Buletin Teknik Elektro dan Informatika
(Bulletin of Electrical Engineering and Informatics)
Vol. 3, No. 2, June 2014, pp. 77~90
ISSN: 2089-3191

Study on the Rough-set-based Clustering Algorithm for Sensor Networks

Fengmei Liang*, Liyuan Zhang, Peng Sun
Taiyuan University of Technology/College of Information Engneering, Tai Yuan, China
*Corresponding author, email: m_liang@163.com

Abstract

The traditional clustering algorithm is a very typical level routing algorithm in wireless sensor networks (WSN). On the basis of the classical LEACH (Low Energy Adaptive Clustering Hierarchy) algorithm, this paper proposes an energy efficient clustering algorithm in WSN. Through the introduction of rough set, the new algorithm mainly introduces how to confirm an optimized strategy to choose the cluster head effectively by the simplified decision table. That is to say, by discrete normalized data preprocessing of attribute value, getting discretization decision table. Finally, the results from simulated experiments show that the clustering algorithm based on rough set theory can optimize the clustering algorithm in network data. That is to say, the rough-set-based clustering algorithm can effectively choose the cluster head, balance the energy of the nodes in the cluster and prolong the lifetime of sensor networks.

Keywords*: WSN, rough set, clustering algorithm, LEACH, network lifetime, energy efficient*

1. Introduction

In the structure of the WSN, routing techniques of the network's level are vitally important. Clustering algorithm is a crucial level routing algorithm of the WSN, which separates all nodes in the network into connected multiple levels by certain rules or methods. Several adjacent nodes constitute a cluster and then choose a cluster head in every cluster; the others are the member nodes. The WSN find fewer nodes as the cluster head nodes by the clustering algorithm as soon as possible. The cluster head nodes are responsible for the nodes connection and routes choose. The other member nodes turn off the transmitter and sleep to achieve the aim of energy protection when they are idle [1].

With advances in processor and wireless communication technologies, WSN will be used everywhere in the future life. It is a type of wireless network and composed of a collection of sensor nodes. Within a WSN, nodes collaborate amongst themselves to accomplish a common task. WSN has enormous potential to improve how we use information from the real world. It is expected to apply the wireless sensor network technology to various application areas such as the military, environment, medical service, and business. Among the various scopes, one of the major applications of sensor network is to collect information periodically from a remote terrain where each node continually sense the environment and sends back this data to the base station which is usually located at considerably far from the target field. However, sensor nodes are typically battery-powered and operate without attendance for a relatively long period of time. Therefore, energy efficiency becomes critical importance. In the wireless sensor network community, a significant focus has been put on increasing energy efficiency. In recent years, one of the major researches in WSN is energy efficient routing protocol. Thus, nodes of WSN are usually running in the dangerous environment where people cannot be close. Energy cannot be replaced. So the core issues of the studies are how to design the most optimized strategy to extend the life cycle of sensor networks [2].

As a new achievement of intelligent information processing technology, the rough set theory is a new method which proposed by Polish scientist professor Z.Pawlak and carries on analysis, inference, study and discovery to incomplete data. Rough set theory is used for data analysis. It copes with the vague and inaccurate problems. Rough set theory can find potential and valuable knowledge from a large number of data mining. It mainly researches the

Received January 24, 2014; Revised April 26, 2014; Accepted May 21, 2014

inaccurate, incomplete and untrue of the objects [3]. The theory provides people with a scientific logic thinking and intelligent information processing.

The traditional LEACH clustering algorithm does not consider the problem of energy consumption and topology of the sensor. This paper based on LEACH algorithm, presents a novel clustering algorithm on the basis of rough set theory. The algorithm reduces the uncertainty of the data; it can effectively choose the cluster head from the cluster, balance the nodes energy and the consumption energy, and prolong the living time of WSN.

2. Review of Present Research Status

2.1. The Research Status and Characteristics of WSN

WSN is constitute on a large number of low-cost micro-sensor nodes where deployed in the monitoring area. It is a multi-hop and self-organization network system that formed by a wireless communication method. And its purpose is perception, acquisition and processing the information of the perceived object in the network coverage area, then the information is sent to the viewer. Sensors, perception object and the observer constitute the three elements of the wireless sensor networks. It is different from the traditional wireless networks, which have high service quality and efficient bandwidth utilization. WSN is first and foremost considered with energy-saving. In the structure of the WSN, routing techniques of the network's level are very important. The level routing system of the WSN, which separates all nodes in the network into connected multiple levels by certain rules or methods. Several adjacent nodes constitute a cluster and then choose a cluster head in every cluster; the others are the member nodes.

As the shown, Figure 1 is a wireless sensor network. Figure 2 is the energy consumptions of the sensor nodes. Figure 3 is the structure of level routing system.

