Scientific Research Publishing

Journal of Computer and Communications, 2023, 11, 165-180
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

A Data Transmission Path Optimization Protocol for Heterogeneous Wireless Sensor Networks Based on Deep Reinforcement Learning

Yu Song[1,2,3], Zhigui Liu[1*], Xiaoli He[4]

[1]School of Information Engineering, South West University of Science and Technology, Mianyang, China
[2]Department of Network Information Management Center, Sichuan University of Science and Engineering, Zigong, China
[3]Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Zigong, China
[4]School of Computer Science, Sichuan University of Science and Engineering, Zigong, China
Email: *zhiguiliu_swust@hotmail.com

How to cite this paper: Song, Y., Liu, Z.G. and He, X.L. (2023) A Data Transmission Path Optimization Protocol for Heterogeneous Wireless Sensor Networks Based on Deep Reinforcement Learning. *Journal of Computer and Communications*, **11**, 165-180.
https://doi.org/10.4236/jcc.2023.118012

Received: July 22, 2023
Accepted: August 28, 2023
Published: August 31, 2023

Abstract

Wireless sensor networks had become a hot research topic in Information science because of their ability to collect and process target information periodically in a harsh or remote environment. However, wireless sensor networks were inherently limited in various software and hardware resources, especially the lack of energy resources, which is the biggest bottleneck restricting their further development. A large amount of research had been conducted to implement various optimization techniques for the problem of data transmission path selection in homogeneous wireless sensor networks. However, there is still great room for improvement in the optimization of data transmission path selection in heterogeneous wireless sensor networks (HWSNs). This paper proposes a data transmission path selection (HDQNs) protocol based on Deep reinforcement learning. In order to solve the energy consumption balance problem of heterogeneous nodes in the data transmission path selection process of HWSNs and shorten the communication distance from nodes to convergence, the protocol proposes a data collection algorithm based on Deep reinforcement learning DQN. The algorithm uses energy heterogeneous super nodes as AGent to take a series of actions against different states of HWSNs and obtain corresponding rewards to find the best data collection route. Simulation analysis shows that the HDQN protocol outperforms mainstream HWSN data transmission path selection protocols such as DEEC and SEP in key performance indicators such as overall energy efficiency, network lifetime, and system robustness.

Keywords

HWSNs, Clusting, Deep Reinforcement Learning, DQN

1. Introduction

Wireless Sensor Networks (WSN) are self-organizing distributed networks composed of a large number of microsensor nodes with sensing, computing, and communication capabilities. Nodes in the network collect and process relevant information about the surrounding environment through autonomous collaboration, and then transmit the information to the base station for further analysis and processing through short-distance multi-hop communication. Usually, the energy of wireless sensor nodes is only provided by batteries with limited capacity, and it is difficult to replenish the energy. Once the sensor nodes run out of energy, it will cause changes in the network topology, resulting in the inability of sensing data to reach external servers and the loss of application functionality. How to improve WSN energy efficiency and maximize network lifetime is the most important issue in the research of wireless sensor network protocols [1].

Heterogeneous Wireless Sensor Networks (HWSNs) are networks composed of various types of sensor nodes with different performance parameters. It is a special type of wireless sensor network composed of different types of nodes, which can have different processing capabilities, memory size, energy supply, sensor types, etc. The significance of heterogeneous wireless sensor networks lies in their ability to improve network performance without increasing costs. Designers can choose different types of nodes based on the implementation functions of the application.

In HWSNs, the heterogeneity of nodes can be set based on specific application requirements and node roles. For example, some nodes in the network may have stronger data processing and communication capabilities, designed as cluster heads or relay nodes responsible for data aggregation and communication with remote servers. Other nodes may be conventional sensor nodes, mainly responsible for data collection and transmission to the cluster head. The main advantage of HWSNs is that they can provide more flexible and efficient solutions to meet different application needs. For example, in situations where a large amount of data processing and transmission is required, they can more effectively utilize those high-capacity nodes to reduce network energy consumption and improve data transmission efficiency. However, the design and management of HWSNs are also more complex, requiring consideration of the differences in capabilities and collaborative work among different nodes.

1.1. Heterogeneous Types of HWSNs

1) Link heterogeneity: Link heterogeneity refers to the presence of broadband links or the ability to provide sufficient long-distance transmission and reception

in heterogeneous nodes of wireless sensors. Heterogeneous nodes in the link can ensure the reliability of data transmission and improve network security. The security issues of HWSNs are more complex than traditional WSNs, involving data confidentiality, integrity, and availability.

