Research Article

Method for Effectively Utilizing Node Energy of WSN for Coal Mine Robot

Xiliang Ma [iD][1,2] **and Ruiqing Mao**[1]

[1]*Xuzhou Institute of Technology, Xuzhou, Jiangsu 221111, China*
[2]*School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China*

Correspondence should be addressed to Xiliang Ma; xlma818@sina.com

As a special application scenario, the data collected by wireless sensor networks of coal mine robot is from vital and dangerous environment. Therefore, the nodes need to work as long as possible. In order to efficiently utilize the node energy of wireless sensor network, this paper proposes a self-organizing routing method for wireless sensor networks based on Q-learning. The method takes many factors into account, such as the hop number, distance, residual energy, and node communication loss and energy. Each node of the wireless sensor networks is mapped into an Agent. Periodic training is carried out to optimize the route choice. Each Agent chooses the optimal path for data transmission according to the calculated Q evaluation value. Simulation results show that the self-organizing sensor networks using Q-learning can balance the energy consumption of the nodes and prolong the lifetime of the networks.

1. Introduction

Coal mine environment is often threatened by toxic gases and high temperature, so coal mine robots [1] often replace human to enter the pit to carry on the detection or the rescue tasks. Coal mine robots need to sense the states of themselves, as well as the environment of the underground. So a variety of sensors need to be deployed outside coal mine robots, such as the sensors to detect temperature, humidity, wind speed, wind direction, wind pressure, light intensity, coal dust and toxic gases, distance, the speed of the robot's left and right tracks, and pitch angle. There are several disadvantages with wired sensors. First, cables for power and signal connections are required, and wiring is a piece of tedious work. Second, temporary data collection cannot be performed and the system expansion is inconvenient., Third, the robot shells need to provide threading holes and location holes, which will reduce the stiffness and strength of the robot shells and reduce the explosion-proof performance of the coal mine robots. Vibration or collision during robots' motion affects the connection and joint quality of the cable. This will affect the reliability of the data detection and even cause coal mine robots

not to work properly in the underground. Wireless sensor network technologies can make up for these deficiencies.

Different from the general wireless sensor networks (WSNs) [2], a wireless sensor network deployed on a coal mine robot consists of a number of small sensor nodes. The distance between the nodes is relatively small. As coordinators and data collectors, these nodes transmit the data to the sink nodes and monitor the environment and the state parameters of the robot in real time. Each node has different data types and must meet the requirements of explosion-proof or intrinsic safety. So the nodes are usually powered by batteries. The data is collected by the nodes if from vital and dangerous environments. Therefore, the nodes need to work as long as possible. As a special application scenario, how to efficiently utilize the node energy of coal mine robot WSN is a difficult problem. Node energy consumption is mostly generated by sensor communications. In order to prolong the lifetime of the network, this paper designs a routing mechanism based on reinforcement learning and applies it to wireless body area networks. This paper also studies the energy efficiency, which further improves the energy efficiency and prolongs the network life cycle.
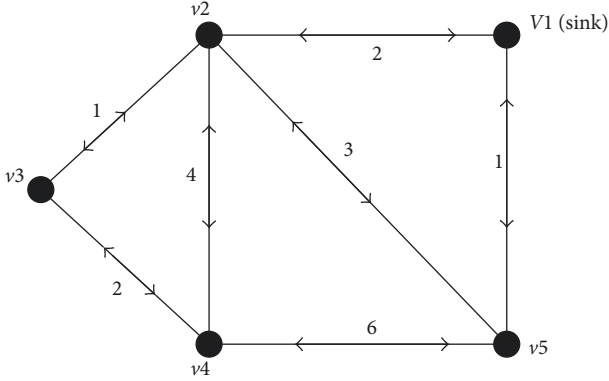
FIGURE 1: Model of wireless sensor network for coal mine robots.

2. Wireless Sensor Networks for Coal Mine Robot

2.1. A Wireless Sensor Network Model for Coal Mine Robot. The coal mine robot wireless sensor network has a single layer structure, which can be represented by a directed graph $G(V, A)$. The graph is shown in Figure 1, where V represents the sensor node set, V_1 is a sink node, and the wireless sensor network is wired to the coal mine robot control system by the sink node. Each node is fixed relative to the coal mine robot. The nodes transmit messages to the sink node in a direct or multihop manner. A represents a set of directed arcs, each of which represents the connection of two nodes in V that can communicate directly. The label beside an edge in the graph represents the corresponding path weight w_{ij}. The path with a smaller weight means that the channel has a better quality. The nodes on this path have more residual energy and the hop consumes less energy.

