Computers, Materials & Continua CMC, vol.64, no.3, pp.1725-1739, 2020

A Nonuniform Clustering Routing Algorithm Based on an Improved K-Means Algorithm

Xinliang Tang[1], Man Zhang[1], Pingping Yu[1], Wei Liu[2], Ning Cao[3,*] and Yunfeng Xu[4]

Abstract: In a large-scale wireless sensor network (WSN), densely distributed sensor nodes process a large amount of data. The aggregation of data in a network can consume a great amount of energy. To balance and reduce the energy consumption of nodes in a WSN and extend the network life, this paper proposes a nonuniform clustering routing algorithm based on the improved K-means algorithm. The algorithm uses a clustering method to form and optimize clusters, and it selects appropriate cluster heads to balance network energy consumption and extend the life cycle of the WSN. To ensure that the cluster head (CH) selection in the network is fair and that the location of the selected CH is not concentrated within a certain range, we chose the appropriate CH competition radius. Simulation results show that, compared with LEACH, LEACH-C, and the DEEC clustering algorithm, this algorithm can effectively balance the energy consumption of the CH and extend the network life.

Keywords: WSN, node energy consumption, nonuniform clustering routing algorithm.

1 Introduction

With the advent of the information society, WSNs have become a popular area of research. Such networks usually consist of a large number of sensor nodes and have low energy consumption and low costs [Akila and Venkatesan (2018)]. WSNs have been applied in many fields, including environmental monitoring, target recognition, and national defense and military. Yu et al. [Yu, Zhong, Xiao et al. (2019); Li, Cui and Liu (2018)] However, the limited energy resources and harsh environment bring great challenges to the application of WSNs [Meng and Zhang (2018)].

Sensor nodes are usually powered by batteries, which have limited capacity and high replacement costs. Therefore, the insufficient energy supply severely limits the network

[1] School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, 050000, China.

[2] Hebei Chemical and Pharmaceutical College, Hebei Seismological Bureau, Shijiazhuang, 050000, China.

[3] School of Internet of Things and Software Technology, Wuxi Vocational College of Science and Technology, Wuxi, 214028, China.

[4] Rutgers Business School-Newark, Washington Park, Newark, NJ 07102, USA.

[*] Corresponding Author: Ning Cao. Email: ning.cao2008@hotmail.com.

Received: 26 February 2020; Accepted: 24 April 2020.

life of the WSNs, which is an important problem in this field [Sundararajan and Arumugam (2018)].

For WSNs with a large number of nodes, the clustering algorithm can effectively improve the network life, and in this way, the requirements of network load balancing can be met [Sohrabi, Gao, Ailawadhi et al. (2000); Sirsikar and Wankhede (2015); Liu (2015); Geetha and Sankar (2018)]. The LEACH clustering algorithm selects a CH based on the concept of a "wheel". However, the selection of the CH is random, which leads to unbalanced CH energy consumption in the whole WSN. In view of the shortcomings of the LEACH protocol, many scholars have done substantial research and proposed many improvement methods. In the literature Heinzelman et al. [Heinzelman, Chandrakasan and Balakrishnan (2002)] proposed the LEACH-C algorithm based on LEACH. This algorithm combines energy cluster routing and media access with application-specific data aggregation and achieves good performance in terms of system life, delay and application-aware quality. However, only two indexes of residual energy and communication cost are considered in the selection of the CH node. The BEEC clustering algorithm Liu et al. [Liu and Hu (2014)] introduced a competition mechanism in the final stage of CH selection. In the cluster construction stage, only the distance of the CH node is considered. The energy consumption of the whole WSN due to unreasonable transmission path planning is not considered. The selection of the CH in the DEEC clustering algorithm [Qing, Zhu and Wang (2006)] is based on the ratio of the remaining energy of the node to the average energy of the network. The algorithm needs to calculate the average energy of the network. The DSBCA algorithm Liao et al. [Liao, Qi and Li (2013)] calculates node weights based on residual energy, node connectivity density and election time parameters, which leads to a more balanced cluster structure and extends the life cycle of the network. However, the CH in the DSBCA algorithm still communicates directly with the base station (BS), which causes the CH to run out of energy quickly. Moreover, the change in the CH is carried out in the cluster only, resulting in unbalanced energy consumption. Another study Hou et al. [Hou, Song and Zhou (2015)] proposed a LEACH optimization clustering algorithm. The residual energy of nodes is considered in CH selection, which increases the probability that nodes with high residual energy will become CHs, thereby improving the energy utilization efficiency of the CH and enhancing the reliability of the network. However, multihop data transmission is adopted between CHs in this algorithm. The CH of the adjacent BS is exhausted prematurely due to the additional data relay tasks. Another study Chang et al. [Chang and Zhang (2016)] proposed an improved nonuniform clustering routing algorithm for calculating WUCH. The algorithm USES applies a different competition radius to produce heterogeneous clusters. The optimal node within the competition radius is called a CH. However, it is difficult to determine the optimal parameters when calculating the competition radius. To deal with the possible uncertainties in the calculation of the competition radius of the CH, one study [Bagci and Yazici (2013)] proposed a fuzzy energy-sensing heterogeneous clustering (EAUCF) algorithm. The EAUCF algorithm inputs the residual energy and the distance between the node and BS into a fuzzy inference system. The election radius of the node is calculated. The node with the highest residual energy in the radius of the election becomes the CH. The simulation results show that the proposed algorithm can effectively extend the network life cycle. However, when the EAUCF algorithm is close to the BS node, the CH close to the BS consumes energy very quickly. A study Zhang et al. [Zhang and Cao (2010)]