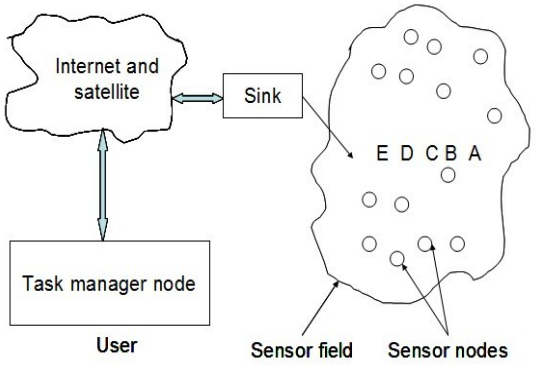

Figure 1. A wireless sensor network

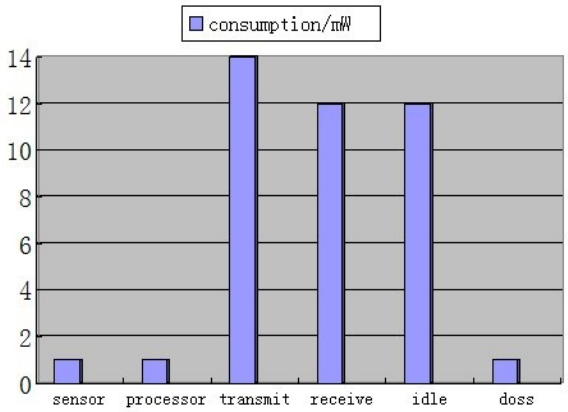

Figure 2. Energy consumption of the sensor nodes

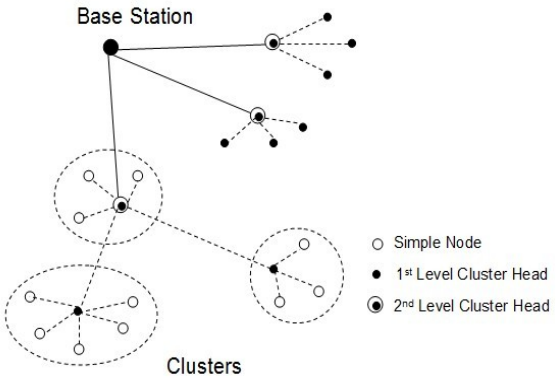

Figure 3. Level routing system structure

The advantage of clustering algorithm for WSN is as follows: firstly, member nodes can turn off the communication module on the most moment. The network, which constitutes an even higher level, is responsible for long-distance routing send of data. This will ensure the original data communications, and to a large extent, saves the network energy as well. Secondly, cluster head combines the data of the member nodes before forwarding. It can reduce the amount of data communication, and then save the network energy. Thirdly, the function of member nodes is relatively simple, no maintenance complex routing information. This greatly reduces the number of network routing control information and reduce the amount of communication. Fourthly, clustering topology structure is easy to manage. It is conducive to the application of distributed algorithms. It can quickly respond to the change of the system. Finally, compared with the flat route, it is easier to overcome the problems of the moving sensor nodes.

2.2. The Basic Clustering Algorithm for WSN

1) The LEACH Algorithm

The LEACH protocol is the earliest classical clustering protocol. It was proposed by Wendi at MIT. It is a low-power adaptive level routing protocol that designed for WSN. The basic idea is that the cluster head is randomly selected by the circular way and the cluster was re-established. The result of the algorithm is that the load energy is average distributed to each node, and then reducing the energy consumption and extending the life cycle.

LEACH makes the reconstruction process of the cluster called 'round'. Every round is divided into two stages: building stage of cluster and stable stage of data transmission. In order to reduce protocol spends, it should make the time of building stage less than stable stage.

In cluster building stage, the sensor nodes generate a random number between 0 and 1. The random number is compared with a threshold value T(n). If the random number is less than the threshold, the node becomes a cluster head in the current round. T(n) is defined as:

$$T(n)=\begin{cases}\dfrac{P}{1-P\times\left[r \bmod (1/P)\right]}, n\in G\\ 0, \qquad\qquad \text{otherwise}\end{cases} \tag{1}$$

In the formula, P is a percentage of the nodes become the cluster head nodes and r is the current round number. The non-selected nodes in the nearest 1/p round belong to set G. When cluster head node is selected, it will broadcast himself as the cluster head, the other nodes decide to join which cluster according to receiving the strength of the broadcast. At the same time, they inform the corresponding cluster head to complete the process of establishing stage. And then, cluster head node assign data transmission time to the members of the cluster by the TDMA way[4].

In the stable stage, the sensor nodes send the collected data to the cluster head nodes. After data fusion, the message will be sent to the gathered node, gathered node transmit data to the monitoring center for data processing [5]. The stable stage lasts for a period of time. The

networks get back into the establishment phase of the next round of cluster reconstruction. This is a continuous cycle.

LEACH assumes that the member nodes send data to the cluster head when the time interval belongs to them. However, the real situation is that the nodes may be sent after monitoring the event. The cluster head nodes of LEACH are randomly generated and it is not considered the residual energy of the nodes. As a result, the energy of the cluster head node which has the less residual energy may be consumed as soon as possible.

The flow chart is as follows:

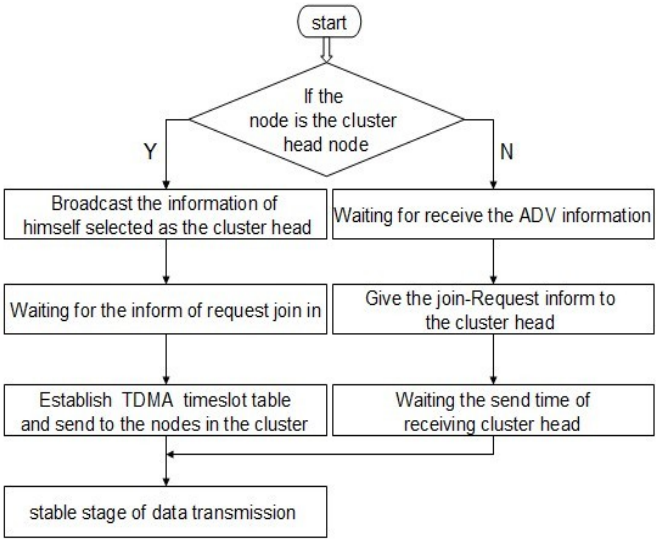

Figure 4. The process of cluster formation

The LEACH protocol is certainly a classical clustering protocol in WSN. Its basic ideas are contained as follows: the cluster head will be selected based on its energy, and the selection will be held in a circular way. Thus, energy consumption of each node tends to be balanced. So it can lower the energy consumption and extend the network lifetime. Though the LEACH protocol has many advantages compared with other protocols, it has some disadvantages. That is, it is generated randomly; the head selection algorithm of LEACH is based on a threshold value in a specific probability [6].

