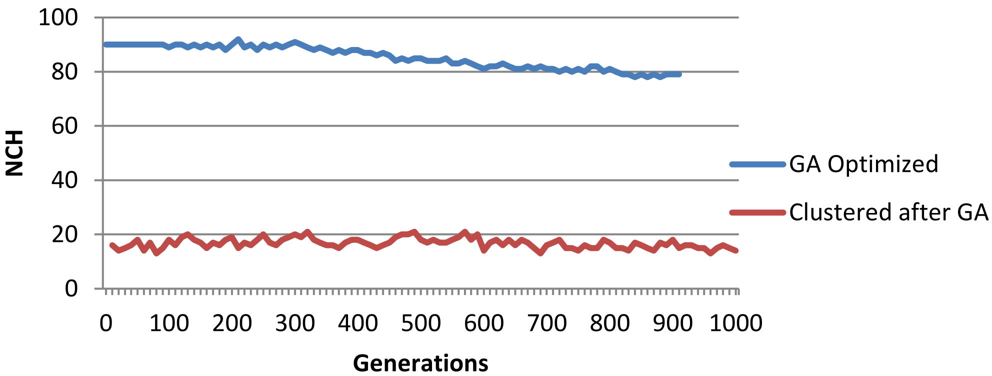

Figure 9. Number of cluster heads count in successive generations

The optimal WSN models resulting from GA are actually optimal with respect to the assumed conditions for their generation. Because the generation process was random and with a specific number of sensor types, it is found that the improvement was also limited. Such improvement has reduced the number of cluster heads by a small amount relative to their initial number. Therefore, the searching for two closely spaced cluster heads with less than the sensing range is indicating that these cases are consuming energy and cost. This case was treated using the clustering approach that merging these two nearby cluster heads together and replacing them with one cluster head, then connecting all their sensors with the nearest cluster head. The use of this method gave very promising results in reorganizing the distribution of sensors in the sensing area with an economical way of small number of cluster heads that credit the connection with all sensors in the area.

Figure 10 shows the results of the clustering process that substitutes the additional cluster heads. Thus, it is found that the number of cluster head has become few and has reached the limit that ensures that all sensors are connected to closest cluster head. These results are greatly reducing the amount of consumed energy as indicated in Figure 11, which describes the average amounts of OE and CE functions for the same two models given in Figures 9 and 10. Therefore, the amount of energy consumption has decreased a lot after applying the clustering process compared to its predecessors, and this confirms the success of the clustering process that was applied to the optimal WSN models to get the best ones.

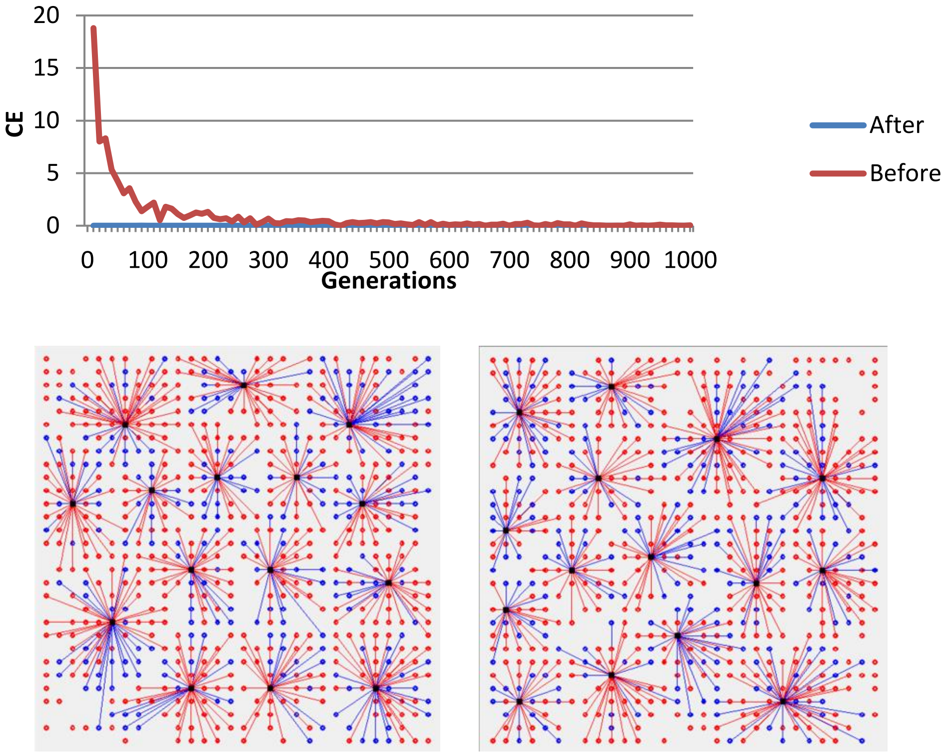

Figure 10. Results of applying the clustering on optimal WSN models

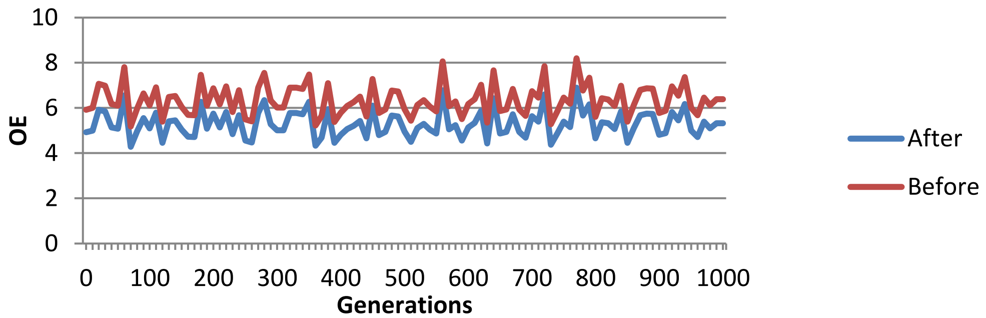

Figure 11. Energy consumption of optimal WSN models before and after clustering

7. Conclusions and suggestions

The random generation of WSN gives unreliable models and improving their performance by redistributing them based on GA was restricted and did not allow access to obtain ideal model. The sensors redistribution by k-means applied on GA based resulted WSN models had given better results, in which the number of cluster heads has been reduced to the lowest value that guarantees the connection availability for all sensors, and the value of the consuming power was lower, which indicates a low cost of building such systems.