proposed a heterogeneous algorithm based on ant colony optimization for the real-time optimization of WSNs. To make the routing in the network more self-organizing and adaptive, the reliability indexes are considered in cluster communication. By prolonging the network life cycle, the packet loss rate is reduced. However, the whole calculation cost of the algorithm is large. Another study Jiang et al. [Jiang, Shi and Tang (2012)] selected CHs according to the time mechanism in heterogeneous clustering to reduce the redundancy of the message in the network topology division stage and the communication energy consumption. In the multihop routing between CHs, a greedy algorithm is used to select the transit nodes to balance the energy consumption of the network. However, the cluster construction in the algorithm is based mainly on the distance factor and does not consider the real-time energy factor of nodes.

Based on research aiming to reduce the energy consumption of WSNs and the existing disadvantages of clustering routing, this paper proposes a nonuniform clustering routing algorithm based on the improved K-means algorithm. The algorithm reduces the energy consumption and improves the unbalanced energy consumption. The rest of this article is arranged as follows. Section 2 presents the deployment model and node energy consumption model in wireless sensors. Section 3 mainly introduces the clustering algorithm based on improved K-means clustering, CH selection and data forwarding. Section 4 compares the advantages and disadvantages of several algorithms through simulation experiments. Finally, Section 5 gives a brief summary of this paper.

2 Correlation model

2.1 Deployment model

The scenario deployment model [Emad and Lon (2018); Wang, Qi and Liu (2018)] is composed of a BS and multiple sensor nodes deployed in a region. As shown in Fig. 1, the sensor node (Nn) first collects data from the monitoring area to the CH, and then the cluster head sends the data to the BS. The locations of the nodes are randomly fixed, and each sensor node is the same. All sensor nodes have the same initial energy. When the initial energy is exhausted, the sensor node dies. The power of the BS is infinite. The sensor node can adjust the transmission power according to the transmission distance.

Figure 1: Scenario deployment model

2.2 Node energy consumption model

The energy consumption of sensor nodes mainly consists of communication energy consumption and data fusion energy consumption. In this paper, the classic wireless communication energy consumption model simplified by Heinzelman et al. [Heinzelman, Chandrakasan and Balakrishnan (2002)], as shown in Fig. 2, is adopted for the communication energy consumption of sensor nodes.

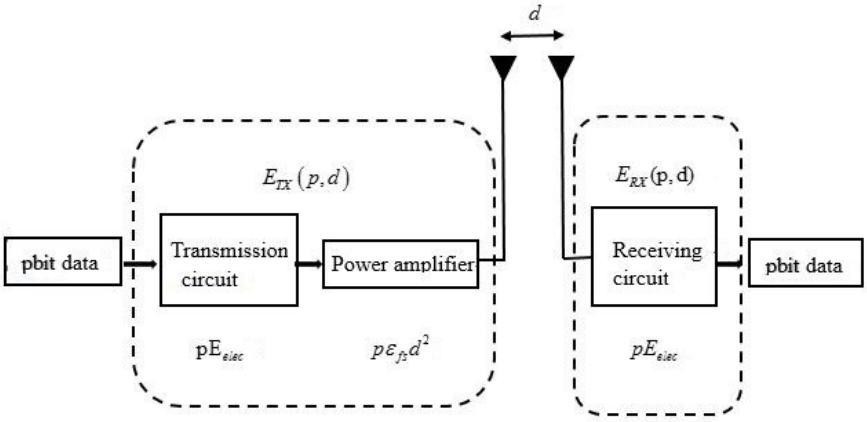

Figure 2: Communication energy consumption model of sensor nodes

The energy consumed during communication between sensor nodes includes the energy consumed when sending data, $E_{TX}(p,d)$, and the energy consumed when receiving data, $E_{RX}(p,d)$. The energy consumption of pbit data sent by the node is:

$$E_{TX}(p,d)\begin{cases} p\cdot E_{elec}+p\cdot\varepsilon_{fs}\cdot d^2 & if \quad d\le d_0 \\ p\cdot E_{elec}+p\cdot\varepsilon_{amp}\cdot d^4 & if \quad d>d_0 \end{cases} \tag{1}$$

The energy consumption of pbit data received by the node is:

$$E_{RX}(p,d)=p\cdot E_{elec} \tag{1}$$

In the above equation, E_{elec} represents the energy consumption of sending and receiving pbit data, $E_{elec}=50nJ/bit$. In addition, ε_{fs} and ε_{amp} are the power amplifier energy consumption coefficients in the free space channel and multipath fading channel, respectively; $\varepsilon_{fs}=10pJ/\left(bit\cdot m^2\right)$, and $\varepsilon_{amp}=1.3\times10^{-3}pJ/\left(bit\cdot m^4\right)$. The transmission distance between nodes is d. The critical distance between the two channel models is $d_0=\sqrt{\varepsilon_{fs}/\varepsilon_{amp}}=87m$.

