

Received January 7, 2020, accepted January 21, 2020, date of publication January 29, 2020, date of current version March 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970208

Analysis of Network Coverage Optimization Based on Feedback K-Means Clustering and Artificial Fish Swarm Algorithm

YINGYING FENG^{ID}, SHASHA ZHAO^{ID}, AND HUI LIU^{ID}

College of Information Engineering, Fuyang Normal University, Fuyang 236041, China

Corresponding author: Shasha Zhao (zhaoshasha_0701@163.com)

This work was supported in part by the Anhui provincial science and technology department fund project “Research on fuzzy multi-threshold tumor image segmentation and recognition based on swarm intelligence optimization” under Grant 1908085MF207, and in part by the Anhui natural science research project “Channel allocation optimization in wireless Mesh networks” under Grant KJ2019A0950.

ABSTRACT There is a certain energy loss in the process of wireless sensor network information collection. Moreover, the current network protocols and network coverage methods are not sufficient to effectively reduce system energy consumption. In order to improve the operating efficiency and service life of wireless sensor networks, this study analyzes the classic LEACH protocol, summarizes the advantages and disadvantages, and proposes a targeted clustering method based on the K-means algorithm. At the same time, in order to maximize the network coverage and minimize the energy consumption on the basis of ensuring the quality of service, a wireless sensor network coverage optimization method based on an improved artificial fish swarm algorithm was proposed. In addition, a controlled experiment is designed to analyze the effectiveness and practical effects of the proposed algorithm. The experimental results show that the method proposed in this paper has certain advantages over traditional methods and can provide theoretical references for subsequent related research.

INDEX TERMS Feedback, K-means clustering, artificial fish swarm algorithm, network coverage, algorithm optimization.

I. INTRODUCTION

With the continuous development of computer technology and communication technology, the information carrying status of wireless sensor networks is increasing. From the current research status of wireless sensor networks, the topology control algorithms can be divided into three types: planar, hierarchical, and heuristic. All nodes in a flat network have the same status and functions. This kind of algorithm is a bit simpler in the network structure, and the network itself has good robustness, but when the transmission data shows a big data burst mode, the network structure cannot adapt to the demand, which directly causes the network to be paralyzed. However, many current studies have not analyzed the energy consumption of wireless sensor networks. Therefore, as the demand for information transmission continues to increase,

wireless sensor networks may not be able to cope with a large number of data transmission needs in the future [1].

The information-carrying capacity of the information society is constantly increasing, and the development of communication technology has also promoted the development of wireless sensor network technology. At present, wireless sensor networks have formed a complete system. From the process of information acquisition, information processing to information transmission, wireless sensor networks play an important role. Its networking structure is simple, and its application flexibility is strong, so it can play an important role in future communications. From the perspective of practical applications, wireless sensor networks can not only realize the real-time transmission of information, but also can be applied to the monitoring of multiple parameters such as noise, natural environment status, and real-time status of buildings. Sensors inevitably have energy consumption in the process of information transmission, and the energy carried by the sensor has priority, so how to reduce the energy

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei^{ID}.

consumption in the information transmission of wireless sensor networks is the fundamental to improve the operating efficiency of wireless sensors [2].

Based on the above analysis, this study analyzes the existing wireless sensor network routing protocols. After analyzing the cut points, we focus on the research of clustered routing, and based on this, we propose a K-means clustering algorithm based on feedback. This algorithm mainly adopts the multi-hop method of cluster head and main cluster head to collect, fuse and transfer information. At the same time, the algorithm combines the cluster head and main cluster head information for each round based on node energy and location concentration weighting to avoid unnecessary energy consumption. For the coverage study of wireless sensor networks, this paper applies the fish school algorithm to reduce the active state of a large number of nodes in the network, thereby effectively reducing energy consumption.

In summary, this paper mainly optimizes the distribution of network nodes through algorithm improvement and uses the least active nodes to improve network energy efficiency, thereby improving network operation efficiency and actual life.

II. RELATED WORK

The WSN routing protocol is mainly to ensure that the sink node and the nodes at other times perform effective data transmission. With the needs of social development, traditional simple protocols have been unable to meet the needs of information dissemination. Moreover, with the advent of the era of big data, a variety of occasions need to be applied to routing protocols, and many scholars have also developed algorithm protocols that can meet multiple needs according to actual needs [3].

In detail, the analysis of wireless sensor network routing protocols can be divided into three types: routing protocols based on quality of service, routing protocols based on geographic location, and cluster routing protocols [4]. At present, plane routing protocols and cluster routing protocols are widely used [5]. The initial energy setting of each node in the flat routing protocol is the same. The difference is that the nodes are distributed in different locations. Representative networks mainly include flood routing [6], chat routing [7], SPIN routing protocol [8], DD protocol [9], Rumor protocol [10], and so on. However, clustered routing protocols are more widely used. Typical clustering routing protocols are: low-energy adaptive clustering protocol [11] (LEACH), PEGASIS protocol [12], TEEN protocol [13], and HEED protocol [14], etc.

Two parameters are mainly set in the TEEN protocol, named soft threshold parameters and hard threshold parameters, respectively. The distinction between the two is used to optimize the network survival time point and node energy consumption [15]. In comparison, different routing protocols have their own advantages and disadvantages. However, cluster routing can effectively achieve a variety of data fusion, and the network nodes in the protocol can be divided into

cluster head nodes and general cluster nodes to achieve the division of labor and cooperation between different nodes. Therefore, the cluster routing protocol is more suitable for the development needs of the big data era.

By analyzing the current status of wireless sensors, many scholars have researched the optimization of cluster routing. Some scholars have proposed a clustering algorithm that applies fuzzy thinking to wireless sensor networks based on the problems of the LEACH protocol algorithm [16]. The algorithm flow is similar to the LEACH protocol, that is, the cluster head is first selected, and the corresponding logical structure is added to make it a complete structure.

In recent years, the further development of wireless sensor networks has prompted many experts and scholars to introduce machine learning to wireless sensor networks. For example, some common computer algorithms have been applied to wireless sensor networks. The most common one is K-means algorithm [17]. The K-means algorithm is mainly to introduce the K-means algorithm into the wireless sensor network and improve and optimize it, thereby further improving the operating performance of the wireless sensor network [18]. In literature, a fuzzy clustering algorithm was proposed to be used in the LEACH protocol [19], and it was proved that the improved algorithm is superior to the LEACH protocol in the first node death time and half the node death time. In literature, the LEACH-K protocol was proposed based on the K-means algorithm. This protocol improves some defects in the LEACH protocol, which makes the improved protocol greatly improve the remaining energy of nodes and the stability of nodes in the cluster. In literature [20], the aggregation degree of nodes was adopted as the main basis for clustering and the principle of dynamic clustering is used to make the network energy consumption more balanced.

Secondly, the research on cluster routing in wireless sensor networks also has some new directions. With the promotion of 5G network technology in the past two years, network coverage has set off a wave of research. For example, the literature [21] analyzed the problems such as possible redundancy in network coverage, and proposed an energy-saving model, which combines node energy consumption and fish swarm algorithm to optimize network coverage. Literature [22] is an improvement on traditional routing. Based on traditional routing analysis, fish school algorithm was introduced, and fish school algorithm was also introduced in network coverage. Literature [23] combined with some shortcomings of artificial fish swarm algorithm to introduce particle swarm algorithm to improve it, and then studied the network coverage of the improved algorithm. Experiments show that the improved algorithm has a qualitative improvement in related performance.

In summary, scholars at home and abroad are currently conducting research on cluster routing, etc., and analyze it from different angles in terms of energy consumption. However, from the actual situation, the research of these experts and scholars has some shortcomings. Most of the researches

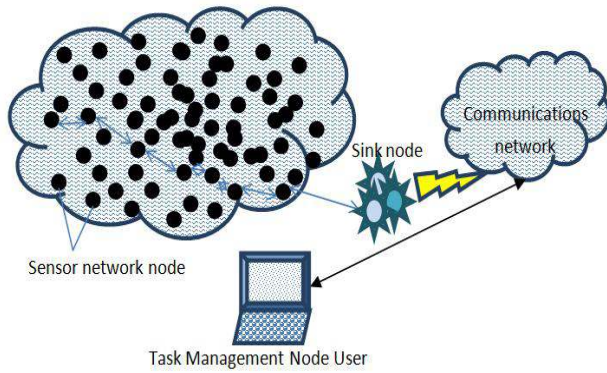


FIGURE 1. The wireless sensor network node model.