2.2. Node Energy Model. The energy consumption of wireless sensor network nodes is divided into 3 parts: perception, communication, and data processing. In this paper, we mainly consider the energy consumption of data communication, that is, the energy consumed by sending and receiving data. This paper refers to the energy model in LEACH [3], as shown in Figure 2. The energy consumption of the sending node includes two parts: the processing circuit and the power amplifier. The energy consumption of the receiving node is only used to process the circuit [4]. Suppose the distance between the receiving point and the sending point is d and the threshold of distance is d_0. When $d < d_0$, the free-space path loss model is adopted. If $d \geq d_0$, multipath fading model [5] is adopted.

In Figure 2, it can be seen that the energy consumption of nodes comes mainly from the process of transmitting and receiving data, which is closely related to the transmission distance. In Figure 2, the meaning of each symbol is as follows: E_{elec} is node energy consumption for sending and receiving 1-bit data. ε_{amp} is the power coefficient of the received data; n is the path decay index, whose typical value is 2 or 4. d is communication distance between two nodes;

k is the number of bytes sent; E_{PA} is the amount of energy consumed by processing received data.

In short distance free-space communication, the transmission power is directly proportional to the transmission distance. And the transmission power can be simply expressed as $E = \lambda d$, in which λ is a fixed parameter. The energy consumption will be four times as much as the original energy when the distance is doubled. Therefore, when the transmission distance is large, the energy consumption will be greatly reduced by adding forwarding nodes. As shown in Figure 2, a power amplifier involves the path decay index n. When the communication distance between nodes $d < d_0$, the path attenuation index n is set to 2 and the energy consumed between two nodes is proportional to the square of their distance. When $d \geq d_0$, n is set to 4. The communication distance between nodes is large. The energy consumed by sending k-bit data to a node at intervals of N is shown in

$$E_{\text{Tx}}(d, k) = \begin{cases} kE_{\text{elec}} + k\varepsilon_{\text{amp}}d^2, & d < d_0 \\ kE_{\text{elec}} + k\varepsilon_{\text{amp}}d^4, & d \geq d_0. \end{cases} \quad (1)$$

The energy consumed by receiving k-bit data is shown in

$$E_{\text{Rx}}(k) = kE_{\text{elec}} + kE_{\text{PA}}. \quad (2)$$

In WSN, the energy consumption of nodes is very sensitive to the number of hops [6, 7]. Therefore, we define the energy function of the number of hops as follows: $E_h = e^{\text{hop}(i,s)/H}$, where $\text{hop}(i, s)$ represents the hop count of the next hop from Agent j to sink. The hop count from Agent i to sink is H. Assuming that Agent i selects Agent j as the next hop, the energy consumption for transmitting k-bit packets between Agent i and Agent j is

$$E_{i,j} = \left(2kE_{\text{elec}} + k\varepsilon_{\text{amp}}d^\theta\right)e^{\text{hop}(i,s)/H} \quad \theta \in [2, 4]. \quad (3)$$

3. Energy Saving Methods

In order to prolong the lifetime of wireless sensor networks for coal mine robots, it is necessary to balance the energy consumption of all nodes. At the same time, the nodes with less remaining energy should be used as few as possible. In this paper, the Q-learning algorithm of reinforcement learning is applied to the wireless sensor networks of coal mine robots. This algorithm takes into account the distance between nodes, hops, communication energy consumption, and the residual energy consumption of nodes.