Data fusion energy consumption is the energy consumed by the CH for the fusion processing of the data sent by all member nodes. The energy consumption model of data fusion in this paper sets the energy consumption of the CH per fused unit bit of data as $E_{da}=5\times10^{-9}J$.

3 Algorithm improvement design

3.1 Clustering algorithm based on improved K-means clustering

When the K-means algorithm is adopted to randomly select nodes in the monitoring area, the clustering center is randomly selected, which directly affects the final clustering result. Moreover, the iteration time is long. At most, the convergence of the algorithm can be guaranteed to the fixed point only, and the convergence results cannot be guaranteed to be complete. Therefore, the traditional K-means clustering algorithm is adopted in the WSN. The algorithm randomly selects k clustering points and iteratively forms clusters. This approach is simple and easy to implement. However, the randomness of the initial clustering points is high, the iteration time is too long, and local optimal solutions are easily obtained. When using the K-means algorithm for clustering, reasonably determining the initial clustering center is an important step in the realization of optimal clusters. Therefore, this paper proposes an improved K-means initial clustering point selection method. This method makes up for the problems existing in the traditional K-means algorithm to a large extent. For example, the least mean squares algorithm and the randomness of the clustering initialization algorithm exist. The steps of the clustering center node selection method proposed in this paper are shown in Fig. 3.

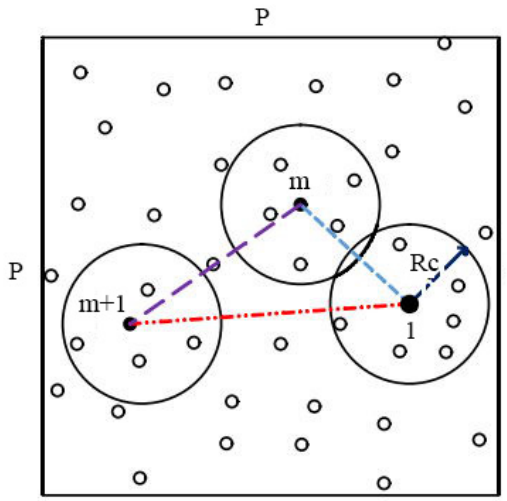

Figure 3: Schematic diagram of initial clustering center point selection

a) Determine the number of initial clusters

In WSNs, the number of clusters plays an important role in the performance of the whole network. If the number of clusters is too small, then more nodes are needed to transmit data to the remote CH. If there are too many clusters, then there will be very few data fused in the cluster, which reduces the advantage of network clustering. Therefore, if the number of network clusters is too small or too large, then network energy consumption grows exponentially. Thus, there is theoretically an optimal CH number K_{CH}.

It is assumed that N sensor nodes are randomly distributed in the perception area of P×P. If there are K_{CH} clusters, then there is an average of N/K_{CH} nodes in each cluster, and each cluster contains a CH and N/K_{CH}-1 member nodes. The CH needs to receive and fuse the data sent by member nodes. The merged data are then sent to the BS. Therefore, the energy consumed by CH in each round of CH selection includes the energy consumed by receiving the data of N/K_{CH}-1 member nodes, the energy consumed by N/K_{CH}-1 nodes within the fusion cluster, and the energy consumed by sending the fused data to the BS, and it is expressed as follows:

$$E_{CH}=\left(\frac{N}{K_{CH}}-1\right)\bullet pE_{elec}^{Rx}+\frac{N}{K_{CH}}\bullet pE_{da}+\left(pE_{elec}^{Tx}+p\varepsilon_{mp}d_{to-BS}^{n}\right) \tag{2}$$

The packet length transmitted by each node is p. The average distance between the CH and BS is d_{to-BS}. The free space energy consumption model is usually used when member nodes transmit data to the CH. The energy consumed by member nodes in each round is:

$$E_{non-CH}=pE_{elec}+p\varepsilon_{fs}d_{to-CH}^{2} \tag{3}$$

The distance between the member node and the CH is d_{to-CH}. The average area of each cluster is P^2/K_{CH}. The CH density is $\rho=K_{CH}/\mathrm{P}^2$. Assuming that the CH is located at the center of the cluster, the radius r of the cluster is approximately $P/\sqrt{\pi K_{CH}}$. The expected distance squared between the member nodes in the cluster and the CH $E\left(d_{to-CH}^{2}\right)$ is:

$$E\left(d_{to-CH}^{2}\right)=\iint r^2\rho(r,\theta)rdrd\theta=\rho\int_0^{2\pi}\int_0^{\frac{P}{\sqrt{\pi K_{CH}}}}r^3drd\theta=\frac{P^2}{2\pi K_{CH}} \tag{4}$$