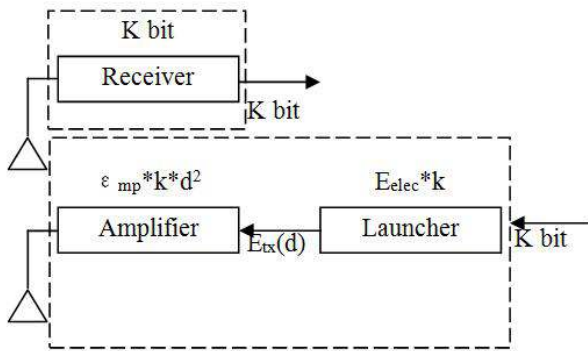


FIGURE 2. Radio energy model.

of experts and scholars are one-sided and cannot adapt to the performance requirements of wireless sensor networks in the era of big data. Especially in the context of centralized information transmission, the routing and network coverage of wireless sensor networks must be optimized. Based on this, this study analyzes the load balancing of network nodes in network coverage planning, etc., and combines the improved algorithm for performance verification.

III. MODEL BUILDING

A. ENERGY CONSUMPTION MODEL

The sensor network is mainly based on the distance when the information is transmitted, and the distance does not have a linear relationship. The functional relationship is generally an exponential function relationship and can be divided into a free space model and a multi-channel transmission model according to the distance. Figure 1 shows the wireless sensor network node model, and Figure 2 shows the current common radio energy model. The model includes a dense network base station, which transmits signals directly to the user side, and also includes the user side and the data processing center.

Assuming the information transmission distance is d , the information transmission process can be expressed as:

$$E_{T_x}(C, d) = E_{T_{x-elec}}(C) + E_{T_{x-amp}}(C, d) = \begin{cases} C * E_{T_{elec}} + C * \epsilon_{fs} * d^2, & d < d_0 \\ C * E_{T_{elec}} + C * \epsilon_{mp} * d^4, & d \geq d_0 \end{cases} \quad (1)$$

In actual operation, if the transmission model is a self-space, the information transmission distance is less than d_0 . If the transmission model is adopted for multi-channel transmission, the information transmission distance is greater than d_0 . Then, the energy consumed by the receiver to receive C -bit information here is:

$$E_{R_x}(C) = E_{R_{x-elec}}(C) = E_{elec} * C \quad (2)$$

d_0 is the threshold, which is the ratio of ϵ_{fs} and ϵ_{mp} .

$$d_0 = \sqrt{\epsilon_{fs} / \epsilon_{mp}} \quad (3)$$

In the above formula, E_{elec} represents the energy consumed for transmitting 1-bit information, ϵ_{fs} represents the energy loss parameter of the power expansion process under the free space model, and ϵ_{mp} represents the energy loss process of the power expansion process under the multipath model.

Assuming that C represents a certain message length, the energy consumed after receiving the desired message can be expressed as [24]:

$$E_{EX} = E_{elec} * C \quad (4)$$

If it is assumed that there is a square region with a side length of m , and there are n sensor nodes in the region, and it is divided into k clusters, the number of sensor nodes contained in each cluster can be expressed as n/k . In each cluster, the energy consumption of the cluster head is the largest. The energy consumption of the cluster head can be expressed as the following formula:

$$E_{CH} = CE_{elec} \left(\frac{n}{k} - 1 \right) + CE_{DA} \frac{n}{k} + CE_{elec} \quad (5)$$

Because the transmission range of nodes in the same cluster is relatively small, free consumption space is often used for transmission in the node consumption model in the cluster. The energy consumption of the nodes in the cluster can be expressed as:

$$E_{in} = CE_{elec} + C\epsilon_{fs}d^2 \quad (6)$$

The average distance from the nodes in the cluster to the head of the cluster is:

$$d = \sqrt{1/2 * \frac{m^2}{k}} \quad (7)$$

The total energy consumed by each cluster is:

$$E_{clu} = E_{CH} + \left(\frac{n}{k} - 1 \right) E_{in} \quad (8)$$

B. CLUSTER GENERATION STAGE

The cluster generation phase of this research is mainly carried out by the K-means algorithm, and the primary cluster candidate area is selected according to the principle of the first base station of the primary cluster and the principle of closeness of the cluster center. The establishment of the primary cluster group can reduce the burden of data transmission by a single

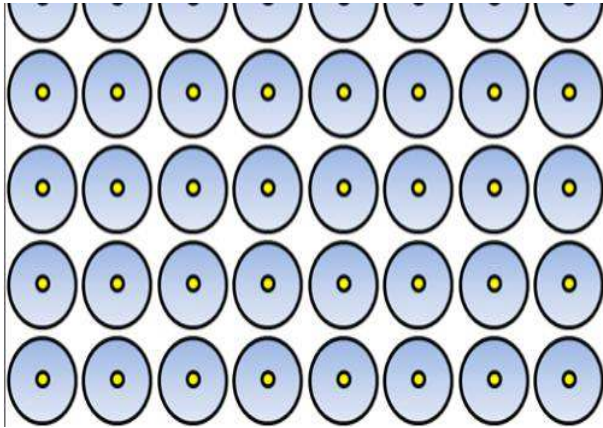


FIGURE 3. Schematic diagram of area coverage.

primary cluster head. From the actual situation, the number of main cluster heads and the number of nodes in the cluster is small. The main reason for this result is that the main cluster is generally only responsible for data fusion and transmission, so it consumes less energy and has fewer alternatives. When the node closest to the center is used as the initial cluster node, the cluster center formula can be expressed as [25]:

$$(Center_j.x, Center_j.y) = v_j = \{x_{vj}, y_{vj}\} \quad (9)$$

In the above formula, $Center_j.x$ is the abscissa of the cluster center, and $Center_j.y$ is the ordinate of the cluster center. In this study, the K-means algorithm is mainly applied to the cluster generation stage. After the initial cluster head is selected, the nodes that are close to the center and the base station are selected as alternatives. When using the K-means algorithm, the following criteria are mainly used to measure whether the clustering is completed:

$$\min \sum_{i=1}^k \sum_{x_j} (x_j - u_i) \quad (10)$$

In the formula, g represents the arithmetic mean of $Z_i = (Z_1, \dots, Z_k)$ in all k clusters, $X = (x_1, \dots, x_c)$ represents the set of real vectors of c dimensions, and u_i is the average of all points $Z_i = (Z_1, \dots, Z_k)$ in k clusters. Figure 3 shows the schematic diagram of area coverage. The yellow part in Figure 3 represents the base station, and the area inside the blue circle is the area radiated by the base station.

The K-means algorithm used in this study has many advantages and the algorithm runs fast. The complexity of the algorithm is $O(n \cdot k \cdot t)$, n is the number of samples, k is the number of clusters, t is the number of iterations, and $k \ll n$. Based on the above analysis, in view of the typical LEACH clustering routing protocol with uneven clustering and random distribution of cluster head nodes leading to high network energy consumption, this paper combines the advantages of the K-means algorithm to propose a K-means clustering protocol based on feedback.

C. THE ESTABLISHMENT OF CLUSTERS IN WIRELESS SENSOR NETWORKS

In this paper, K-means clustering is used for clustering in wireless sensor network node processing. Moreover, this paper chooses a linear weighted distance from the base station and the weight of the center distance of each cluster from the overall network and draws several minimum nodes as the main cluster head candidate. Through this centralized establishment of the main cluster head group to reduce the energy consumption of a single main cluster head for data transmission, the following definitions can be made:

$$G_{SCH} = 0.1L_b + 0.9L_c \quad (11)$$

In the formula, L_c represents the distance between the node and the center of each cluster, and L_b represents the distance between the node and the base station. G_{SCH} is calculated and the smallest node is selected as the main cluster head. The above formula mainly considers the following points:

A large distance between the cluster head node and the base station will result in higher energy consumption. However, if the distance between them is close, choosing the main cluster head will seem meaningless. The node corresponding to the minimum value of G_{SCH} calculated at the beginning of the study is the main cluster head at the beginning. After that, each round is similar to the process of selecting the cluster head. In the calculation, a dynamic feedback function is required to carry out the calculation.