3.1. Agent Reinforcement Learning. Broadly speaking, an intelligent Agent senses the environment and performs an action, reinforcement learning, which is derived from animal learning and stochastic approximation [8]. It is an unsupervised machine learning technique. It can use the uncertain environment reward to find optimal behavior sequence and realize on-line learning in dynamic environment. Agents are independent, cooperative, and self-learning. They do not require human control. Agent nodes perceive the change of the environment and get the maximum reward value and get
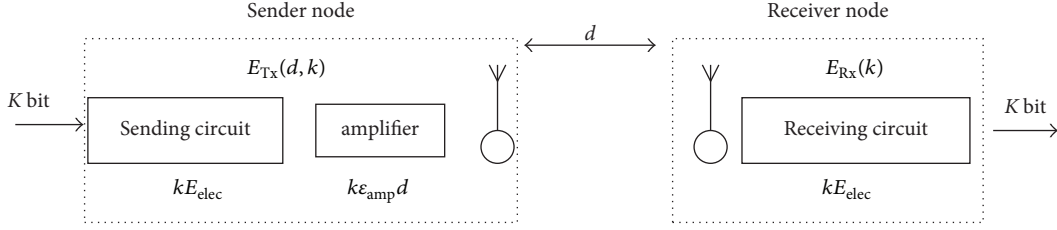
FIGURE 2: Energy consumption model of wireless sensor nodes in coal mine robot.
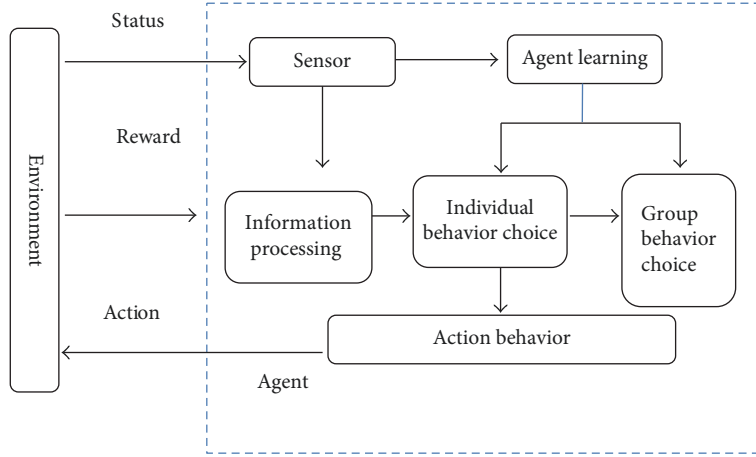

FIGURE 3: Basic principles of Agent's reinforcement learning.

an optimal strategy for corresponding decision actions. The principle of Agent's enhanced learning is shown in Figure 3. As can be seen from the diagram, Agent needs to interact with the environment. First, Agent senses information from complex environments. Then, Agent processes information, improves the performance, chooses a behavior, and makes group behavior choices. According to its individual and group behavior choices, Agent makes a decision, selects an action, and influences the environment.

3.2. Q-Learning Algorithm. Q-learning [9] is the most widely used algorithm in reinforcement learning algorithms. It is an unsupervised learning method whose input is a feedback from constantly changing and complex environment. Q-learning can be modeled by Markov decision process (MDP) [10]. MDP can be defined as a four-tuple (S, A, P, R), where S represents the set of states, A represents a set of actions, P represents the state transition function of the environment, and R is the reward function of the environment.

In MDP, the state transition function P and the reward function R are only related to the current state and actions and not related to the previous states and actions. The purpose of Agent's reinforcement learning is to learn a strategy [11] $\pi : S \rightarrow A$. Agent takes an action a_t according to the current status S_t; that is, $\pi(S_t) = a_t$. By following an arbitrary policy,

the cumulative value $V^\pi(S_t)$ obtained from any initial state S_t is

$$V^\pi (S_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots = r_t + \gamma V^\pi (S_{t+1})$$
$$= \sum_{i=0}^{\infty} \gamma^i r_{t+i}. \tag{4}$$

In formula (4), r represents the return value, and γ represents the discount factor which reflects the relative proportion of the delay return to the immediate return. The goal of Agent is to learn a strategy to make $V^\pi(s_t)$ maximum; this strategy is called an optimal policy and is represented by π^*:

$$\pi^* = \arg\max_\pi V^\pi (S_t), \quad \forall s. \tag{5}$$

Q-learning is a model independent reinforcement learning algorithm. It learns the optimal strategy directly. Q-learning neither needs a prior knowledge model about the state transfer function P and the return function R nor needs to learn these models in the course of learning. According to the characteristics of the coal mine robots' wireless sensor networks, a model-free Q-learning algorithm is proposed in this paper. The algorithm is simple, fast, and easy to use.