Substitute the above formula into Eq. (6), the energy consumed by member nodes in a single cluster in each round can be obtained as follows:

$$E_{non-CH}=lE_{elec}+l\varepsilon_{fs}P^2/\left(2\pi K_{CH}\right) \tag{5}$$

The energy consumed in a single cluster within a round is:

$$E_{cluster}=E_{CH}+\left(\frac{N}{K_{CH}}-1\right)E_{non-CH} \tag{6}$$

Therefore, the energy consumption of the entire network in each round is:

$$E_{\text{round}}=K_{CH}E_{\text{cluster}}=2\mathrm{NpE}_{\text{elec}}+NpE_{DA}+K_{CH}\left(l\varepsilon_{\text{mp}}d_{\text{to-BS}}^{n}-pE_{\text{elec}}\right)+\frac{Np\varepsilon_{fs}P^2}{2\pi K_{CH}}-\frac{p\varepsilon_{\text{fs}}P^2}{2\pi} \tag{8}$$

The first and second derivatives of the above equation E_{round} are obtained for K_{CH} as follows:

$$\frac{\partial E_{\text{round}}}{\partial K_{CH}}=l\varepsilon_{\text{mp}}d_{\text{to-BS}}^{n}-pE_{\text{elec}}-\frac{Np\varepsilon_{fs}P^2}{2\pi K_{CH}^2} \tag{7}$$

$$\frac{\partial^2 E_{\text{round}}}{\partial^2 K_{CH}} = \frac{Np\varepsilon_{fs}P^2}{\pi K_{CH}^3} > 0 \tag{8}$$

According to Eq. (11), the second derivative is greater than 0. When the first derivative is 0, the optimal number of network clusters is:

$$K_{CH} = \sqrt{\frac{N}{2\pi}} \cdot \sqrt{\frac{\varepsilon_{fs}P^2}{\varepsilon_{mp}d_{to-BS}^n - E_{elec}}} \tag{9}$$

According to different communication channel models adopted by network nodes, the optimal number of network clusters is:

$$K_{CH} = \begin{cases} \sqrt{\frac{N}{2\pi}} \cdot \sqrt{\frac{\varepsilon_{fs}P^2}{\varepsilon_{mp}d_{to-BS}^2 - E_{elec}}} & d_{to-BS} \le d_0 \\ \sqrt{\frac{N}{2\pi}} \cdot \sqrt{\frac{\varepsilon_{fs}P^2}{\varepsilon_{mp}d_{to-BS}^4 - E_{elec}}} & d_{to-BS} > d_0 \end{cases} \tag{10}$$

b) Calculation of the collection of adjacent nodes

The set of n nodes in the P×P monitoring area is expressed as $E = \{e_1, e_2, e_3, \ldots, e_n\}$. When the network is initialized, all nodes broadcast their location information $e_i(x_i, y_i)$. The BS assigns a unique network ID to all nodes in the network. The set of adjacent nodes of each node in the monitoring area is expressed as the number of nodes in the circle, with the node as the central point in a certain distance. The set of adjacent nodes of node e_i in the ring CH competition radius R_c is expressed as $e_i.Vic$, as shown in Eq. (13):

$$e_i.Vic = \left\{ e_j \mid \mathrm{d}\left(e_i, e_j\right) < R_c \right\} \quad (\mathrm{i}, \mathrm{j} = 1, 2, ..., \mathrm{n} \quad \mathrm{i} \neq \mathrm{j}) \tag{11}$$

To ensure that CH selection in the network is fair and that the location of the selected CH is not concentrated in a certain range, an appropriate CH competition radius must be selected. It should not only include a certain number of nodes but also avoid other nodes being affected during neighborhood selection. If the competition radius of selected CHs is too large, then the competition range of temporary CHs will be large, and the number of generated final CHs will decrease. In contrast, if the selected CH competition radius is too small, then the final CH number in the network will increase. Next, the best RC value is selected via MATLAB simulation. Fig. 4 shows the relationship between the CH competition radius and the number of rounds corresponding to the network life cycle.

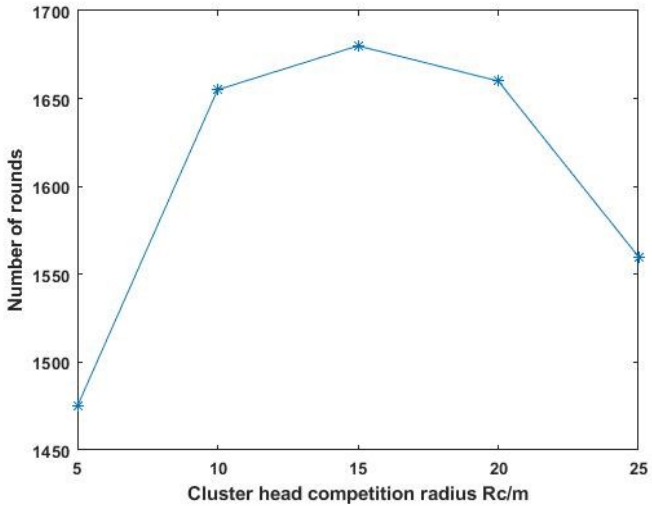

Figure 4: Relationship between the CH competition radius and the number of rounds corresponding to the network life cycle