In actual calculation, the iteration can be automatically stopped as long as the objective function is optimal. There are certain differences in the objective function for different distance measures. If the Euclidean distance is used for optimization, the goal is generally to minimize the sum of squares of the object to its cluster center. When solving the optimal solution of a variable, it is mainly carried out by solving its partial derivative. After each iteration, the SSE decreases and eventually converges.

In the cluster head election phase, the cluster generation process is generally performed only once, and the cluster head selection is performed once each time a cluster is generated. Based on this, this research assumes that the initial energy values of the nodes in the sensor system network are the same, and after the nodes are arranged in the area, the central node is determined by combining the cluster head and main cluster head selection methods described above. In the energy consumption calculation, the cluster head and the main cluster head are selected by weighting the position distance and the feedback function of the remaining energy of each node. Through research, it is found that with the demise of network nodes, gradually increasing the distance weighting coefficient and decreasing the energy weighting coefficient can effectively improve the network operation efficiency and further enhance the life of the wireless sensor network. The feedback scoring function can be

TABLE 1. Formula parameters.

parameter	Parameter meaning
E0	Node's initial energy
Eres(node _i)	Residual energy of node i
D(node _i ,Centriod-C _j)	Distance from node to cluster center
dmax	The total distance from all nodes in the cluster to the cluster center
nodesel	Selected cluster head node

expressed as:

$$w_{node_1} = \alpha(t) \frac{E_{res}(node_1)}{E_0} + [1 - \alpha(t)] \times \frac{d_{max} - d(node_i, Centriod_C_j)}{d_{max}} \quad (12)$$

The cluster head selection formula is:

$$node_{sel} = \arg \max (w_{node_i}) \quad (13)$$

Through research, it is found that when the network operation discourse is increasing, the use of residual energy weighted reduction and location concentration weighted increase can effectively improve the system operation life. Therefore, $\alpha(t)$ should be a subtraction function. Meanwhile, $\alpha(t)$ is a number in the range between [0, 1], so the formula used in this article is:

$$\alpha(t) = \frac{1}{1 + e^{-kt}} \quad (14)$$

In the above formula, k represents a control variable, and t represents the total number of nodes currently contained in the network during system operation. From the analysis of the above formula, it is not difficult to see that at a certain moment, the smaller the cluster center is from the cluster center and the more the remaining energy of the node, the easier it is to be elected as the cluster head at that moment. The cluster head selected by this method satisfies the current optimal conditions of the cluster, but it does not consider the global situation. After each selection of the cluster head, it will send out the relevant information as the cluster head. However, other unselected nodes are always working, and these nodes collect and pass information to the cluster head. The system gradually enters a steady state. After that, the cluster heads are re-selected in each round of the network node and related operations are performed until the next round. The meaning of each parameter is shown in Table 1, and the clustering principle is shown in Figure 4.

IV. ARTIFICIAL FISH SWARM ALGORITHM

A. BASIC MODEL OF ARTIFICIAL FISH SWARM ALGORITHM

Artificial fish swarm algorithm is a bionic algorithm that imitates fish foraging in the ecological environment. The algorithm achieves optimal target search by simulating fish swarm foraging, rear-end collision, clustering and other methods, and achieves global optimization through local optimization. In the simulation, the competition, survival and coordination

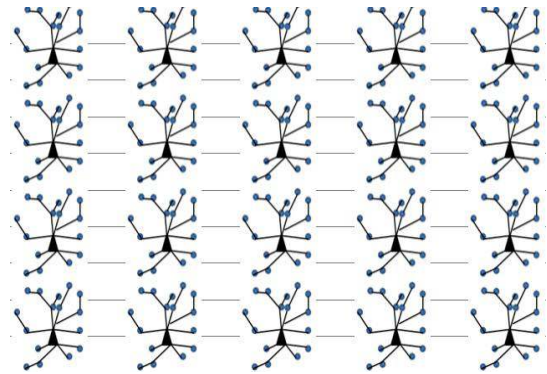


FIGURE 4. Schematic diagram of clustering.

mechanism of the school of fish is simulated during the process of foraging and mutual cooperation of the school of fish to achieve the purpose of maximizing efficiency. Judging from the actual situation, the school of fish has excellent evasive ability and searching ability when foraging. Applying the fish swarm algorithm to wireless sensor network coverage optimization does not need to know the value of the objective function and the gradient and has a certain adaptive type in the search space. Moreover, it does not require high initial value settings and is not sensitive to parameter values. Therefore, it has certain practical effects.

The artificial fish swarm algorithm in the optimization mainly goes through the iterative layer to iteratively until it gets the optimal solution. The fish school algorithm is applied to the optimization of wireless sensor networks, and an object-oriented processing method is adopted, and a large class containing member functions and variables is set as an artificial fish. The parameters in this model mainly include the representation of the individual state of the artificial fish X, the artificial fish step size, the crowding factor, the perceived distance, and the number of retries. Moreover, the system member functions mainly include the artificial fish food concentration at the current position, and the behavior of the artificial fish. After the system parameter setting is completed, the information of the artificial fish is encapsulated similar to the label form, and the entire system includes a school of fish composed of multiple fish. These individual states can be perceived by other fish. Therefore, in a certain n-dimensional search space, there are N artificial fish, and the state of the artificial fish is expressed as:

$$X = \{X_1, X_2, \dots, X_n\} \quad (15)$$

Among them, $X_i = (1, 2, \dots, n)$ is the variable to be optimized. The current food concentration of artificial fish is:

$$Y = f(x) \quad (16)$$

In the formula, Y is the objective function, and the distance between individual artificial fish is

$$d_{i,j} = \|x_i - x_j\| \quad (17)$$

The state of each fish is a solution, which is not necessarily the optimal solution. The state is substituted into formula (16), and the result can be used to compare the status of each artificial fish.

B. BEHAVIOR DESCRIPTION OF ARTIFICIAL FISH IN WIRELESS SENSOR NETWORK

1) FORAGING BEHAVIOR

In wireless sensor networks, simulation is mainly performed by simulating the foraging behavior of fish swarms. During the foraging process, fish swarms mainly observe the concentration of things to determine the subsequent behavior. Specifically, if there are many things found during the school of fish movement, it will preferentially swim in that direction, and it can simulate the process and set related parameters. If the current state of the school of fish can be expressed as X_i , the food concentration can be used as the main factor to guide the school of fish to move forward. If the food concentration is higher than the current one, the school of fish swims to the place with high food concentration. In contrast, the school of fish re-selects a new random state X_j . If the forward condition is less than the try-number, the search is continued until the condition is satisfied. After that, some parameter formulas are given:

$$X_j = X_i + Visual * Rand () \quad (18)$$

In the formula, $Rand ()$ is a random number between 0 and 1. During the solution of the maximum value, if $Y_j > Y_i$, the artificial fish school moves forward in that direction (in the solution of the minimum value, if $Y_j < Y_i$, the artificial fish school moves forward in that direction).

$$X_i^{t+1} = X_i^t + \frac{X_j - X_i^t}{\|X_j - X_i^t\|} * Step * Rand () \quad (19)$$

If the next state that satisfies the conditions has not been found, the artificial fish moves one step randomly.

$$X_i^{t+1} = X_i^t + Visual * Rand () \quad (20)$$

2) SOCIAL BEHAVIOR

Fish are usually gathered together in the water. This method not only helps to ensure the safety of one's own life, but also moves to the center as much as possible when gathering, and all individuals in the school of fish can interact with surrounding individuals. Therefore, the aggregation mode can be represented by a single individual state. The artificial fish swarm algorithm can also be expressed in the form of interaction during the execution process to avoid fish being overcrowded during the process of progress.