2) Heterogeneous computing and storage: Heterogeneous nodes selected for different scenarios have different configurations. Heterogeneous nodes have more powerful microprocessors and storage space, which can complete special tasks that meet the conditions and provide the network with more powerful and complex node data computing and storage capabilities.

3) Node energy heterogeneity: Refers to the process of parameter initialization of nodes, where each heterogeneous node carries different amounts of energy or the battery can be replaced. In the application of nodes, heterogeneous nodes are located in different environments and application scenarios, resulting in different energy consumption in the network. And energy heterogeneity is the foundation for building link heterogeneity and computing storage heterogeneity, and the lack of node energy heterogeneity will lead to the inability of the entire network to be heterogeneous and changes in its lifecycle. The research on heterogeneous node energy mainly includes the application and optimization of energy consumption models, energy optimization strategies, energy scheduling, and other fields [2].

The features possessed by HWSNs can be combined with many fields in research and application to improve the progress and development of each other, such as wireless communication, which is the basis for the study of heterogeneous wireless sensor networks; distributed computing, where the nodes of a heterogeneous wireless sensor network need to work together, and the network's operation mechanism can be designed and analyzed with the help of the concept of distributed computing. There are also the fields of data fusion and information processing, optimization theory, and network security. In addition to the cross-fertilization of the above research areas, data transmission protocols for HWSNs are also a focus of current research. Such as the development and research of LEACH protocol, SEP protocol and DEEC algorithm.

1.2. Challenges Faced by HWSNs

1) Node management and coordination: There are different types and capabilities of nodes in heterogeneous wireless sensor networks, which require effective node management and coordination. This includes issues such as task allocation, routing, data aggregation, and network optimization, requiring the design of appropriate algorithms and protocols to achieve effective network operations. When heterogeneous wireless sensor networks work, most nodes are randomly deployed in the region. Due to the different working states of each node, some nodes may exhaust their energy prematurely. At this point, nodes will stop working, leading to coverage vulnerabilities and affecting the normal operation of the network [3].

2) Clustered data collection: Clustering algorithms aim to solve the efficiency

problem of data transmission in HWSN, and the collection of clustered data is one of the important tasks in HWSN data processing and performance optimization. Due to the differences in initial energy, computational efficiency, storage capacity, and path selection ability of sensor nodes, the process of clustering data collection is full of challenges [4].

3) Data security: Nodes in heterogeneous wireless sensor networks may face security and privacy threats. Due to the different functions and capabilities of nodes in the network, some of them may be more susceptible to becoming attack targets. Protecting data security and privacy in the network and designing effective security mechanisms is an important challenge.

4) Network scale and deployment: Heterogeneous wireless sensor networks may need to cover a large area and contain a large number of sensor nodes. The complexity and cost of managing and maintaining large-scale networks are high. At the same time, the deployment method and location selection of nodes can also affect the performance and coverage of the network, requiring reasonable deployment planning [5].

2. Related Work

2.1. Research Status of Heterogeneous Wireless Sensor Networks at Home and Abroad

The research status of routing optimization in heterogeneous wireless networks mainly includes the use of routing discovery optimization, routing Selection algorithm, self-organizing networks, software-defined networks and machine learning. Using routing discovery optimization to improve routing discovery efficiency and reduce routing overhead, mainstream protocols include Ad hoc On-Demand Distance Vector Routing (AODV), Dynamic Source Routing (DSR), and Optimized Link State Routing (OLSR). The shortest path, load balancing and other routing Selection algorithms are used to consider factors such as network topology and node energy. Self-organizing networks achieve network configuration, optimization, and recovery through intelligent means. Related research includes wireless adaptive routing based on Q-learning. The software definition of OpenFlow enabled wireless routing enables network programmability by separating the data plane and control plane. Wireless routing optimization introduces machine learning technology. Common applications include intelligent routing based on Deep reinforcement learning and routing prediction based on graph convolutional networks. Mobile edge computing can effectively reduce wireless network traffic and routing overhead.

In 2015, Tanwar S *et al.* [6] first introduced the concepts of Low Energy Adaptive Clustering Hi array (LEACH) and Stable Election Protocol (SEP). The clustering hierarchy method was used for CH selection in homogeneous LEACH networks, but when the environment is heterogeneous, it can easily lead to a decrease in network performance. For simple heterogeneous networks, SEP consists of two types of nodes. The selection of CH is based on

weighted selection probability. The Weighted Election Probability (WEP) solves the problem of LEACH's inability to maintain performance in heterogeneous networks, but SEP cannot guarantee stability in multi-level heterogeneous networks.