Fig. 4 shows that R_c =15 m corresponds to the longest network life cycle. If the CH competition radius is small, then the average number of nodes in the cluster is small. If the number of network CHs is too large, then more nodes need to communicate directly with the BS, which reduces the advantage of the clustering routing algorithm in terms of network energy savings. When the competition radius is large, the number of CHs is small, and on average, each cluster contains more nodes. At this time, the CH needs to consume energy to receive and fuse the data in the cluster, which can cause the first node of the network to die prematurely.

c) Select the initial cluster center node

In Step a, the number of initial clustering center points is preliminarily obtained. In Step b, the size of the node CH competition radius is obtained. In this step, k clustering centers are selected. According to the above discussion and analysis, the traditional K-means clustering algorithm has advantages in terms of simplicity. However, when the algorithm is processing a large number of samples, a local optimal solution with unreasonable final clustering results can be easily obtained. Therefore, after improving the traditional K-means clustering algorithm, k clustering center points are selected from the node set; see Fig. 3.

In Step b, the value of CH competition radius R_c is obtained via analysis. In addition, this approach obtains the number of nodes in the radius. However, not all nodes in the step can compete for the number of cluster center points. To reduce the complexity of the algorithm and to ensure fairness, a threshold of the number of neighboring nodes should be set. It is necessary to select the node ID whose number is larger than the threshold and to participate in the competition for the cluster center point to reduce the randomness of the traditional K-means algorithm. To ensure that the appropriate cluster central node is selected, blind iteration must be avoided, and the number of iterations must be reduced. The threshold value of the number of adjacent nodes is shown in Eq. (14):

$$\eta = \alpha_2 \frac{1}{n} \sum_{i=1}^{n} e_i.Vic \tag{12}$$

In the formula, the weight coefficient is A, which equals 0.6 according to the experiment. When the number of adjacent nodes meets the following conditions.

$$e_j.Vic \geq \eta, j = 1,2,...,n \tag{13}$$

Node e_j is saved to collection C_{S_1}, and the iterative process continues. All nodes that meet the above criteria are added to collection C_{S_1}, and then a node is randomly selected from the set C_{S_1} as the first initial clustering center C_1 (as shown in the no. 1 node in Fig. 3). When the m ($2 \leq m \leq k$) initial clustering center node C_m is selected, for the remaining ordinary nodes e_j, the distance from the other m-1 clustering center nodes must be greater than R_C, and the number of neighboring nodes must be greater than η. As shown in Eq. (16), node e_j is saved into set C_{S_m}:

$$d(e_j, c_i) > R_c \quad e_j.Vic \geq \eta \quad (1 \leq i \leq m-1) \tag{14}$$

Select a node randomly from set A as the mth initial cluster center (node m in Fig. 3). Repeat this step until k initial cluster points are selected. The first initial clustering point is expressed as C_1 (l=1, 2, …, k).

d) Compute the cluster centers

After the iteration of the improved K-means clustering algorithm, k cluster center nodes are obtained. According to the above equation, the selection of clustering points is related to the density of the nodes. If the node density is high, then the probability of clustering points will be higher. However, the probabilities of clustering points in places with low densities will be much smaller, which improves the similarity of a cluster and separates the objects of different clusters. By improving the method of selecting initial cluster points, the problem of local convergence due to random selection of traditional algorithms is solved. After the clustering center node is selected, the remaining nodes are added to the cluster according to their distance to all the clustering points, and finally, k clusters are formed. Then, the center of mass of each cluster set is calculated by Eq. (17). The centroid of the first cluster set is expressed as $CE_l = \left(\overline{x_l}, \overline{y_l}\right)$:

$$\overline{x_l} = \frac{\sum_{i=1}^{N_l} x_i}{N_l} \qquad \overline{y_l} = \frac{\sum_{i=1}^{N_l} y_i}{N_l} \tag{15}$$

e) Cluster formation

After selecting the cluster central node, the Euler formula is used to assign the remaining nodes to their respective cluster central nodes. The center of each cluster is then recalculated based on the existing objects in the cluster.

The traditional K-means algorithm is a simple iterative method. The results are highly random and cannot guarantee the convergence of the final solution. The final result depends on the central location of the cluster. In addition, before using the modified algorithm, the initial value of k must be clear. If the initial k value is not appropriate, then it is difficult to obtain effective clustering results. For the K-means algorithm, the initial partition also needs to be determined according to the clustering center and then optimized. Therefore, the selection of the initial clustering center has a great impact on the cluster results. In this section, the improved K-means algorithm combines the node CH competition radius and the number of adjacent nodes for the selection of the clustering center node. This approach ensures that the initial cluster points are not clustered in a region. In this way, the distribution of center nodes is even, and outlier center nodes are avoided. This approach reduces the number of iterations and effectively improves the iteration efficiency.