2) Hausdorff Algorithm

The Hausdorff algorithm is a distributed data collection algorithm that based on the location of the node, communication effectiveness and network connectivity. The algorithm is divided into three parts: firstly, the Hausdorff algorithm divides nodes into several static clusters. Then using a Euclidean distance to calculate the distance between two points and using the Hausdorff distance to calculate the distance between the two points set. Hausdorff distance is being a measure of clustering. Secondly, in the life of the entire network carve, the processing of clustering only executed once, and the cluster head in each cluster is optimal rotation scheduling by cluster members. The algorithm is based on the closeness of the adjacent node and the node's residual energy to select the cluster head in each cycle. Finally, the network skeleton which constitutes the cluster heads is on a regular basis to the collection, integration and forward data to the base station, the communication between them is using the minimum power routing algorithm, which use the Dijkstra shortest path algorithm.

3) RDCA Algorithm

The response distribution load balancing clustering algorithm (RDCA) is also a distributed algorithm. The algorithm has been improved the DCA algorithm. It does not need to know the location information of the nodes in advance, but only according to the local topology

information quickly distributed cluster head election and divided the cluster according to the cost function. The algorithm has good load balance performance and smaller protocol overhead. Compared with LEACH protocol, it can reduce energy consumption and extend the network lifetime.

4) The Clustering Algorithm based on RBF

The algorithm is a centralized and location-based clustering algorithm, which suitable for small-scale WSN. Clustering decisions calculated by the base station and each node have been advised of the location before the algorithm execution. When the base station collects the residual energy and the location information of all nodes in the network, set up RBF neural network consisted with the three levels of the output layer, hidden layer and output layer. Then, calculate the probability of each node becomes a cluster head.

3. Rough Set Theory

Rough set is a new data analysis theory that proposed by Pawlak in 1982. It always solved the vague and inaccurate problems. Rough set theory provides a strict mathematics theory method. It is an effective tool which handles imperfect information such as inaccurate, inconsistent, incomplete and so on. The core of rough set describes uncertainty of the thing by the upper approximation and the lower approximation. The concept of rough set can be defined by means of topological operations [7]. In this approach, vagueness is expressed by a boundary region of a set.

Rough set theory has many characteristics, they are: mainly deals with incomplete data and has many variable data; copes with imprecise data, including the certainty and uncertainty possible; gets the minimum reduce and various different granularities level of knowledge; announces the pattern of the easy concept and operation from experiment data; produces precise and easy to check and confirm the rules, applicable to automatically generated rules in the intelligent control. Now, introduce some basic concepts of the rough set theory.

3.1. The Information System and the Decision Table

Rough set is an object-attribute relationship operation based on the form of an information system. The information system is a set of objects describes a set of attributes, usually expresses as a form of binary information table, and uses a four-group to define, that is S=(U, A, V, f). U is a set of non-empty finite objects (called universe). $U=\{x_1,x_2,\cdots,x_n\}$, x_i is objects (binary group). A is a non-empty finite set that described objects, divided into two disjoint subset. That is the condition attribute C and the decision attribute D. $A=C\cup D$. V is a set of attribute values. $V=\cup V_a$. V_a is the value of attribute a, f is a function, that is $U\bullet A\to V$ is an information function, it is specified the value of every object x under the attribute a in the U, that is $\forall a\in A$, $x_i\in U$, $f(x_i,a)\in V_a$.

If the attribute set can be divided into condition attribute C and decision attribute D, that is $A=C\cup D$. The information system S is also called the knowledge express system or the attribute value table [8]. The decision table is a special knowledge express system, rough set based data analysis starts from a data table called a decision table, means that the knowledge express system what has condition attribute and decision attribute can be expressed decision table. If a decision rule uniquely determines the decision in terms of conditions the decision rule is certain. Otherwise the decision rule is uncertain. Decision rules are closely connected with approximations. Roughly speaking, certain decision rules describe lower approximation of decisions in terms of conditions, whereas uncertain decision rules refer to the boundary region of decisions[8].

3.2. The Concept of Rough Sets

Rough set regards knowledge as the division of the universe. Thus knowledge is thought to possess granularity and the imprecision of knowledge is caused by large knowledge granularity. If $\underline{R}X\neq\overline{RX}$, X is called the R-term rough set, otherwise X is called the R-term precise set. R is the family of equivalence relation in the universe U. Obviously rough sets, in contrast to precise sets, cannot be characterized in terms of information about their elements.

With any rough set a pair of precise sets, called the lower and the upper approximation of the rough set, is associated.

3.3. Approximation

Approximations are fundamental concepts of rough set theory [9]. The lower approximation consists of all objects which surely belong to the set and the upper approximation contains all objects which possibly belong to the set. The boundary region of the rough set is composed of the difference set between the upper and the lower approximation. The mathematical definition is as follows:

In the given knowledge base $K = (U, R)$, there is an equivalence relation $R \in IND(K)$ with regard to $X \neq \varnothing$ and $X \subseteq U$, then $\underline{R}X = \cup\{Y \in U / R | Y \in X\}$ is called the R-term lower approximation of X and $\overline{R}X = \cup\{Y \in U / R | Y \cap X \neq \varnothing\}$ is called the R-term upper approximation of X [10].