Assuming the state of the artificial fish set by the current system can be expressed as X_i , other individuals within its field of view can be expressed as n_f . Can be expressed as $d_{i,j} < Visual$, The status of the center position is X_c . If $Y_c/n_f > \delta Y_i$ is satisfied, it means that the food concentration around the artificial fish is high and the location is

not crowded. After that, the individual can move forward to the surrounding artificial fish center. If this requirement is not met, then foraging behavior continues on this basis. The process is expressed as:

$$X_i^{t+1} = X_i^t + \frac{X_c - X_i^t}{\|X_c - X_i^t\|} * Step * Rand () \quad (21)$$

3) REAR-END BEHAVIOR

In a school of fish, if n fish find food at the same time, and the surrounding artificial fish will follow the food location, this situation is described as rear-end behavior. The rear-end Tail behavior is a more common tailing phenomenon. In the system, this situation indicates that all nodes are looking for the optimal direction, that is, all artificial fish want to move in the optimal direction. We assume that the current state of artificial fish is X_i , and the maximum food concentration of other artificial fish companions in its field of view is X_{max} . If $X_{max} > Y_i$ and $Y_{max}/n_f > \delta Y_i$ are satisfied, the surrounding companion has the highest food concentration X_{max} , and the artificial fish moves forward in that direction, otherwise it will perform foraging behavior. The mathematical expression is:

$$X_i^{t+1} = X_i^t + Step * Rand () * \frac{X_{max} - X_i}{\|X_{max} - X_i\|} \quad (22)$$

4) RANDOM BEHAVIOR

There is some randomness in the state of fish when swimming in the water, and this behavior is conducive to the whole fish to search for food. Generally speaking, artificial fish will randomly select a state within its recognition range, and then move one step in that direction. Its movement process consists of multiple small processes, which is a common foraging behavior.

5) BULLETIN BOARD

The function of the bulletin board is to record the optimal historical state of artificial fish, that is, the highest concentration of things in the group during the movement of the school of fish. During the operation of artificial fish in the system, the operating information of all artificial fish will be recorded, and all artificial fish information will be compared and analyzed. If the new value obtained is higher than the original value, its status is replaced with the current value, otherwise the original value is kept unchanged. After the fish school algorithm is executed, the bulletin board will also record the optimal value obtained by it.

C. PROCESS OF ARTIFICIAL FISH SWARM ALGORITHM

The following figure shows the artificial fish swarm algorithm. The operation steps of the algorithm are: (1) First, algorithm parameter settings are performed, mainly including parameter settings such as fish school scale and step size, and the maximum visual field range, number of iterations, and congestion factor are specified. (2) Generate school

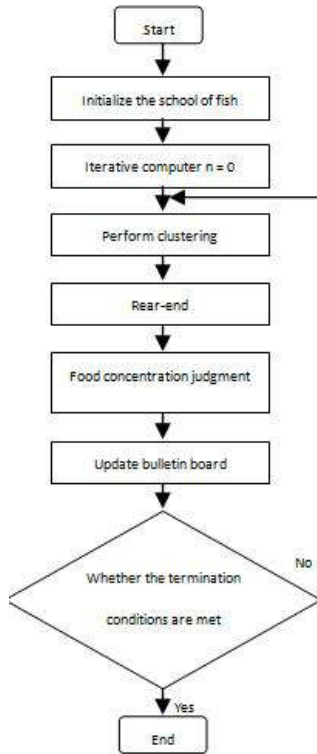


FIGURE 5. Artificial fish swarm algorithm flow.

of fish. By setting parameters, several artificial fish are generated within a certain range. After that, the school of fish is initialized, and the number of iterations is set. (3) Food concentration of artificial fish after initial purification is calculated, and related parameters are recorded on the bulletin board. (4) Foraging behavior of fish and food concentration distribution were simulated. All fish individuals need to perform this process simulation, and we need to choose whether to continue to proceed based on the nearby food concentration. (5) After the artificial fish school has performed its corresponding behavior, the bulletin board records its optimal record. After that, the food concentration is compared, and if the food concentration is higher than the original value, the current value is replaced on the bulletin board. (6) If the maximum number of iterations is reached, it will automatically stop, and the bulletin board value cannot be output. If the maximum number of iterations is not reached, iteration continues.

If N nodes are randomly arranged in a monitoring area, and the initial basic information of each node is the same, the geographical location (i.e., coordinates) of each sensor node is known, and the monitoring radius of the node is r , the sensor node in the area can be expressed as $C = \{c_1, c_2, \dots, c_N\}$. In the formula, $c_i = (x, y, r)$ indicates that this is a circle with a radius of r . Because we have no way to calculate how much to monitor an area, we discretize the area into $m \times n$ pixels. How much pixels are covered can be calculated. If a pixel (x, y) is covered by the sensing area of the i -th sensor node, the event is called r_i , and the probability of this event

is $P\{r_i\}$. The following is its mathematical expression.

$$P\{r_i\} = P_{\text{cov}}(x, y, c_i) = \begin{cases} 1, & \text{if } (x - x_i)^2 + (y - y_i)^2 \leq r^2 \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

If the distance from a pixel point (x, y) to a certain node i is less than a certain transmission range r , we consider that the pixel point (x, y) is covered by the i sensor node.

$$P\{\bar{r}_i\} = 1 - P\{r_i\} = 1 - P_{\text{cov}}(x, y, c_i) \quad (24)$$

If r_i and r_j are irrelevant, there is the relationship shown in formula (25):

$$P\{r_i \cup r_j\} = 1 - P\{\bar{r}_i \cap \bar{r}_j\} = 1 - P\{\bar{r}_i\} \cdot P\{\bar{r}_j\} \quad (25)$$

Similarly, in a node set, as long as the sensing range of one of the sensor nodes covers the pixel point (x, y) , we can assume that the pixel point (x, y) is covered by this node set. Therefore, the probability that the pixel (x, y) is covered by the node set is the union of r_i . If it is assumed that each random event r_i is independent, the coverage of a node set C is:

$$P_{\text{cov}}(x, y, c_i) = P\left\{\bigcup_{i=1}^N r_i\right\} = 1 - P\left\{\bigcap_{i=1}^N \bar{r}_i\right\} = 1 - \prod_{i=1}^N (1 - P_{\text{cov}}(x, y, c_i)) \quad (26)$$

If no node in the node set covers the pixel (x, y) , the pixel (x, y) is not covered. Otherwise, it is covered.

We assume that there are $m \times n$ pixels in the monitored area, and assume that each pixel area is 1, and its area is represented by $\Delta x \times \Delta y$. The probability $P_{\text{cov}}(x, y, c_i)$ of the joint measurement of each node set indicates whether the pixel is covered by the node set. At the same time, we refer to the ratio of the area $A_{\text{area}}(C)$ covered by the node set C to the area of the monitored area A as the node set area coverage, expressed by $R_{\text{area}}(C)$. The specific mathematical formula is as follows:

$$R_{\text{area}}(C) = \frac{A_{\text{area}}(C)}{A_s} = \frac{\sum_x \sum_y P_{\text{cov}}(x, y, C)}{m \times n} \quad (27)$$

According to formula (23), the coverage of a pixel to the sensor nodes in the network can be calculated. According to the formula ((26), the joint coverage of a pixel to the sensor node set in the network can be calculated. According to formula (27), the area coverage of the sensor node set in the network can be calculated.

A control vector $X = \{a_1, a_2, \dots, a_N\}$ is defined to describe the working state of sensor nodes in the network. If $a_i = 1$, it means that the i -th node is working. If $a_i = 0$, it indicates that the i -th node is sleeping. The coverage of sensors in the network is expressed as follows:

$$f_1(X) = A_{\text{area}}(C) / (m \times n) \quad (28)$$

The node utilization is expressed as follows:

$$f_2(X) = \sum_{i=1}^N a_i / N \quad (29)$$

For the optimization of the network coverage of the sensor network, two aspects are considered: one is to maximize the network coverage, and the other is to minimize the node utilization. Therefore, coverage optimization is a multi-objective optimization combination problem, and a linear objective function is formed by weighting. Therefore, the overall objective optimization function is:

$$F(X) = w_1 f_1(X) + w_2 (1 - f_2(X)) \quad (30)$$

In formula, $0 < w_1, w_1 + w_2 = 1$, and $w_1 = 0.6, w_2 = 0.4$.