3.2 CH selection stage

After the optimal cluster is formed, the BS broadcasts the cluster message to each node. The message includes the ID number of the cluster, the central location of the cluster, and the average energy of the cluster. After receiving the message from BS broadcast, all nodes begin to campaign for CH in their own cluster. After the distributed algorithm is used to form a cluster in the LEACH algorithm, the nodes make autonomous decisions without centralized control. In this algorithm, the goal is to evenly distribute the traffic load among all nodes in the network to avoid increasing the loads of the nodes and causing the nodes to run out of energy prematurely. The energy consumption of CH nodes is much larger than that of other nodes. Therefore, each node is required to compete to become a CH according to certain rules. Based on the above analysis, the voting right value of CH in clustering was set, as given by Eq. (18):

$$T_h(i) = \lambda_1 \frac{E_i}{\overline{E_h}} + \lambda_2 \frac{e_i \cdot V_{ic}}{\overline{e_h \cdot V_{ic}}} + \lambda_3 \frac{d_{avg}^h}{d_{(i,NCE_h)}} \tag{16}$$

The probability of nodes in the h-th cluster becoming CHs is expressed as $T_h(i)$. Here, $h \in [1,k]$. The remaining energy of node i in cluster h is represented by E_i. The average energy of all surviving nodes in the h-th cluster is expressed as $\overline{E_h}$. The average number of adjacent nodes of all surviving nodes in the cluster is expressed as $\overline{e_h.Vic}$. The average distance of all nodes in the h-th cluster to the center of the cluster is denoted by d_{avg}^h. The distance between node i in the cluster and the center of the cluster is denoted by $d_{(\mathrm{i,NCE_h})}$. The regulatory factors are λ_1, λ_2, and λ_3, and they are responsible for adjusting the weight of each parameter; $\lambda_1 + \lambda_2 + \lambda_3 = 1$. The CH set is represented as CH_{h}. The selection of CH in the cluster group is related to the number of neighboring nodes, the residual energy of the nodes and the centroid of the cluster. Moreover, in the cluster, the number of neighboring nodes is large, and the remaining energy is large. The closer the node is to the cluster center, the greater the probability of becoming the CH.

3.3 Data forwarding stage

After completing the CH selection in the cluster, each cluster CH node broadcasts the message that it is competing to become a cluster CH to the monitoring area. Other common member nodes in the network receive the CH broadcast information and adaptively join each CH node to complete the establishment of the final cluster according to the level of signal reception strength. In intracluster communication, cluster members need to send the monitored data and the data forwarded by the agent to the CH node. Then, the CH node receives the data of all nodes in the cluster and performs data fusion processing to remove redundant data. For the communication between the CHs, the distance between the CHs and BS is uncertain. If single-hop transmission is adopted, then the transmission power will increase, and a large amount of energy will be consumed. Therefore, multihop routing is adopted to transmit data between clusters to BSs.

4 Simulation and analysis

In a WSN, it is difficult to implement a routing algorithm in a real environment when the nodes in the monitoring region are deployed on a large scale. Therefore, this paper adopts a simulation experiment to study the protocol test of the algorithm and analyzes its performance on a computer. This paper uses MATLAB to write a simulation program to analyze and evaluate the performance of the algorithm.

Assume that 300 uniformly configured nodes are randomly distributed in a 300 m×300 m square area. The BS is located outside the square area (150, 400). The performance of the improved algorithm is analyzed in terms of clustering structure, energy consumption and number of surviving nodes. Other simulation parameters are set as shown in Tab. 1.

Table 1: Simulation parameter settings

Parameters	Value
Monitoring area $P \times P$	300 m×300 m
The BS location	(150, 400)
Number of sensor nodes	300
Initial energy of node	0.5 J
E_{elec}	50 nJ / bit
ε_{fs}	10 $pJ / \left(bit \cdot m^2\right)$
ε_{mp}	1.3×10^{-3} $pJ / \left(bit \cdot m^2\right)$
Data fusion energy consumption	5 $pJ / \left(bit \cdot m^2\right)$
$\lambda_1 \quad \lambda_2 \quad \lambda_3$	$0.7 + 0.2 + 0.1 = 1$
Packet size	4000 bit