The single step Q value updating formula of the Q-learning algorithm is

$$
\begin{aligned}
Q_{t+1}\left(s_t, \alpha_t\right) \\
= (1 - \alpha) Q_t\left(s_t, \alpha_t\right) \\
+ \alpha\left[r_{t+1} + \gamma \max Q_t\left(s_{t+1}, \alpha'\right) - Q_t\left(s, \alpha_t\right)\right].
\end{aligned}
\tag{6}
$$

The contents of the Q table are constantly updated through formula (6) in the learning process [12]. In the formula, $Q_t(s_t, \alpha_t)$ is the function of status-action pair (s_t, α_t) at moment t; α and γ are learning rates and discount factors, respectively. s_t is the state of the environment at moment t. r_{t+1} is the return value of the state given by the environment at moment $t + 1$, $\max Q_t(s_{t+1}, \alpha')$ is the maximum Q value of the environment state s_{t+1} at the moment $t + 1$, and α' is any action taken by Agent in the environment state. The Q-learning algorithm converges to an optimal solution by iteratively searching the state space.

3.3. QLSORP Algorithm. In this paper, a routing algorithm for wireless sensor networks based on Q-learning is designed. The algorithm is applied to the routing protocol of wireless sensor networks for coal mine robots. The routing protocol is called QLSORP (Q-Learning-based Self-Organization Routing Protocol). Q-learning is applied to the routing algorithms in wireless sensor networks.

The process of finding the optimal path in wireless sensor networks is equivalent to a Markov process. In each step, an instantaneous reward value is generated. The prerequisite for implementing the routing algorithm is that the return value R, the status s, and the action a must be determined. The return value is the ratio of the residual energy of Agent nodes to the energy consumed by communication. With R_j representing the current remaining energy of the node, according to formula (3), the return value of the Agent node j can be calculated as follows:

$$
V^\pi(j) = \frac{R^j}{E_{i,j}} = \frac{R^j}{\left(2kE_{\text{elec}} + k\varepsilon_{\text{amp}}d^\theta\right) e^{\text{hop}(i,s)/H}}
\tag{7}
$$

$$
\theta \in [2, 4].
$$

In the formula, the return value takes into account the remaining energy in the network and the energy consumption required for node communication. Then, a key point is found to balance the energy loss of nodes. State changes with time. Action refers to the process in which Agent nodes choose a path as the optimal one. Considering the communication energy consumption and the remaining energy, the Agent nodes on the routing path continuously send learning packets to the neighbor nodes to obtain the desired return value. The nodes will select the path with the highest Q value as the optimal one. Thus, the whole energy of the network is balanced and the network life cycle is prolonged under the premise of guaranteeing the efficiency of data transmission. The algorithm is described as follows.

Step 1. The sink node sends a learning evaluation message to its neighboring nodes in the same cycle. The wireless sensor network is initialized, and all sensor nodes are started. Each sensor node records the number of its neighboring nodes and energy consumption to its neighboring nodes. At the same time, the energy consumption threshold of each sensor node is set and the return value of the node is set to 0.

Step 2. Define a set D to store the information of the Agent nodes that have been learned during this cycle to prevent infinite loops in the path creation process. The source node periodically sends learning information to its neighboring nodes and determines whether the neighboring nodes exist or not in the set D; at the same time, the return value to each neighboring node is calculated. The neighboring node with a high return value is selected as the next hop routing node.

Step 3. Repeat Step 2. The neighboring node of the selected node is calculated and then probes the route of the next hop. The sink nodes are found successively.

Step 4. When the return value of other nodes is received, the sink node updates the Q value table by formula (6) according to the Q-learning mechanism in the reinforcement learning algorithm. The path is selected according to the Q value in the table. When the energy of the sensor nodes in the selected path is lower than the set threshold or the sensor node no longer has any effect, a message is sent to the source node along the opposite direction of the selected path. Then the source sensor node gives up choosing this path. Instead, the path with the second largest Q value will be chosen to transmit the information.

Step 5. The source node selects the path with large Q value to transmit the information stably to the sink node. The residual energy information of each sensor node is updated simultaneously.

Step 6. The sink node periodically sends learning message. According to the message, the source node probes the path, selects the path, and sends the message to the sink node. The value of Q varies with the return value of the node. The Q table is updated and stored in the sink node.

Figure 4 is the flow chart of the QLSORP algorithm.