V. IMPROVED ARTIFICIAL FISH SWARM ALGORITHM

If it is assumed that there are n nodes in the wireless sensor network in this study, we can know that there is a total of 2^n selection methods through the probability formula. When n is large, it cannot be solved by the ergodic method. Therefore, it is necessary to construct a corresponding model to solve the problem by combining the artificial fish school algorithm in this study. Through the artificial fish school algorithm, the optimization problem of the wireless sensor network can be transformed into a good target search problem, which can get rid of the problem of being limited to extreme values. For wireless sensor networks, when the space is relatively flat, the artificial fish school algorithm is in a natural swimming state in this state, so the process of finding the optimal solution may be prolonged. Therefore, the algorithm needs to be optimized. The optimization process is as follows

In fish school algorithm, artificial fish vision and swimming step are two important parameters, which have a great impact on the algorithm results. When the field of view is large, the convergence speed is faster, but the algorithm accuracy is low, and at this time, the number of fish to be searched is large. However, the algorithm converges slowly in a small field of view, but its accuracy is higher. Therefore, there are certain advantages and disadvantages in both the large field of view and the small field of view. Therefore, it is necessary to combine the two to improve, to achieve the characteristics of high accuracy and fast convergence.

The improvement ideas of this research are: In the initial stage of the algorithm, a large field of view is used to quickly find a rough solution, which effectively reduces the convergence time and determines the target range. As the iterative steps continue to increase, the field of view is gradually reduced, and a fine target search is performed until the optimal solution is found.

Model extraction: We set the wireless sensor network node coverage to C , the node's dormancy rate to S , and K to a real number greater than 0 and less than 1. Then, the function to be optimized is:

$$F = KC + (1 - K) \times S \quad (31)$$

0	1	1	...	0	0
1	2	3	...	n-1	n
L					

FIGURE 6. Example of a node selection scheme (an artificial fish).

In the above formula, K represents the degree of preference for coverage. Assuming that the number of nodes is n , 1 indicates that the node is selected, and 0 indicates that the node is dormant (not selected), then the independent variable can be represented by a binary sequence of length n , as shown in Figure 6:

The following is the representation of the artificial fish swimming field of view and step size, and formula (32) is the expression of the artificial fish field of vision:

$$Visual = Visual_o \times a + Visual_{min} \quad (32)$$

In the formula, $Visual$ is the current visual field representation of artificial fish, $Visual_o$ is the initial value of artificial fish visual field, $Visual_{min}$ is generally determined according to the accuracy of different problems, and a is the change factor of network operation. The following formula is the expression of the artificial fish step size:

$$Step = Step_o \times a + Step_{min} \quad (33)$$

In the formula, $Step$ is the current step size of the artificial fish, $Step_o$ is the initial value of the artificial fish step size, $Step_{min}$ is the value determined according to the accuracy of different problems, and a is the change factor of the network operation.

$$a = \exp(-10((gen - 1) / Max_Gen) \wedge 3) \quad (34)$$

In the formula, gen represents the number of rounds performed by the artificial fish school, and Max_Gen represents the maximum number of rounds of fish school behavior. At the beginning of the operation, the field of view (step size) is given a maximum value. With the operation of the network, the field of view (step size) gradually decreases as the parameter a change.

The algorithm pseudo code is as follows:

The population size of the initialized school of fish and the value of the initialization bulletin board F (0 is fine)

For $gen = 1 : Max_Gen \% Max_Gen$ represents the maximum number of rounds of fish school behavior.

For $j = 1 : Fish_Num \% Fish_Num$ represents the size of the school of fish/

For $(j) \%$ represents the value of F corresponding to the state of the artificial fish

END

END

Through the artificial fish school algorithm, the optimization problem of the wireless sensor network can be transformed into a good target finding problem, which can get rid of the problem of being limited to extreme values. After the model clusters and rear-ends once, the artificial fish and

its corresponding F value after two behaviors are obtained, and the largest one and the corresponding artificial fish are selected as the final result. Furthermore, the current F value and the bulletin board F value are compared, and the largest of the two is recorded on the bulletin board. From the above analysis, we know that the artificial fish is added to the dynamic field of view analysis in this study, which effectively improves the convergence speed and speeds up the optimal solution query speed. However, compared with the fish school algorithm with a smaller field of view, this study does not have an accuracy advantage. The results obtained may be the same as the traditional fish school algorithm, but it has advantages in convergence speed and running time.

When the judgment value obtained by the operation reaches the maximum value set by the system, iteration conditions are satisfied, and the median iteration is then performed. The improvement idea of this research is: in the initial stage of the algorithm, use a large field of view to quickly find a rough solution, effectively reduce the convergence time, determine the target range, and gradually improve the iteration step, gradually reduce the field of view, and perform a fine target search until the optimal solution is found.

VI. ALGORITHM SIMULATION EXPERIMENT AND PERFORMANCE ANALYSIS

During system verification, the simulation of the K-means clustering algorithm based on feedback is mainly divided into the following steps:

(1) First, the network initialization process is performed, system parameter settings are made, and wireless sensor nodes are randomly distributed into the corresponding area. (2) By using the K-means algorithm, the main cluster head candidate area is found, and the initial main cluster head and cluster head are determined. (3) During data transmission, the cluster head mainly performs intra-cluster information fusion and transmits information until it is passed to the base station. (4) If the death rate of nodes in the system exceeds 70%, the system is disabled by default and the process is terminated, otherwise the scoring function is run and the cluster head and main cluster head are continuously selected. The system operation process can be expressed as shown in Figure 7.

For the purpose of detecting the validity of the protocol, the K-means clustering algorithm based on feedback is compared with the K-means algorithm and the LEACH algorithm in the test analysis of this research. The experimental software is matlab. Experimental research shows that the algorithm in this paper can effectively reduce the system energy consumption of the wireless sensor network and effectively increase the system life.

In this research experiment, a monitoring area of size $0 \leq x \leq 100, 0 \leq y \leq 100$ is selected, and its base station position is selected at a coordinate (50, 50), and 100 sensor nodes are randomly distributed throughout the experimental area. Table 2 shows the experimental communication parameters. There are 5 cluster head candidate nodes, and 4 clusters are

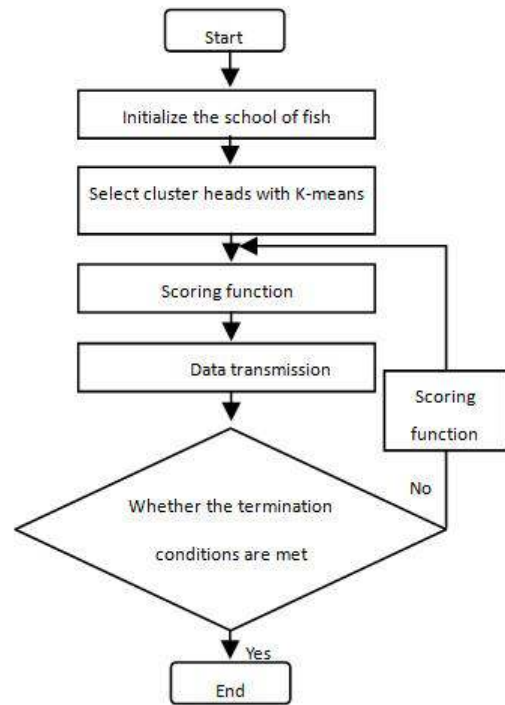


FIGURE 7. K-means algorithm flowchart based on feedback.

TABLE 2. Simulation parameters.