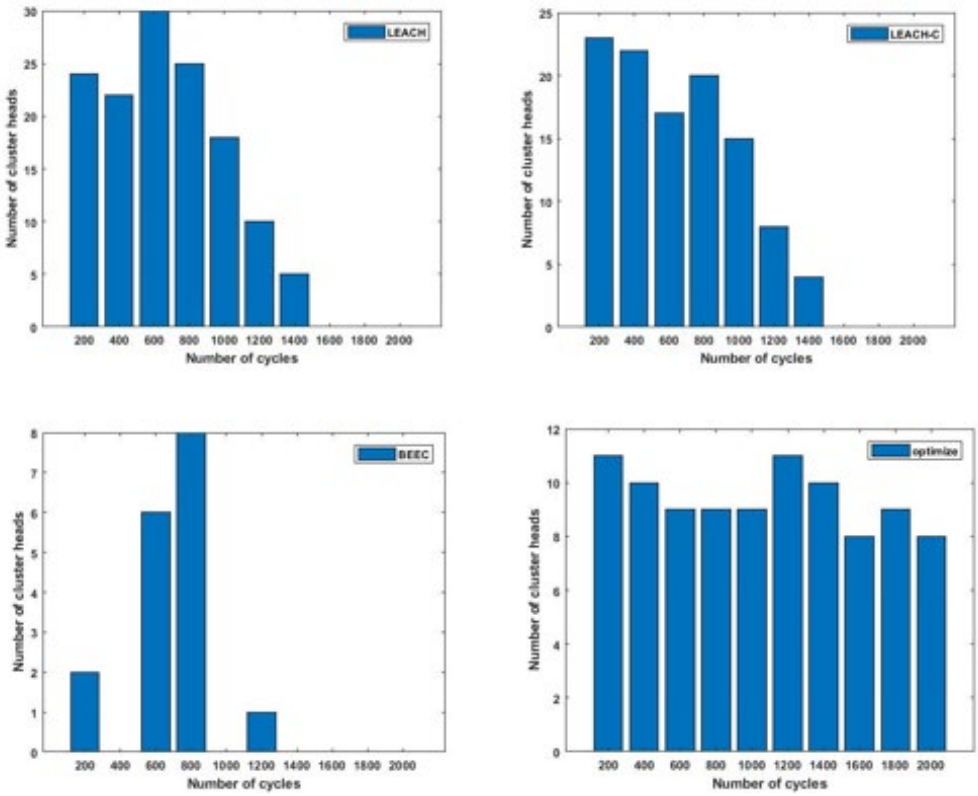

Figure 5: Statistical graph of the number of CHs generated by various algorithms

As seen from Fig. 5, the number of CHs selected by the LEACH and LEACH-C algorithms fluctuates greatly due to the CH selection mechanism of the random number and threshold. Although the BEEC algorithm adopts a local nonuniform competition mechanism, the candidate CHs are still randomly selected. As a result, the number of CHs is unstable and cannot cover the whole network. The algorithm in this paper is based on the global network and considers the distance between nodes and sink nodes. Therefore, the number of CHs is relatively stable.

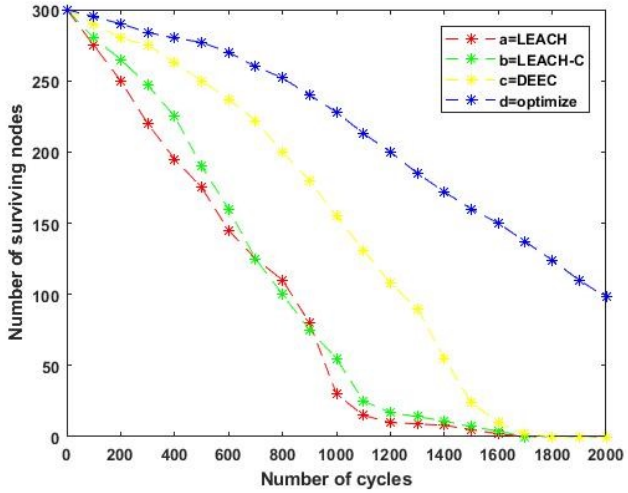

Figure 6: Relationship between the number of surviving nodes and the number of cycles

Fig. 6 shows the relationship between the number of surviving nodes and the number of cycles in the network. The slope of the number of surviving nodes versus the number of cycles during the operation of the improved algorithm is the smallest of all the slopes, and there is no obvious inflection point. The results show that, compared with LEACH, LEACH-C and BEEC, the improved algorithm can balance the energy consumption of the whole network nodes well and is very energy efficient. When the four algorithms all run approximately 1700 rounds, the nodes of LEACH, LEACH-C and BEEC have all died, while the algorithm proposed in this paper has nearly 140 surviving nodes. The improved algorithm can effectively extend the network life cycle.

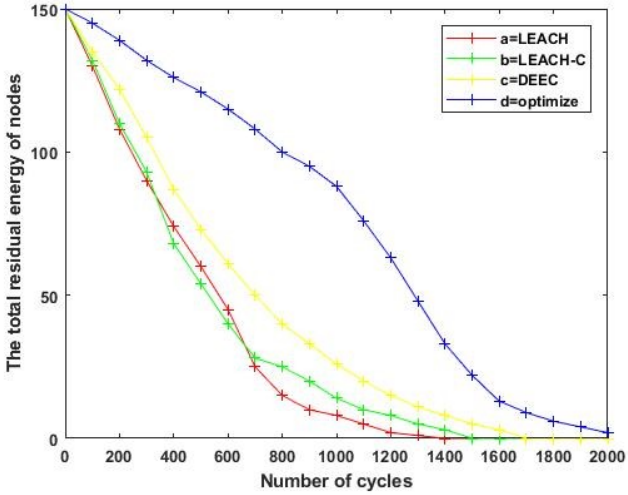

Figure 7: Relationship between the total residual energy of nodes and the number of cycles

From the diagram of the relationship between the total residual energy of network nodes and the number of cycles in Fig. 7, it can be concluded that the network energy decreases with additional cycles. In contrast, the total energy consumed by using the proposed algorithm is less than that of the other three algorithms for the same number of iterations. At the same initial energy, the energy of the other three algorithms has been exhausted, while the improved algorithm still has energy in the 2000^{th} round. The results show that the proposed algorithm can effectively save energy.