3.4. Indiscernible Relationship

The indiscernible relationship is used to express that it cannot distinguish with some objects of the known information system because of lacking certain knowledge. In fact, it is an equivalence relation. The definition is as follows:

For approximation space K= (U, R), amusing $P \subseteq R$ and $P \neq \varnothing$, the intersection of all equivalence relation from P is an equivalence relation, called an indiscernible relationship of P[9]. Recorded as $IND(P)$, that is:

$$[x]_{IND(P)} = \cap [x]_R \quad (2)$$

3.5. Knowledge Reduction and the Core of Rough Set

Knowledge reduction is a significant concept of rough set used for data analysis. It is divided into attribute reduction and attributes value reduction. Under the premise that the indiscernibility relation, classification, and decision-making ability of decision table information system is unchanged, attribute reduction refers to eliminate irrelevant or unimportant redundant attributes, and get the minimum subset of condition attribute. Attribute reduction is based on the importance to reduce. The attributes of an information system in the decision table are not all important. Some attributes may be redundant, can delete after simplified; only part of the conditions property must be retained.

For an attribute subset $P \subseteq A$, if there exists the relation $Q = P - r$, satisfying $Q \subset P$ and making $IND(Q) = IND(P)$ tenable, while Q is the best subset, then Q is called the reduction of P, denoted with $red(P)$. The intersection of all reduction attribute sets in P is called the core of P, denoted with $core(P)$ [10].

The specific concepts in the theory of rough sets are as follows in Table 1 [11].

Table 1. Rough set concepts

Symbol	Concept	Definition
U	Universe of Discourse	Set of all objects in the System
A	Set of all Attributes	Attributes of each object $A = C \cup D$
C	Set of Condition Attributes	$C \subseteq A$
D	Set of Decision Attributes	$D \subseteq A$
Ind	Indiscernibility Relation	$Ind(P) = \cap P, P \subseteq R$ and $P \neq \varnothing$
$\underline{R}$	Lower Approximation	$\underline{R}(X) = \cup\{Y_i \in U | Ind(R) : Y_i \subseteq X\}$
$\overline{R}$	Upper Approximation	$\overline{R}(X) = \cup\{Y_i \in U | Ind(R) : Y_i \cap X \neq \varnothing\}$

4. Case Analysis of the Algorithm Based on Rough Set Theory

The application of power control technology in WSN, there are many factors to decide many sensor nodes in which progression to determine the attenuation and increase its own power, now need to select the cluster head. The process of the cluster election, it is usually considered the information such as total energy, surplus energy, geographical position, signal strength, stability, and communication costs and so on. With the traditional selection methods, it supposes that the information is accurate. However, the WSN is a dynamic network. The different environments and backgrounds also can make impact on the selection of the cluster head [13].

Rough set theory provides an effective and novel theory to the field of data mining (DM) and knowledge discovery in databases (KDD). DM is mainly focus on the database and management information systems, statistical community and the profession of data analysis, while KDD focuses primarily on artificial intelligence and machine learning. Rough set is based on the classification ability of the object to determine the reliability of the subject of knowledge. In the paper, rough set theory is applied to the WSN power control technology analysis [12].

The process of DM with rough set theory, amusing identifies the cluster head node receives the node information of the surplus energy, the size of the signal strength, communication costs and geographical distances in four areas. Through discrete normalized data preprocessing of attribute value, get discretization decision table. Now, introduce the new clustering algorithm based on rough set and present how to use the rough set theory to choose a cluster head. The special conditions are as shown in Table 2.

Table 2. The WSN decision table

Nodes (U)	Surplus energy (y_1)	Communication costs(y_2)	signal strength (y_3)	geographical position (y_4)	Decision D(d)
x_1	1	1	1	0	1
x_2	0	0	0	1	0
x_3	0	1	1	0	1
x_4	1	1	0	0	0
x_5	1	0	0	1	0
x_6	0	1	1	1	0
x_7	1	1	1	1	1
x_8	0	1	0	1	0

In the situation, x1, x2, ⋯, x8 is the selected nodes for WSN. In the surplus energy, 1 is considered that the energy is sufficient, 0 means that the energy is inadequate; In the signal strength, 1 represents the strong signal strength, 0 is poor; In the communication costs, 1 is defined as the small communication costs, 0 represents that the cost of communication is large; 1 of the geographical position represents that geographical location is near, 0 is considered far.

4.1. The Calculation of Lower Approximation and Dependency in Equivalent Set

a) The equivalent set of conditional attribute set C(y1, y2, y3, y4).

Each object in condition attribute set C does not exist equivalence relation, so each group constitutes an equivalent set, a total of eight. That is X1={x1}, X2={x2}, X3={x3}, X4={x4}, X5={x5}, X6={x6}, X7={x7}, X8={x8}

b) The equivalent relationship of decision attribute set D(d).

According to the decision attribute, the value of each object is a number of 0 or 1; divide into two kinds, there are two equivalent sets in the decision attribute set D (d). That is Y1={x1, x3,x7}, Y2={ x2,x4, x5,x6,x8}.

c) The lower approximation sets of the decision attribute equivalent set about condition attribute set C is

C_*Y_1 ={X_1, X_3, X_7} ={x_1, x_3, x_7}

C_*Y_2 ={ X_2,X_4,X_6, X_5,X_8}={ x_2,x_4, x_5,x_6,x_8}

d) The calculation of $POS(C,D)$ and $\gamma(C,D)$

$POS(C,D) = C_*Y_1 \cup C_*Y_2$ ={ x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8}

Because $Card[POS(C,D)]=8$, $Card(U)=8$

$\gamma(C,D) = Card[POS(C,D)] / Card(U) = 1$

According to the condition attributes set C, all objects of decision attribute set D can have an accurately classification. Now considers the possibility of reducing the condition attributes set

.