Types of	parameter	value
Network	Network size	100*100
	Topology	Total number of nodes Number of clusters Number of nodes in the main cluster head group Base station location node
Radio model	Energy of each initial node	0.02J
	Received and transmitted energy %	50*0.00000000001J / bit
	Dissipated energy%	10*0.00000000001J. bit
	Fusion energy	5*0.00000000001J / bit
Other parameters	Cluster head probability	0.1
	Fusion rate	0.6
	Controlling message size	32bit
	Data information size	4000bit

also specified. The definition is given, the clustering opportunity of each node in the cluster is set to 10%, and one message is 4000bit. The degree of information fusion is 0.6, and each node starts with.02J energy. The distribution of nodes is non-directional, and the layout is likely to be different each time. This paper chooses a graph of experimental results as a reference. When the number of nodes in the network is only 30%, the algorithm defaults to the network being unable to work. There are four clusters in FIG. 8, and the nodes of each cluster have a specific shape. The cluster head in each

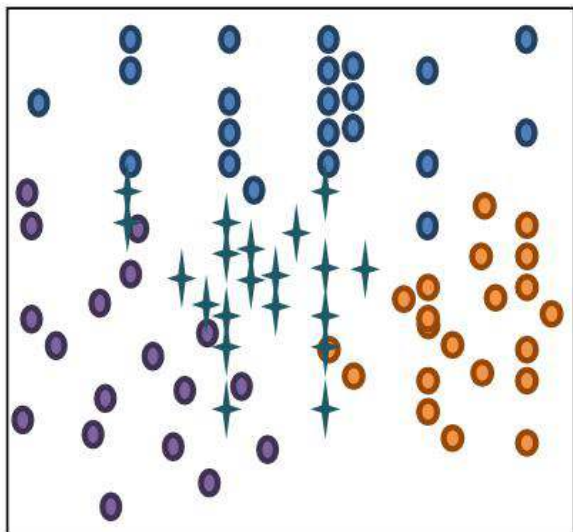


FIGURE 8. Clustering result image.

cluster is different from the other nodes. Secondly, there are 5 circular points that are the main cluster head selection range.

Similarly, a design experiment was carried out to prove the artificial fish school algorithm. This study builds a 100 * 100 pixel test area. There are 200 nodes in this area, and the radius of each node is about 15. In the experiment, several points were selected to increase the coverage while minimizing the node utilization. The length of an artificial fish in the experiment is about 200, so the distance between two artificial fish is expressed as the Hamming distance of two binary sequences. We set the number of fish to 20, the maximum number of cycles to 30, the field of view to 60, and the maximum number of attempts to 50. After setting the above parameters, the test is started.

A. SIMULATION RESULTS AND ANALYSIS

Due to the chance of the experiments, the number of experiments was increased to 50, and the results of 50 experiments were averaged to obtain the corresponding results. Figure 9 shows the number of death rounds of the first node. It is not difficult to see from the figure that the algorithm proposed in this study has a significantly longer death time for the first time in network operation than the original K-means algorithm and the LEACH algorithm. From the overall effect, the network runtime determines the node survival time in a sense.

It can be known from Figure 10 that the proposed algorithm shows that half of the nodes disappear in the network operation time is longer than that based on the original K-means algorithm and LEACH algorithm, and the data is longer than 13.6% and 30%, respectively. Based on the conclusion that the first node consumes time, it can be concluded that the proposed algorithm is more stable.

The following conclusions are drawn from Figure 11: the operation of nodes in the network is an irregular process and does not follow any function curve. However, from the figure,

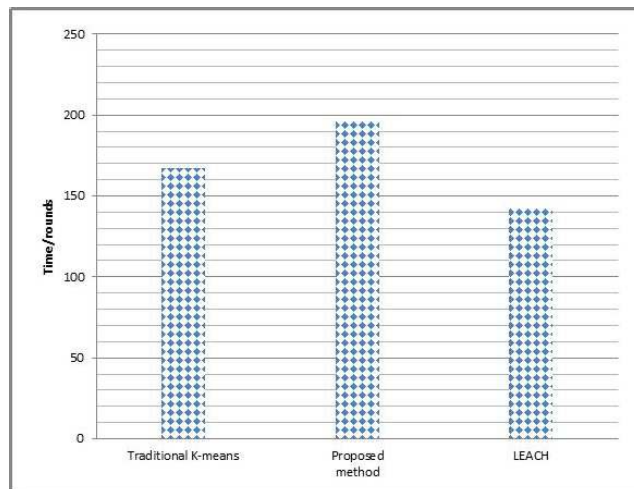


FIGURE 9. Number of death rounds for the first node.

we can still see that the curve of the proposed algorithm has a smaller overall slope, a smaller slope, and more rounds, so it is more stable than the other two algorithms. The most obvious feature from the figure is that the proposed algorithm takes longer than the original K-means and LEACH algorithms.

This algorithm defines that 70% of the nodes die is the network paralysis. Then, from Figure 12, the following conclusions are drawn: 70% of the nodes die in the network operation for longer than the original K-means algorithm and LEACH algorithm, and the data are longer than 25% and 35%, respectively. Therefore, the algorithm proposed in this paper has a longer life and achieves the required purpose of reducing energy consumption.

Based on the above, the algorithm proposed in this paper is superior to the original K-means algorithm and the LEACH protocol in terms of running time, the death of the first node, the death of half of the nodes, and the death of 70% of the nodes, or the stability of the entire network.

At present, it is always a research hotspot to reduce the energy consumption of the network as much as possible and increase the utilization of node energy. In this paper, we use the K-means algorithm to determine the required clusters in the network, and then select the current cluster head and main cluster head according to the feedback function each round. The cluster head fuses the information and transmits it to the main cluster head, and the main cluster head is only responsible for transmitting the information to the base station. The experiments show that the feedback-based K-means clustering routing in this paper has improved the running time compared with the original K-means algorithm and LEACH algorithm, and the network stability is good.

As shown in the figure shown in Figure 13, the square area represents the area to be detected, the blue circle represents the node, and the node perception range is the gray circle area. It is not difficult to see from the figure that there are overlapping areas between different nodes. If no node selection is performed, it will cause a serious waste of resources, so the

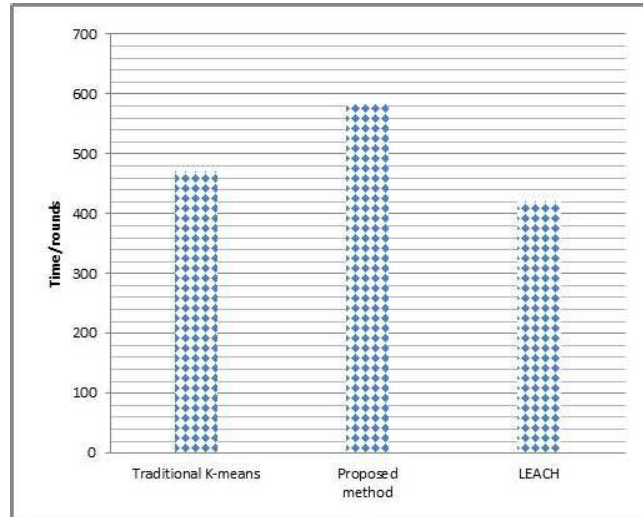


FIGURE 10. Number of rounds when 50% of nodes die.

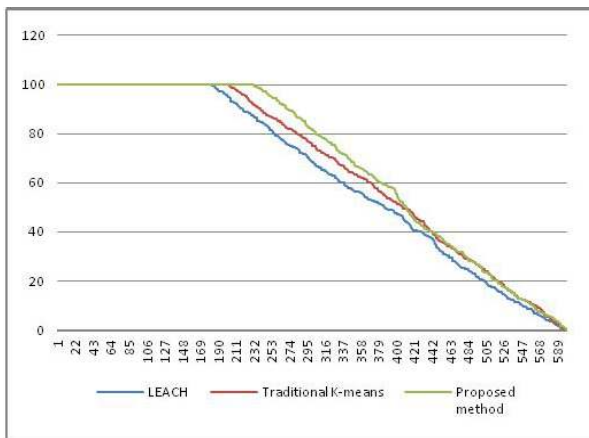


FIGURE 11. Number of nodes during the entire network operation.