4. Simulation Experiment Analysis

4.1. Experiment General Situation. Network simulation is a basic method for wireless sensor network research. The wireless sensor network of coal mine robot mainly involves temperature, humidity, wind speed, wind direction, light intensity, coal dust, oxygen, gas, distance sensor, track speed, and pitching angle of robot. Communication between wireless sensor nodes and between nodes and sink nodes constitutes a wireless sensor network. The topology is shown in Figure 5. Simulation analysis is carried out to verify the efficiency of routing algorithm based on Q-learning for wireless sensor networks. It is assumed that the transmission
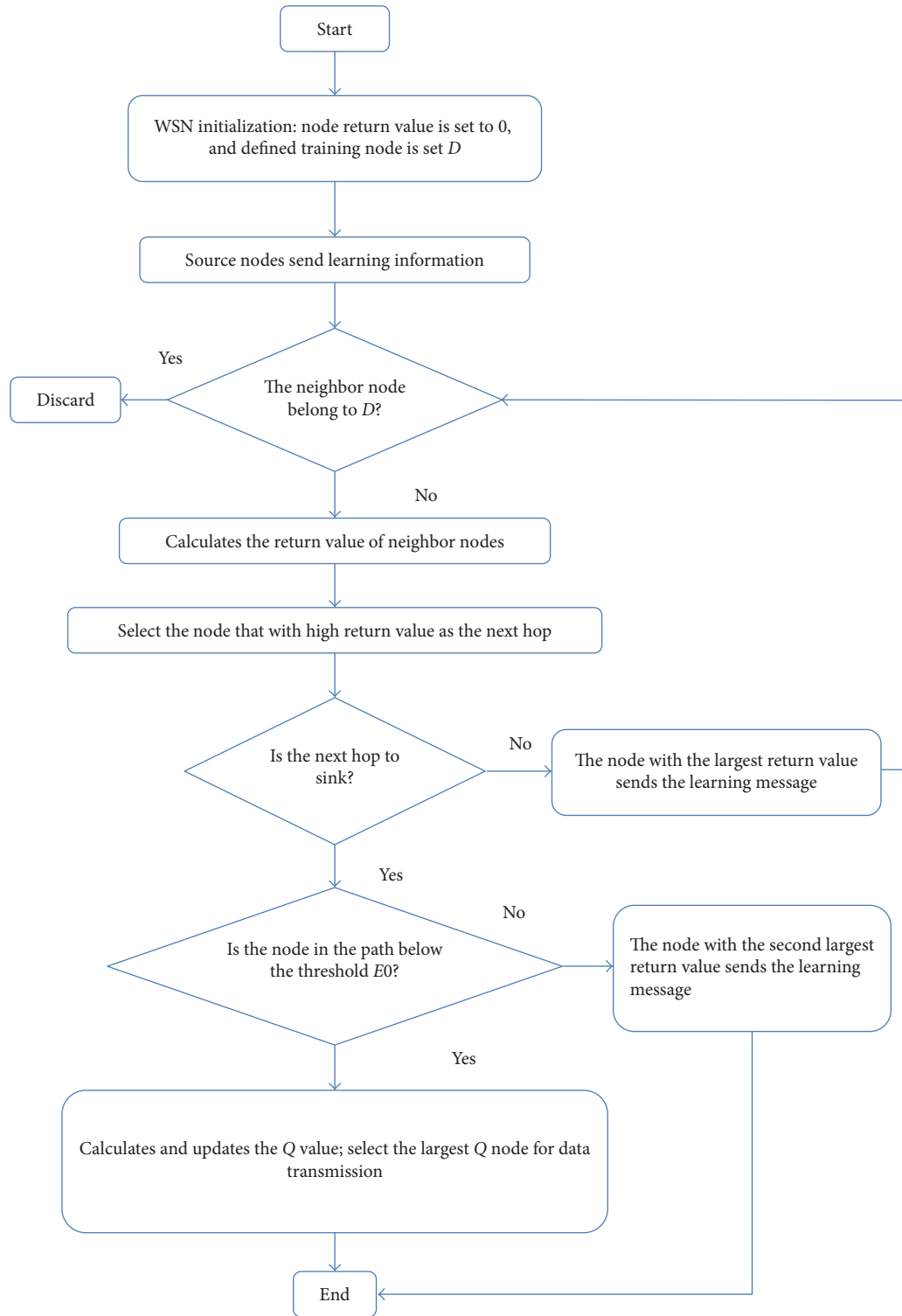
FIGURE 4: Flow chart of the algorithm.