4.2. The Computation of Each Attribute Importance of Condition Attributes C

The computation for y1:

a) The equivalent set of condition attributes set $C-\{y_1\}$.

When lacks of attribute y1 in condition attribute C, x1 and x3, x2 and x5, x6 and x7 has equivalent relationship, total 5. The equivalent set of $C'(y_2,y_3,y_4)$ is:

X1={x1, x3}, X2={x2, x5}, X3={x4}, X4={x6, x7}, X5={x8}.

b) Equivalence of decision attribute set Y_1={x_1, x_3, x_7}, Y_2={x_2, x_4, x_5, x_6, x_8}.

c) The equivalent episode of decision attribute of lower approximation set in condition attribute set $C'(y_2,y_3,y_4)$.

$C'_*Y_1 = \{X_1\} = \{x_1, x_3\}$

$C'_*Y_2 = \{X_2, X_3, X_5\} = \{x_2, x_4, x_5, x_8\}$

d) Calculation

$POS(C-\{y_1\},D) = C'_*Y_1 \cup C'_*Y_2$ ={x1,x2,x3,x4,x5,x8};

$Card(C-\{y_1\},D) = 6$, $Card(U)=8$

$\gamma(C-\{y_1\},D) = Card(C-\{y_1\},D) / Card(U) = 3/4$

e) The importance of attribute y_1

$SGF(C-\{y_1\},D)=\gamma(C,D)-\gamma(C-\{y_1\},D)=1-3/4=1/4\neq 0$

So y1 is not the omission. The similarity to others, the condition attributes y3 and y4 is not the omission, condition attribute y2 is may be omitted.

Finally according to the data reduction, the result shows as Table 3.

Table 3. The simplified decision table

Nodes (U)	surplus energy (y_1)	signal strength (y_3)	geographical position (y_4)	Decision D(d)
x_1	1	1	0	1
x_2	0	0	1	0
x_3	0	1	0	1
x_4	1	0	0	0
x_5	1	0	1	0
x_6	0	1	1	0
x_7	1	1	1	1
x_8	0	0	1	0

The picking of the cluster head should put the surplus energy and geographical position as more important indicators for judging. It is relatively appropriate to choose cluster head on the place where nodes are more intensive. On these conditions, x_5 and x_7 node which corresponding the higher surplus energy becomes the cluster head is a better choice, and contact information of the signal intensity, finally choosing x_7 out as a cluster head.

5. The Introduce of Simulation Software

This paper uses Network Simulator version 2 (NS2) simulation platforms, selects the survival node living time and the cluster head number per round to compare LEACH algorithm with the new algorithm.

NS2 was developed by the University of UC Berkeley. It is an analog network technology free software platform. Its source code is open. NS2 is an object-oriented and discrete-time driver network simulator; in essence, it is a discrete event simulator [28]. It can use complementary C++ and Otcl language as the development language. In NS2, the simulation has been achieved, in terms of network transmission protocol, can support both TCP and UDP transport protocol; generator aspects of the flow of the traffic sources, can simulate the business model such as FTP, Telnet, CBR and VBR; in routing queue management mechanisms, can simulate Droptail, RED and CBQ. NS2 is also achieving multicast and the MAC sub-layer protocol on the LAN (Local Area Network) emulation. NS2 now has become one of the first selected software to implement network simulation in the academic field.

NS2 contains many components required for network simulation, the basis is node, simulator, link, packed format, and so on.

The steps of network simulation using NS2 are as follows [14].

Step 1: Write an OTcl script program.

Step 2: Write supplementary design program, set the time of the end of the network simulation.

Step 3: Interpretation using NS2 and begin to run the OTcl script program.

Step 4: Analysis the trace file in the simulations, obtain useful data for their own.

Step 5: The data are in the process of drawing, obtain the performance of the simulated network.

Step 6: Modifies the parameters such as the topology of the network and business model, then simulates again.

Figure 6 shows the basic procedure of network simulation with NS2[15].

6. The Simulation Experimental Results and Analysis

6.1. The Simulation Environment

Now it has a variety of tools for the simulation of WSN routing protocol, such as OPNET, MATLAB, OMNET++ and NS, etc. The simulation environment is in the range of 100 * 100. This paper uses NS2 to simulate LEACH protocol and the new strategy, and shows the performance of two protocols by comparing.

The settings of simulation parameters are as follows:

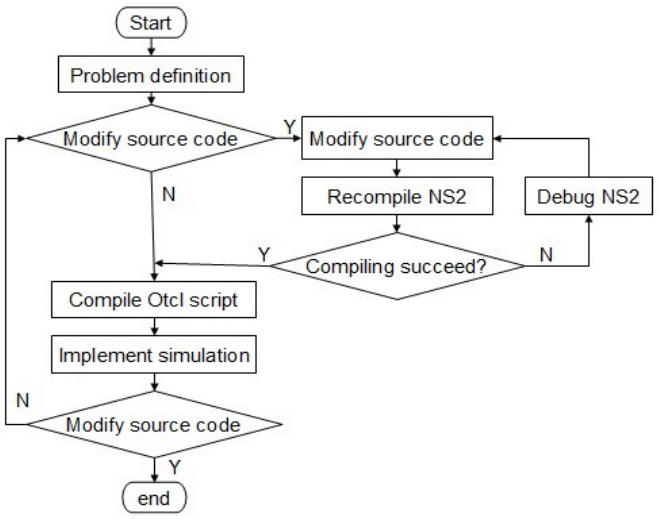

Figure 5. Process of network simulation using NS2

From this range, it randomly distributes with the scope of the 100 nodes and 500 nodes, respectively. Initial energy is 0.5 J. The number of passing the message is 2000 rounds. As showed in Figure 6 and Figure 7.