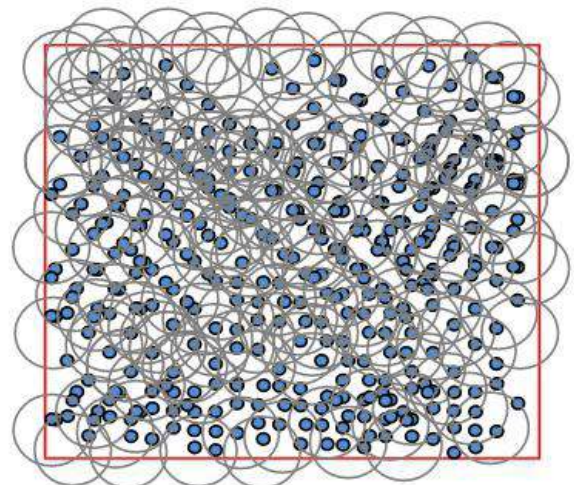


FIGURE 13. Schematic diagram of node selection (not preferred).

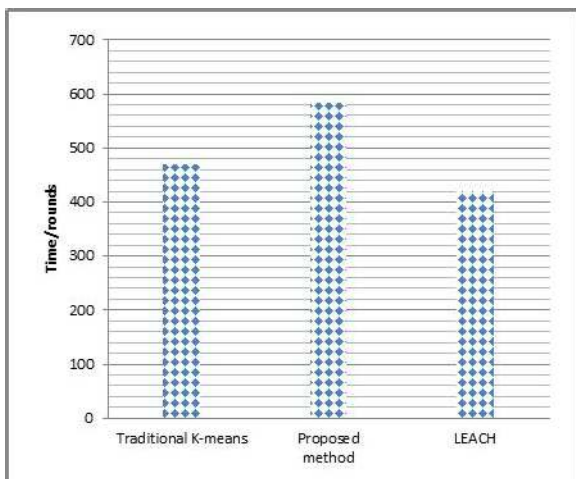


FIGURE 12. Number of rounds when 70% of nodes die.

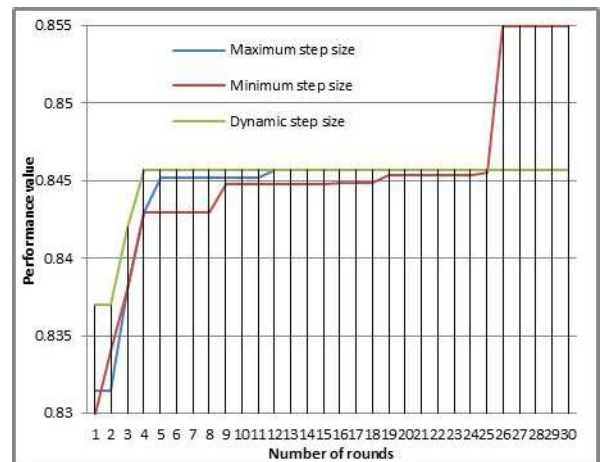


FIGURE 14. Effect of artificial fish vision on performance.

fish school algorithm is applied to the node optimization selection process.

Figure 14 shows the effect of artificial fish school field of vision on performance obtained by dynamic field of vision.

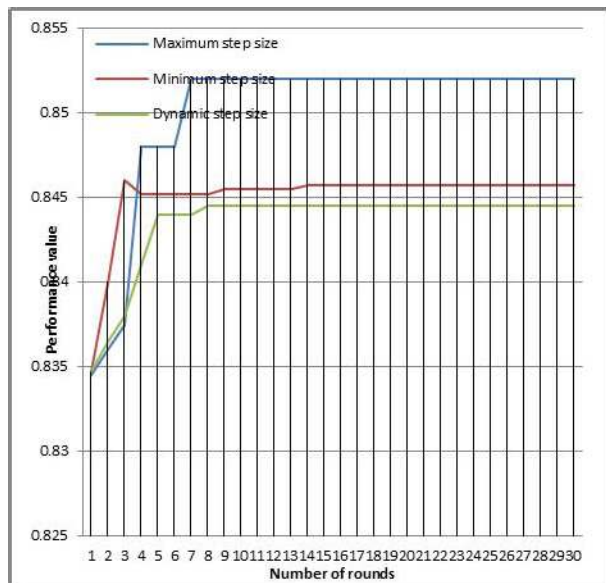


FIGURE 15. Effect of artificial fish swimming step size on performance.

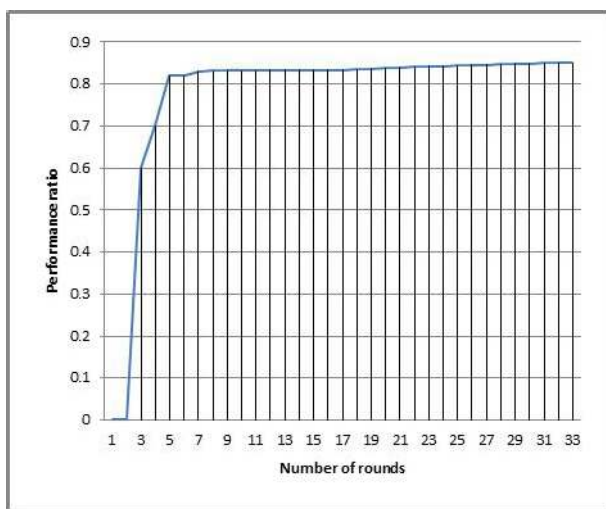


FIGURE 16. Performance image under dynamic step size and field of view.

As can be seen from the figure, the performance values obtained in the minimum field of view are slightly better than the performance values in the maximum field of view and the dynamic field of view. The best performance (corresponding to the convergence position) is found first in the dynamic field of view, followed by the largest field of view, and the smallest field of view finds the best performance value at the latest.

Figure 15 is a comparison chart of the number of rounds of performance with and without dynamic step size. It can be seen from Figure that the performance value obtained under the dynamic step size is slightly better than the performance values under the maximum step size and the minimum step size. The best performance (corresponding to the convergence position) is found first in the dynamic field of view, followed by the largest field of view, and the smallest field of view finds the best performance value at the latest.

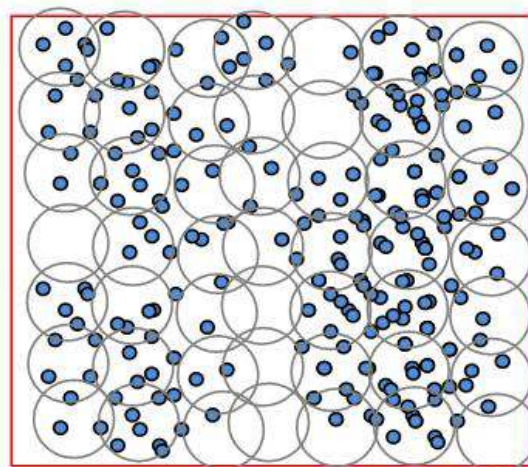


FIGURE 17. Final node selection in dynamic step size and field of view.

Combining Figure 14 and Figure 15, it can be found that a satisfactory solution can be found most quickly using dynamic parameters. Under the largest parameters, the reason for not being able to find the solution quickly may be the best advantage being jumped over. Under the minimum parameters, the reason why the best solution cannot be obtained may be that it is trapped in the local optimal solution and cannot be jumped out. It is not difficult to find that dynamic parameters can alleviate the two shortcomings of solving speed and falling into local optimum.

Figure 16 is a performance-round graph obtained by using both dynamic field of view and dynamic step size. Figure 17 is the final node selection result. It can be seen from Figure 16 that the initial parameter is large, and the performance value increases rapidly. As the number of rounds increases, the parameter decreases, and fine tuning is started. It can be seen from Figure 17 that with only 35.5% nodes, 98% of the area to be monitored can be achieved. Under the basic premise of full coverage, the use of nodes is greatly saved, and energy consumption is reduced. That is, without affecting the network coverage, the network coverage can be reduced as much as possible to meet the required requirements.