Declaration of competing interest

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

References

[1] P. Nayak, K. Kavitha and N. Khan, "Cluster Head Selection in Wireless Sensor Network Using Bio-Inspired Algorithm," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), 2019, pp. 1690-1696, doi: 10.1109/TENCON.2019.8929440.

[2] A. P. C. a. H. B. W. B. Heinzelman, "An application-specific protocol architecture for wireless microsensor networks," IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 660-670, Oct. 2002, doi: 10.1109/TWC.2002.804190.

[3] Y. Cao, J. Liu and Z. Xu, "A Dynamic Multiagent Genetic Algorithm for Optimal Charging in Wireless Rechargeable Sensor Networks," 2021 IEEE Congress on Evolutionary Computation (CEC), 2021, pp. 1856-1863, doi: 10.1109/CEC45853.2021.9504883.

[4] M. Anoop and P. Sripriya, "Ensembled Adaptive Fuzzy K-Means With Stochastic Extreme Gradient Boost Big Data Clustering on Geo-Social Networks," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 498-502, doi: 10.1109/ICACITE51222.2021.9404574.

[5] Z. Qu, C. Liu and H. Liu, "Deep neural network localization algorithm based on GA optimization K-means," 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), 2021, pp. 1065-1069, doi: 10.1109/AEECA52519.2021.9574133.

[6] K.-L. Wu and M.-S. Yang, "Alternative c-means clustering algorithms," Pattern recognition, vol. 35, no. 10, pp. 2267-2278, 2002.

[7] B. K. Khotimah, F. Irhamni, and T. Sundarwati, "A Genetic algorithm for optimized initial centers K-means clustering in SMEs," Journal of Theoretical and Applied Information Technology, vol. 90, no. 1, p. 23, 2016.

[8] M. Feng and Z. Wang, "A genetic k-means clustering algorithm based on the optimized initial centers," Computer and information science, vol. 4, no. 3, p. 88, 2011.

[9] B. Ghimire, A. Mahmood and K. Elleithy, "Hybrid Quantum Approximate Optimization Using Enhanced Ant Colony Optimization to Solve Large-Scale Combinatorial Optimization Problems," *2021 8th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, 2021, pp. 108-113, doi: 10.1109/ISCMI53840.2021.9654824.

[10] M. N. Iqbal, Muhammad; Anpalagan, Alagan; Ahmed, Ashfaq; Azam, Muhammad, "Wireless Sensor Network Optimization: Multi-Objective Paradigm," Journal contribution, 2015, doi: https://doi.org/10.32920/14640045.v1.

[11] S. Sen, S. Narasimhan, and K. Deb, "Sensor network design of linear processes using genetic algorithms," Computers & chemical engineering, vol. 22, no. 3, pp. 385-390, 1998.

[12] M. Al-Obaidy, A. Ayesh, and A. F. Sheta, "Optimizing the communication distance of an ad hoc wireless sensor networks by genetic algorithms," Artificial Intelligence Review, vol. 29, no. 3, pp. 183-194, 2008.

[13] K. P. Ferentinos and T. A. Tsiligiridis, "Adaptive design optimization of wireless sensor networks using genetic algorithms," Computer Networks, vol. 51, no. 4, pp. 1031-1051, 2007.

[14] M. Iqbal, M. Naeem, A. Anpalagan, A. Ahmed, and M. Azam, "Wireless sensor network optimization: Multi-objective paradigm," Sensors, vol. 15, no. 7, pp. 17572-17620, 2015.

[15] "What is a Genetic Algorithm?, ." [Online]. Available: https://www.generativedesign.org/02-deeper-dive/02-04_genetic-algorithms/02-04-01_what-is-a-genetic-algorithm.

[16] A. Pugazhenthi and L. S. Kumar, "Selection of Optimal Number of Clusters and Centroids for K-means and Fuzzy C-means Clustering: A Review," *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, 2020, pp. 1-4, doi: 10.1109/ICCCS49678.2020.9276978.

[17] ""Analogy between a numerical GA and biological genetics", ." [Online]. Available: https://www.semanticscholar.org/paper/The-Binary-Genetic-Algorithm-Haupt-Haupt/fae177cd739db7d386ab655b3481bc5c33202742/figure/0.

[18] H. Z. M. B. M. E. H. A. Ajith Abraham, "An improved dynamic deployment technique based-on genetic algorithm (IDDT-GA) for maximizing coverage in wireless sensor networks," Vol.:(0123456789)1 3Journal of Ambient Intelligence and Humanized Computing, 14 January 2020, doi: org/10.1007/s12652-020-01698-5.

[19] K. P. Sinaga, I. Hussain and M. -S. Yang, "Entropy K-Means Clustering With Feature Reduction Under Unknown Number of Clusters," in *IEEE Access*, vol. 9, pp. 67736-67751, 2021, doi: 10.1109/ACCESS.2021.3077622.