Figure 6 shows randomly distributed plot of 100 nodes in the 100m * 100m monitoring area of network. We can see from that although the location of the generated nodes is sparse or dense, there is always a forwarding node to forward data.

Figure 7 shows randomly distributed plot of 500 nodes in the 100m * 100m monitoring area of network. From this figure, we can see the spares and dense better than Figure 6.

In order to reflect the density of nodes more intuitive, this paper further simplified simulation of the map based on Figure 7. The result is as showed in Figure 8.

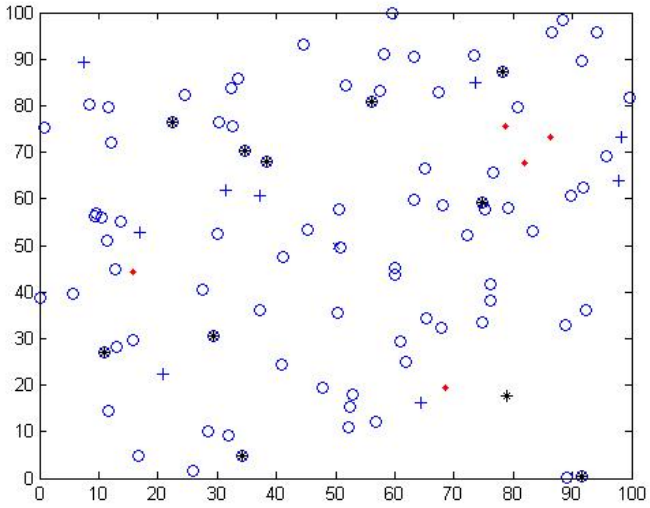

Figure 6. The process of wireless sensor network of 100 nodes

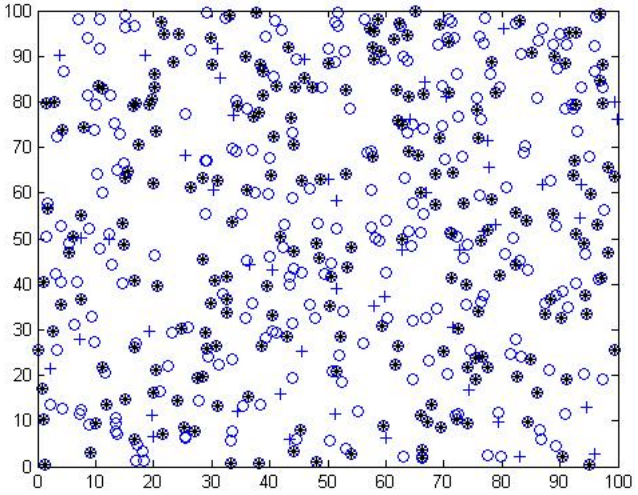

Figure 7. The process of wireless sensor network of 500 nodes

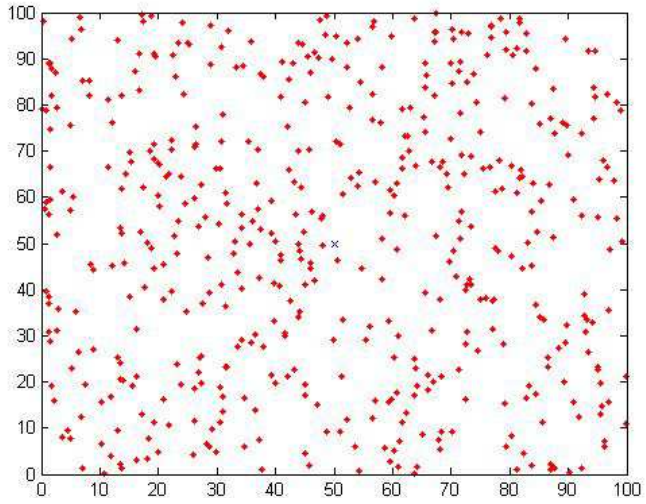

Figure 8. The simplified simulation of 500 nodes

Through several simulations, we get the average value, which is used to analyze. This article focuses on simulating and analyzing the number of nodes alive and energy consumption within the prescribed time.

6.2. Results and Analysis

The algorithm is simulated by NS2. Now makes a contrast between two agreements under the operational status.

Figure 9 and Figure 10 is simulated the node dead phenomenon of 100 nodes, while Figure 11 and Figure 12 is simulated the phenomenon of 500 nodes.

By contrast, in Figure 9, it can be found that when the operation of LEACH algorithm is less than 1000 round, it already appears node death phenomenon. In Figure 10, when using rough set algorithm, it begins to appear at 1200 wheel. It is prolonging the survival node living time. The fact shows that the algorithm based on rough set is much better than the traditional LEACH algorithm.

Similarly, Figure 12 shows that the node death phenomenon appears at 1200 wheels compared with figure 11. Meanwhile, the plot of 500 nodes is more accurate and intuitive. The four figures fully show that the Rough-set-based Clustering Algorithm is much better than LEACH algorithm. It effectively prolongs the lifetime of the network.

From these graph, it can be seen that compared to conventional LEACH algorithm, the Rough-set-based algorithm get more smoothly survival nodes curves and total energy consumption curve, indicating that the improved algorithm in the network life cycle is relatively more stable and energy consumption is more evenly, while the number of cluster head node generated by the traditional LEACH algorithm is not fixed due to the uncertainty of random selection, the consumed energy gap is relatively large, the curve is not relatively smooth.