radius of the sensor node is 20–60 cm. The transmitted packet size is 10 K bits. The initial energy of the nodes is 5 J. The nodes send data to the sink node in a single hop or multihop. The sink node is wired to the control system of the coal mine robot, whose energy is always constant. The movement of the nodes is not considered because the wireless sensor network is fixed after the arrangement of the sensor and the robot body. The QLSORP algorithm was developed and implemented in the environment of MATLAB 2016; other simulation parameters are set as follows: the size of the
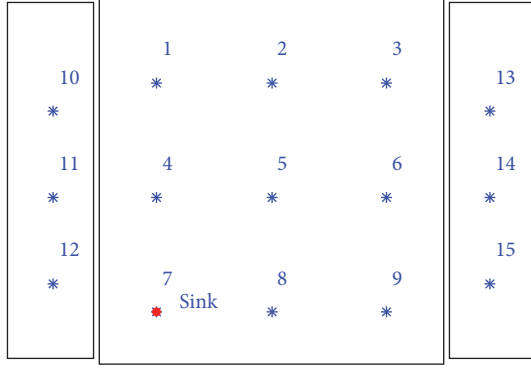
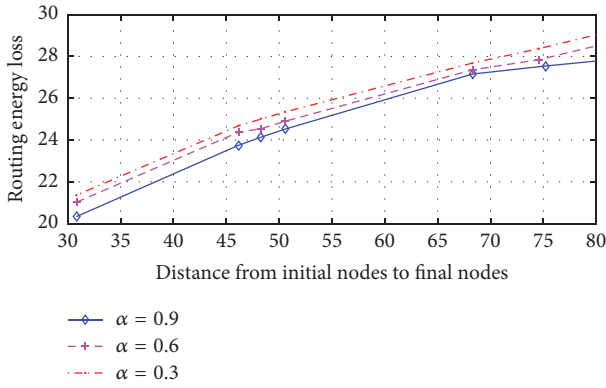FIGURE 5: Node topology of wireless sensor network for coal mine robot.

FIGURE 6: Routing energy loss.

FIGURE 7: Residual energy of node.

FIGURE 8: The impact of alpha values on Q.

network area is $120 \, \text{cm} * 200 \, \text{cm}$, E_{elec} is $50 \, \text{nJ/bit}$, E_{PA} is $100 \, \text{pJ/bit/cm2}$, and $d0$ is $100 \, \text{cm}$.

4.2. Experimental Results and Analysis. From formula (6), it can be seen that in the Q-learning based self-organizing routing method for wireless sensor networks (QLSOP), the two parameters γ and α remain unknown and need to be preset. In order to reduce the computational complexity of the routing algorithm, $\gamma = 1$ is used to reduce the delay. α is desirable for any number of 0-1, such as 0.9, 0.6, and 0.3. The simulations are conducted to verify the influence of parameter α on the lifetime of wireless sensor networks, the choice of node paths, the routing energy loss, and the residual energy of nodes.

Figure 6 shows the routing loss of the source nodes at different distances to sink nodes. It is obvious that the energy loss of the nodes routing to the sink nodes is proportional to the distance between them. The energy consumed by routing increases as the distance between the nodes and the sink nodes increases. α is the learning rate. Its size has a certain influence on the loss of routing energy. The smaller the learning rate is, the more the loss of routing is.

The number of nodes also affects the routing energy consumption of the source node routing to the sink node. The residual energy of the node is obtained by simulation, as shown in Figure 7. When the number of nodes in wireless
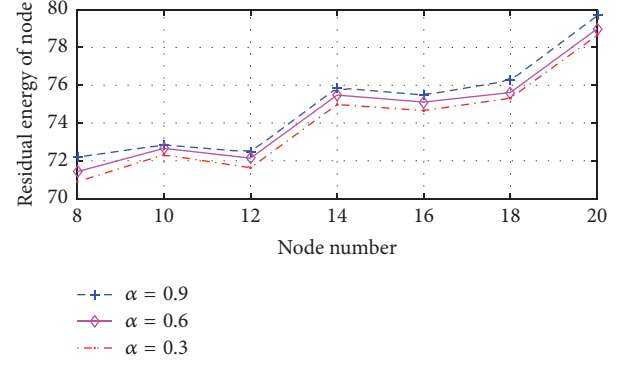
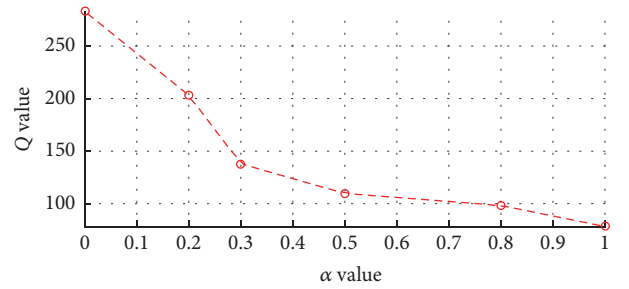sensor networks increases, the residual energy in the nodes becomes larger. The higher the learning factor, the more the residual energy the nodes have.