5 Conclusion

This paper introduces the application background of a routing algorithm in terms of WSN energy savings and the advantages and disadvantages of clustering routing. Then, the classical clustering algorithm and related improved clustering algorithm are analyzed. Finally, to make up for the unbalanced energy consumption in the network, a nonuniform energy-saving clustering routing algorithm based on improved K-means clustering is designed. This approach improves the clustering point selection mechanism of the traditional K-means algorithm, and the cluster structure is optimized. Finally, by comparing the method with LEACH, LEACH-C and KAEC, the improved algorithm is shown to balance the network energy well and exhibit a long network life cycle.

Funding Statement: This research was funded by the Science and Technology Support Plan Project of Hebei Province (grant numbers 17210803D and 19273703D), the Science and Technology Spark Project of the Hebei Seismological Bureau (grant number DZ20180402056), the Education Department of Hebei Province (grant number QN2018095), and the Polytechnic College of Hebei University of Science and Technology.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Akila, I. S.; Venkatesan, R. A. (2018): Fuzzy based energy-aware clustering architecture for cooperative communication in WSN. *Computer Journal*, vol. 59, no. 10, pp. 1551-1562.

Bagci, H.; Yazici, A. (2013): An energy aware fuzzy approach to unequal clustering in wireless sensor networks. *Applied Soft Computing*, vol. 13 no. 4, pp. 1741-1749.

Chang, X. Q.; Zhang, D. H. (2016): A new heterogeneous clustering algorithm for wireless sensor networks. *Journal of Jilin University (Science Edition)*, vol. 54, no. 6, pp. 1388-1394.

Emad, A.; Lon, M. (2018): New energy efficient multi-hop routing techniques for wireless sensor networks: static and dynamic techniques. *Sensors*, vol. 18 no. 6, pp. 1863-1870.

Geetha, N.; Sankar, A. (2018): A multi criterion fuzzy based energy efficient routing protocol for Ad hoc networks. *Intelligent Automation and Soft Computing*, vol. 24, no. 4, pp. 711-719.

Heinzelman, W. B.; Chandrakasan, A. P.; Balakrishnan, H. (2002): An application-specific protocol architector for wireless microsensor networks. *IEEE Transactions on Wireless Communication*, vol. 1, no. 4, pp. 660-670.

Hou, H.; Song, B.; Zhou, W. Y. (2015): Clustering routing optimization algorithm. *Microelectronics and Computer Energy Efficient*, vol. 32, no. 7, pp. 121-124.

Jiang, C. J.; Shi, W. R.; Tang, X. L. (2012): Energy balanced non-uniform cluster routing protocol for wireless sensor networks. *Journal of Software*, vol. 34, no. 5, pp. 1222-1232.

Li, S. S.; Cui, T. J.; Liu, J. (2018): Research on the clustering analysis and similarity in factor space. *Computer Systems Science and Engineering*, vol. 33, no. 5, pp. 397-404.

Liao, Y.; Qi, H.; Li, W. Q. (2013): Load-Balanced clustering algorithm with distributed self-organization for wireless sensor networks. *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1498-1506.

Liu, J. J.; Hu, Y. J. (2014): A balanced and energy-efficient algorithm for heterogeneous wireless sensor networks. *IEEE Wireless Communications and Signal Processing*, pp. 1-6.

Liu, X. X. (2015): A typical hierarchical routing protocols for wireless sensor network: a review. *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5372-5383.

Meng, Q.; Zhang, J. J. (2019): Optimization and application of artificial intelligence routing algorithm. *Software Tools and Applications*, vol. 22, no. 4.

Qing, L.; Zhu, X. Q.; Wang, M. W. (2006): Distributed energy efficient clustering algorithm for heterogeneous sensor networks. *Journal of Software*, vol. 17, no. 3, pp. 481-489.

Sirsikar, S.; Wankhede, K. (2015): Comparison of clustering algorithms to design new clustering approach, *Communication and Control*, vol. 15, no. 3, pp. 147-154.

Sohrabi, K.; Gao, J.; Ailawadhi, V. (2000): Protocols for self-organization of a wireless sensor network. *IEEE Personal Communications*, vol. 7, no. 5, pp. 16-27.

Sundararajan, R. K.; Arumugam, U. (2018): Event detection and information passing using LEACH lrotocol in wireless sensor networks. *Wireless Personal Communications*, vol. 101, no. 3, pp. 1703-1714.

Wang, L.; Qi, L.; Qi, J. Y.; Liu, Z. Z. (2018): An improved energy optimization algorithm of hierarchical sensor network for cluster head election. *Measurement and Control Technology*, vol. 37, no. 1, pp. 92-115.

Yu, L.; Zhong, Y.; Xiao, Y. Y.; Liu, G. C.; Hu, B. (2019): A novel multi-hop algorithm for wireless network with unevenly distributed nodes. *Computers, Materials & Continua*, vol. 58, no. 1, pp. 79-100.

Zhang, R.; Cao, J. (2010): Uneven clustering routing algorithm for wireless sensor networks based on ant colony optimization. *Journal of Xian Jiaotong University*, vol. 3, no. 6, pp. 67-71.