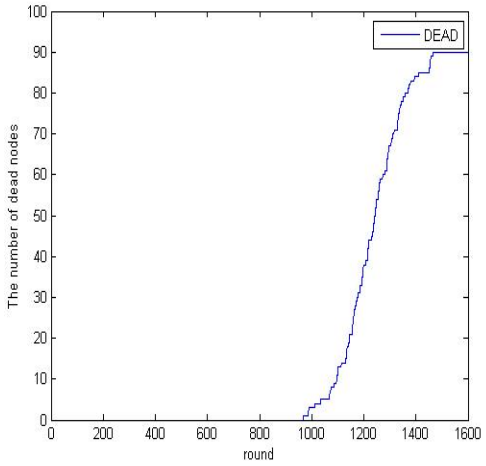

Figure 9. The dead nodes using leach per round with 100 nodes

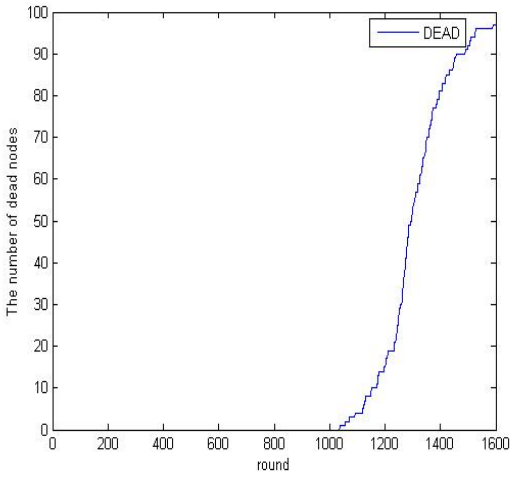

Figure 10. The dead nodes using rough sets per round with 100 nodes

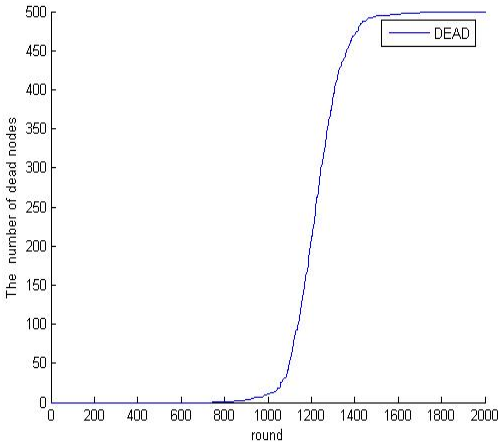

Figure 11. The dead nodes using leach per round with 500 nodes

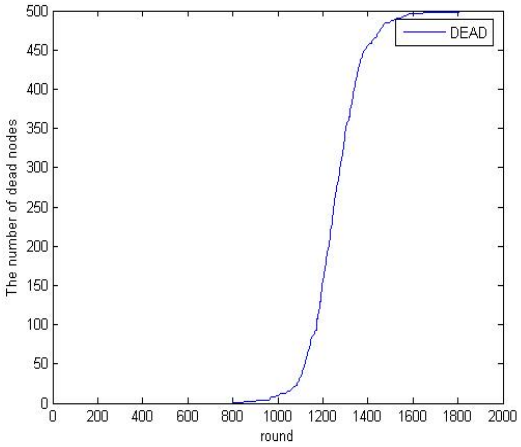

Figure 12. The dead nodes using rough sets per round with 500 nodes

Now contrast the next four figures. Figure 13 is the cluster head number per round in the running of the LEACH agreement in the case of 100 nodes. Figure 14 is the cluster head number per round in running process of rough set algorithm in the case of 100 nodes. Figure 15 is the cluster head number per round in the running of the LEACH agreement in the case of 500 nodes. Figure 16 is the cluster head number per round in running process of rough set algorithm in the case of 500 nodes. From the two contracts, it can be seen that the cluster head nodes number in each round is more average using rough set algorithm in the process of the network operation.

It can be seen from these Figure, the election of cluster head using traditional LEACH algorithm is a randomly generated process which the center of ideal numbers. The generated cluster head number is unstable and the magnitude of changes in the number of cluster head is large. Then if the number of cluster head nodes generated currently is too small, the cluster head node will consume too much energy in the current rounds due to the large loading. Thus, nodes are dead early. While if the current number of cluster head node is too much and then the number of clusters is also too much, making the overall energy of the cluster head consumed too large, and it is not conducive to saving energy. The improved methods deal with

the problem perfectly. As showed in figures, it produces solid number of cluster heads in each period of time. They are the optimal number of cluster head in the ideal condition. The moving cluster head ensures that the optimal number of cluster heads when part of the node has been dead in the same current situation. As a result, it reduces unnecessary energy consumption; on the other hand, it ensures the reasonableness of the number of clusters per round.