α is the learning rate, its value affects convergence of Q-learning. The effects of different α on Q value are shown in Figure 8. With the increasing of α, the value of Q in the fixed path decreases gradually. Formula (6) shows that when α value is close to 0, the value of Q depends on the preserved Agent node's Q value. When α value is close to 1, Q value depends on the maximum Q value of the Agent's neighboring node and the return value from the neighboring node.

The lifetime of the wireless sensor network for the coal mine robot is shown in Figure 9. Figure 9(a) is a 3D surface for node number, communication path distance, and network lifetime in wireless sensor networks. The shorter the communication path, the less the energy consumed by the nodes and the longer the network lifetime. In addition, with the increasing number of nodes in wireless sensor networks, the number of alternative routes through which the source node sends data to the sink node increases. By continuously updating the return value, the optimal path is selected and the consumption of the routing energy is reduced, which results in a longer network lifetime. Figure 9(b) shows the comparison between the DSR algorithm and the QLSORP algorithm proposed in this paper. The curve at the bottom of the graph represents the network lifetime of the DSR under the same simulation conditions. Its performance is lower than that of QLSORP. The reason is that the QLSORP algorithm balances the energy between the nodes of wireless sensor networks. From Figures 9(a) and 9(b) we can see that when
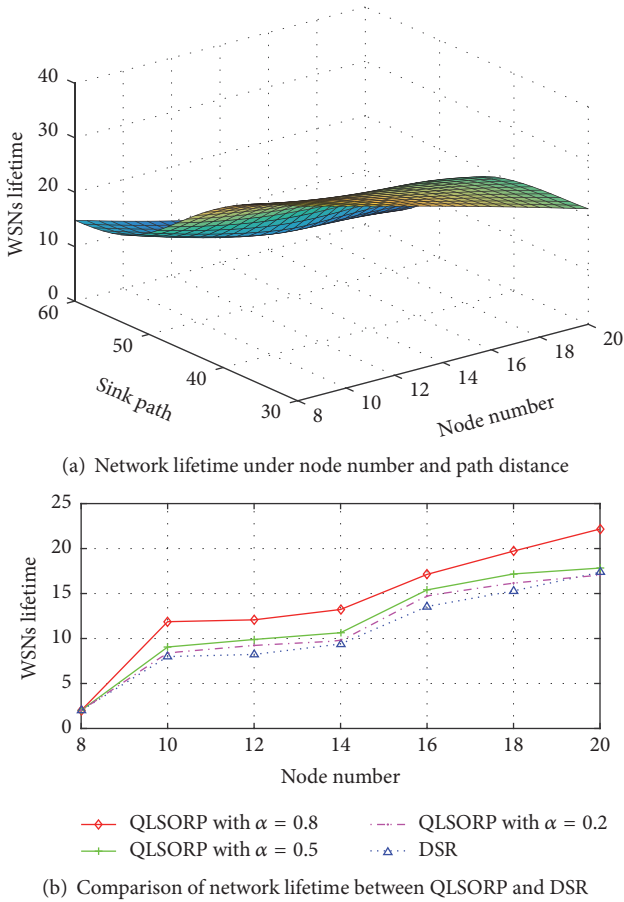
(a) Network lifetime under node number and path distance

(b) Comparison of network lifetime between QLSORP and DSR

Figure 9: Network lifetime.

the number of nodes is near 18, with the increase of α, the lifetime of WSNs also increases accordingly. It is noteworthy that the optimal path is invariable, regardless of the value of α.