VII. CONCLUSION

This study analyzes the protocol routing of wireless sensor networks and starts from the perspective of cluster routing to explore the classic ACH protocol and analyze its advantages and disadvantages. Based on this, a clustering process based on the K-means algorithm is proposed, and a scoring function is added to the algorithm to dynamically adjust the system to achieve stable operation of the system and transmit information while reducing system energy consumption. Secondly, in view of network node consumption, this article uses artificial fish swarm algorithm to judge the working condition of the nodes. In addition, under the premise of determining the distribution of the nodes, the utilization rate of the nodes is reduced, and the network coverage is not changed, and the

convergence speed is improved without reducing the accuracy, thereby achieving the expected goals. The innovation of the article proposes an improved artificial fish swarm algorithm for the coverage problem of wireless sensor networks. This algorithm can effectively improve the system life, reduce the running time, and quickly obtain the optimal solution point set. From the experimental solution results, the research algorithm has certain practical effects, and it can be applied to subsequent wireless sensor network coverage optimization.

REFERENCES

- [1] S. Diwakaran, B. Perumal, and K. Vimala Devi, "A cluster prediction model-based data collection for energy efficient wireless sensor network," *J. Supercomput.*, vol. 75, no. 6, pp. 3302–3316, Jun. 2019.
- [2] C. Sun, "A time variant log-linear learning approach to the SET K-COVER problem in wireless sensor networks," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1316–1325, Apr. 2018.
- [3] W. D. Fang, F. R. Li, and L. H. Shan, "Anonymous communication technology for wireless sensor network: A survey," *J. Beijing Univ. Posts Telecommun.*, vol. 40, no. 1, pp. 1–17, 2017.
- [4] J. Roselin, P. Latha, and S. Benitta, "Maximizing the wireless sensor networks lifetime through energy efficient connected coverage," *Ad Hoc Netw.*, vol. 62, pp. 1–10, Jul. 2017.
- [5] H. Wang, X. Ding, and L. Wang, "Target connected coverage algorithm for wireless sensor networks," *Nanjing Li Gong Daxue Xuebao/J. Nanjing Univ. Sci. Technol.*, vol. 41, no. 3, pp. 285–293, 2017.
- [6] X. Gao, Z. Chen, F. Wu, and G. Chen, "Energy efficient algorithms for k -sink minimum movement target coverage problem in mobile sensor network," *IEEE/ACM Trans. Netw.*, vol. 25, no. 6, pp. 3616–3627, Dec. 2017.
- [7] F. Yap and H.-H. Yen, "Novel visual sensor coverage and deployment in time aware PTZ wireless visual sensor networks," *Sensors*, vol. 17, no. 12, p. 64, Dec. 2016.
- [8] M. Usman, N. Yang, M. A. Jan, X. He, M. Xu, and K.-M. Lam, "A joint framework for QoS and QoE for video transmission over wireless multimedia sensor networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 4, pp. 746–759, Apr. 2018.
- [9] C. I. Weng, C. Y. Chang, and C. Y. Hsiao, "On-supporting energy balanced k -barrier coverage in wireless sensor networks," *IEEE Access*, vol. 6, pp. 13261–13274, 2018.
- [10] F. El Hajji, C. Leghris, and K. Douzi, "Adaptive routing protocol for lifetime maximization in multi-constraint wireless sensor networks," *J. Commun. Inf. Netw.*, vol. 3, no. 1, pp. 67–83, Mar. 2018.
- [11] S. Lykov, Y. Asakura, and S. Hanaoka, "Positioning in wireless sensor network for human sensing problem," *Transp. Res. Procedia*, vol. 21, pp. 56–64, 2017.
- [12] H. Prihtyadi and M. Djamal, "The reliability of wireless sensor network on pipeline monitoring system," *J. Math. Fundam. Sci.*, vol. 49, no. 1, pp. 51–56, 2017.
- [13] T. Dong, "Assessment of data reliability of wireless sensor network for bioinformatics," *Int. J. Bioautomat.*, vol. 21, no. 3, pp. 241–250, 2017.
- [14] N. Kaleeswari and K. Baskaran, "An optimized secure data gathering scheme for wireless sensor network," *J. Comput. Theor. Nanosci.*, vol. 14, no. 3, pp. 1653–1660, Mar. 2017.
- [15] N. Qin and K. Chen, "A wireless sensor network location algorithm based on insufficient fingerprint information," *Mod. Phys. Lett. B*, vol. 32, nos. 34–36, Dec. 2018, Art. no. 1840093.
- [16] Y. Alsaba, S. K. A. Rahim, and C. Y. Leow, "Beamforming in wireless energy harvesting communications systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1329–1360, 2nd Quart., 2018.
- [17] L. Lombardo, S. Corbellini, M. Parvis, A. Elsayed, E. Angelini, and S. Grassini, "Wireless sensor network for distributed environmental monitoring," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1214–1222, May 2018.
- [18] L. Dong, G. Liu, X. Cui, and T. Li, "G-skyline query over data stream in wireless sensor network," *Wireless Netw.*, vol. 26, no. 1, pp. 129–144, Jan. 2020, doi: 10.1007/s11276-018-1784-2.
- [19] S. A. Kumar and P. Ilango, "The impact of wireless sensor network in the field of precision agriculture: A review," *Wireless Pers. Commun.*, vol. 98, no. 23, pp. 685–698, 2017.
- [20] L. Bracciale, A. Catini, G. Gentile, and P. Loreti, "Delay tolerant wireless sensor network for animal monitoring: The pink iguana case," in *Proc. Int. Conf. Appl. Electron. Pervading Ind., Environ. Soc.*, vol. 429, Cham, Switzerland: Springer, 2017, pp. 18–26.
- [21] K. Martinez, J. K. Hart, P. J. Basford, G. M. Bragg, T. Ward, and D. S. Young, "A geophone wireless sensor network for investigating glacier stick-slip motion," *Comput. Geosci.*, vol. 105, pp. 103–112, Aug. 2017.
- [22] R. R. Sahoo, S. Ray, S. Sarkar, and S. K. Bhoi, "Guard against trust management vulnerabilities in Wireless Sensor Network," *Arabian J. Sci. Eng.*, vol. 43, no. 12, pp. 7229–7251, Dec. 2018.
- [23] M. Neshat, G. Sepidnam, M. Sargolzaei, and A. N. Toosi, "Artificial fish swarm algorithm: A survey of the state-of-the-art, hybridization, combinatorial and indicative applications," *Artif. Intell. Rev.*, vol. 42, no. 4, pp. 965–997, Dec. 2014.
- [24] S. Thompson, M. E. Celebi, and K. H. Buck, "Fast color quantization using MacQueen's k -means algorithm," *J. Real-Time Image Process.*, vol. 16, no. 64, pp. 1–16, Oct. 2019.
- [25] H. Sun and J. Zhao, "Application of particle sharing based particle swarm frog leaping hybrid optimization algorithm in wireless sensor network coverage optimization," *J. Inf. Comput. Sci.*, vol. 8, no. 14, pp. 3181–3188, 2011.



YINGYING FENG was born in Shandong, China, in 1982. She received the bachelor's degree from Changjiang University, in 2004, and the master's degree from the China University of Geosciences, in 2007. She currently works at Fuyang Normal University. She has published five papers, one of which has been indexed by SCI. Her research interests are included computer graphics, and image processing and video analysis.



SHASHA ZHAO received the M.S. degree in computer application of professional engineering from the China University of Geosciences, Wuhan, China. She is currently an Associate Professor with the Information Engineering College, Fuyang Normal College. Her research interest is mainly in the area of computer software and computer networks. She has published several research articles in scholarly journals in the above research areas.



HUI LIU received the M.S. degree in computer software and theory from the Hefei University of Technology, Hefei, China. He is currently a Lecturer with the Information Engineering College, Fuyang Normal College. His research interest is mainly in the areas of digital image processing and intelligence algorithm. He has published several research articles in scholarly journals in the above research areas.

...