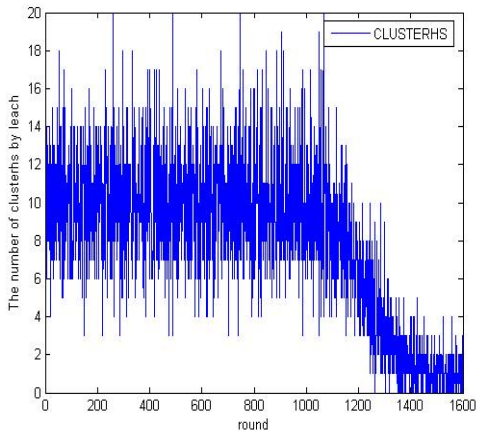

Figure 13. The numbers of clusters using leach with 100 nodes

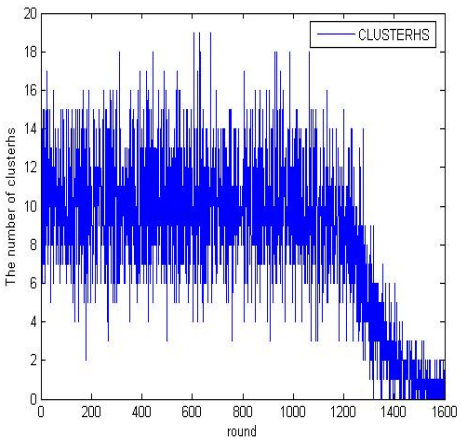

Figure 14. The numbers of clusters using rough sets with 100 nodes

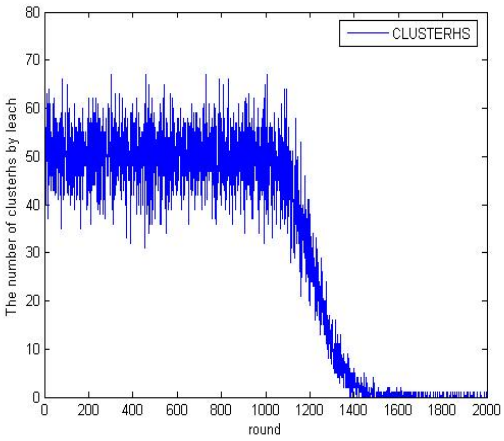

Figure 15. The numbers of clusters using leach with 500 nodes

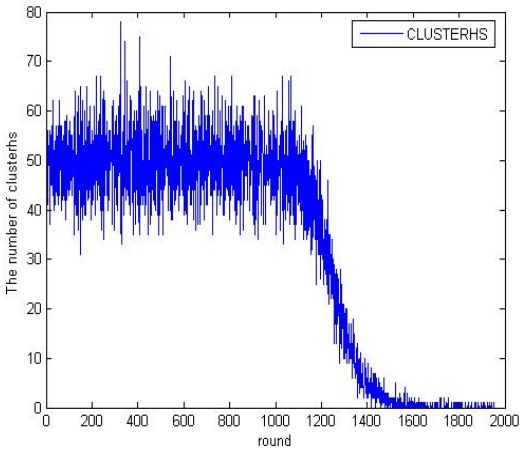

Figure 16. The numbers of clusters using rough sets with 500 nodes

7. Conclusion

WSN has been applied widely in various fields such as military defense, environmental monitoring, biological and medical, emergency rescue and disaster relief, and commercial application, etc., which also possesses a great strategic meaning in overall improving the modernize level of precision agriculture. Routing protocol is one of the key technologies of WSN, which has become one of the hot issues in present research of WSN. Because of its own characteristics, WSN makes its routing protocol research harder than that of previous network. Based on the study of cluster strategy in sensor network, the paper uses the method of the rough set to choice the cluster head in WSN. Rough set theory is used for the data fusion mechanism of WSN, the calculation is small but it can make a quick correct judgment for

making a choice of effective information, and it can effectively balance the energy of the nodes in the cluster, prolong the network survival cycle.

Acknowledgment

The authors are grateful to the lab members for providing helpful comments and information.

References

[1] Limin Sun. Wireless sensor network. Beijing: Tsinghua University Publishing House. 2005.
[2] DING Yue, DING Yong. An Energy-Efficient Clustering Routing Algorithm with Improved Quality of Cluster in Wireless Sensor Networks. *Chinese Journal of Sensor and Actuators*, China. 2012; 2: 258.
[3] HUA Ertian, LIU Kehong. Rough-set-based infant trolley design rules extraction and application. *Computer Engineering and Applications*, Beijing. 2013; 5: 117.
[4] Lu Chunfeng, REN Quanhui. Energy Balanced Clustering Algorithm Applications in Sensor Networks. *Science Technology and Engineering*, Beijing. 2013; 2: 351-352.
[5] NIU Yu-qing, SHI Wei. The design and implementation of wireless granary monitoring system based on 433MHz band. *Science paper Online*. 2011.
[6] Huiling Zhou, Chi Zhang. Improvement of leach protocol based on uneven clustering algorithm. *Computer Society*, 2012. DOI:10.1109/ICNIDC.2012.6418704.
[7] Dai Zhifeng. Research on intelligent information processing in wireless sensor networks using rough set theory. *Application Research of Computers*. 2007; 10: 75-78.
[8] Luo Jun, Xue Shengjun. New clustering algorithm based on rough sets for wireless sensor networks. *Application Research of Computers*. 2009; 1: 121-123.
[9] LIU Bin, LIANG Fengmei. Design of A MAC Protocol in Wireless Sensor Network Based on Cross Layer Optimization. *Mechanical Engineering and Automation*. 2011; 5: 9-11.
[10] Wang biao, Duan chanlun. The studies and applications of rough set and vague set. Beijing: *Publishing House of Electronics Industry*. 2008.
[11] Sadra Abedinzadeh, Samira Sadaoui. A Rough Set Approach to Agent Trust Management. *Computer society*, 2010.DOI:10.1109/SocialCom. 2010; 158: 1065.
[12] Guo qiang. *Rough set theory simple algorithm*. Wuhan: Wuhan university of technology. 2011.
[13] Wang Jue. *Rough set theory and its application*, Xian: Xi'an Electronic Science and Technology University. 2007.
[14] Huang huaji. *NS network simulation and protocol simulation*. Beijing: Posts and Telecommunications Press. 2010.
[15] Junguo ZHANG, Wenbin LI, Dongxu CUI, Xueliang ZHAO, Zhongxing YIN. The NS2-based Simulation and Research on Wireless Sensor Network Route Protocol. *Computer society*. 2009.