5. Conclusions

This paper proposes a QLSORP routing algorithm for wireless sensor network in coal mine robot based on Q-learning. The wireless sensor network nodes are regarded as intelligent Agent nodes. A self-organizing routing method based on Q-learning for wireless sensor networks is designed. It takes into account the hop number, distance, residual energy, communication loss, energy, and so on. By calculating the return value of neighboring nodes, Q value of the path is updated constantly. The updated Q value table is kept in the sink node. The sink node searches the optimal path by comparing Q values and selects the path with the largest Q value for data transmission. When the node in the path is accidentally dead or the node energy is exhausted, the path with the second largest Q value is selected for data transmission. This balances the energy between nodes. Compared with the DSR algorithm, the QLSORP algorithm can significantly reduce the energy consumption of the network. The algorithm can prevent the premature death of some nodes in the network.

The coal mine robot wireless sensor network node energy can be utilized effectively. The lifetime of wireless sensor networks is prolonged.

This paper mainly studies the effectiveness of the routing algorithm based on Q-learning to balance the node energy of the WSN and improve the lifetime of the whole network. The robustness of WSN has not been considered, such as the impact of environmental noise interference on WSN, or the coverage hole problem of WSN [13] caused by the failure of some nodes or the exhaustion of energy. These problems all affect the practical effect of WSN. The next work will further study the topology self-cure algorithm [14] for node failure or energy exhaustion and improve the robustness to environmental noise [15] and other disturbances.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] X. Ma and H. Zhu, "Gas concentration prediction based on the measured data of a coal mine rescue robot," *Journal of Robotics*, vol. 2016, pp. 1–10, 2016.

[2] A. E. Forster and A. L. Murphy, "Exploiting reinforcement learning for multiple sink routing in WSN," in *Proceedings of the IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems, MASS*, Italy, October 2007.

[3] P. Ciciriello, L. Mottola, and G. P. Picco, "Efficient routing from multiple sources to multiple sinks in wireless sensor networks," *Wireless Sensor Networks*, vol. 4373, pp. 34–50, 2007.

[4] M. Meng, X. Wu, B.-S. Jeong, S. Lee, and Y.-K. Lee, "Energy efficient routing in multiple sink sensor networks," in *Proceedings of the 5th International Conference on Computational Science and Its Applications (ICCSA '07)*, pp. 561–566, IEEE, August 2007.

[5] J.-M. Dricot, S. Van Roy, G. Ferrari, F. Horlin, and P. De Doncker, "Impact of the environment and the topology on the performance of hierarchical body area networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, pp. 1–17, 2011.

[6] X. Yan, Y. Zhang, H. Tang, and S. Li, "An ETBG optimization algorithm based on analytic hierarchy process in WSS," in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, China, March 2013.

[7] R. R. Yager, "Weighted maximum entropy OWA aggregation with applications to decision making under risk," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 39, no. 3, pp. 183–189, 2009.

[8] J. K. Goyal and K. S. Nagla, "A new approach of path planning for mobile robots," in *Proceedings of the 3rd International Conference on Advances in Computing, Communications and Informatics (ICACCI '14)*, pp. 863–867, IEEE, New Delhi, India, September 2014.

[9] X. Shi, Q. Ran, M. Fan, H. Yu, and L. Wang, "Dynamic weighted DV-Distance algorithm for wireless sensor networks," *Chinese Journal of Scientific Instrument*, vol. 34, no. 9, pp. 1975–1981, 2013.

[10] S. L. Zheng, H. Che, Y. Q. Fan, D. Hu, and X. Xiao, "Agent-based mobile sink routing algorithm in wireless sensor networks,"

Journal of Electronic Measurement and Instrument, vol. 27, no. 2, pp. 127–134, 2013.

[11] H. Zhao, Z. H. Hu, and Y. Wwn, "ACS based differentiated service routing algorithm in wireless sensor network," *Journal on Communications*, vol. 34, no. 10, pp. 106–115, 2013.

[12] J. Elias, "Optimal design of energy-efficient and cost-effective wireless body area networks," *Ad Hoc Networks*, vol. 13, pp. 560–574, 2014.

[13] L. H. Yan and Y. Y. HE, "Robust approach foe holes recovery of wireless sensor networks," *Computer Science*, vol. 44, no. 2, pp. 123–128, 2017.

[14] L. F. Liu, J. G. Wu, Z. Q. Zou, H. Chen, and L. Niu, "opology self-cure algorithm aiming at node failure problem in wireless sensor networks," *Journal of Southeast University*, vol. 39, no. 4, pp. 695–699, 2009.

[15] A. Fortuna, L. Frasca, M. Iachello, M., and V. T. Pham, "Robustness to noise in synchronization of network motifs: Experimental results," *An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 4, Article ID 043106, 2012.