In 2017, Chatap A *et al.* [7] first analyzed and elaborated on isomorphic clustering, explaining the differences and difficulties between isomorphic and heterogeneous networks. The initial energy and hardware complexity of all sensor nodes in isomorphic networks are the same, but heterogeneous nodes have two or more types of nodes, and the initial nodes are not the same. In the work of the network, the energy consumption of nodes determines the entire network lifecycle. Comparative analysis was conducted on different methods of extending network lifespan and stability using protocols such as Distributed Energy Efficient Clustering (DEEC), Developed Distributed Energy Efficient Clustering (DDEEC), and Enhanced Distributed Energy Efficient Clustering (EDEEC). It is concluded that energy efficiency is the main problem faced by heterogeneous wireless sensor networks, and the protocols studied and analyzed above have overcome the problem of energy consumption and are widely used in practical applications.

In 2022, Abdul-Qawy A. S. H. *et al.* [8] proposed a new classification called bridall together, which covers all energy-saving applications in traditional HWSNs and IoT-based wireless sensor technologies. In the application of sensors in the fields of environmental monitoring, precision agriculture, and security in the Internet of Things, their heterogeneous devices usually have limited electricity or energy. In order to meet user needs and established goals, extending the network life is crucial. Use energy sensing technology to balance the energy load between all nodes, coordinate their interactions, and add additional relay nodes to reduce the energy of data transmission.

2.2. Deep Reinforcement Learning

Deep reinforcement learning combines deep neural networks with Reinforcement learning, uses deep neural networks to approximate value function, and uses Reinforcement learning method to update [9]. For Reinforcement learning, the input of many practical application problems is high-dimensional. When we think of achieving a set goal, we directly use the original data as the state. The dimensions are high and the number of states is large, leading to difficulties in achieving the goal. Based on Reinforcement learning, Deep reinforcement learning abstracts features from high-dimensional data as states. The role of deep neural networks is like this. It can not only fit the idea of classification or regression function, but also fit the value function and strategy function. The decision-making process of deep reinforcement learning is shown in **Figure 1**.

In 2015, Yan Zhang, *et al.* [10] constructed a network with dynamic adjustment and adaptability to diverse business needs based on Markov chains, presenting a polymorphic layer of routing instance sets with multimodal characteristics

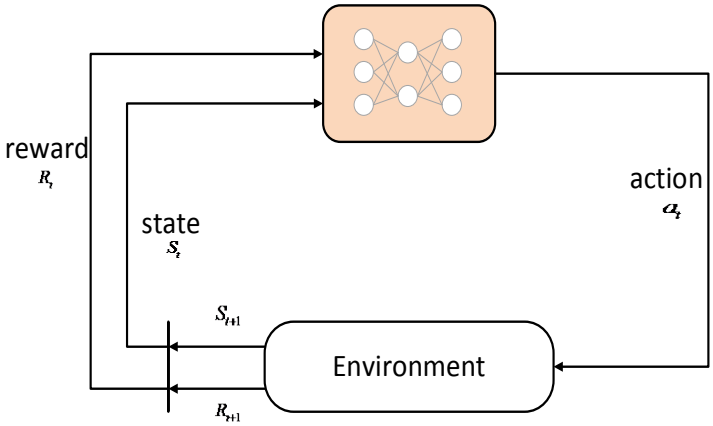

Figure 1. The decision-making process of deep reinforcement learning.

such as functionality-specific, security-specific, and service quality specific. In the process of polymorphic derivation, the selection of routing function properties is not only related to application requirements, but also related to the output of the last selected routing function. Therefore, a Decision model is constructed using Markov properties, which are jointly determined by the state space, action set, transition probability, reward function and criterion function Co-determination. This protocol greatly improves the routing transmission capability of the network and allows for flexible formulation of routing protocols based on different business needs.

In 2019, Sun P *et al.* [11] An intelligent network control architecture is designed, which is based on data platform, control plane and artificial intelligence plane, and can output appropriate strategies according to the change of network traffic distribution. And a Deep reinforcement learning method based on RNN (Recurrent Neural Network) is designed. Compared with traditional routing algorithms, this method can quickly process a large number of network status data. It proves that AI has great potential in Transportation engineering, but in different scenarios, the balance between performance and communication overhead and information collection still needs further research.

3. System Model

In this paper, we assume that the sensors are randomly deployed within a specific area, with the Base Station (BS) located at the center, as shown in **Figure 2**. In this network, the node transmit their respective data to the cluster, which then forwards it to the base station either directly or through a relay mechanism. Note that all sensor nodes are heterogeneous, consisting of normal nodes, advanced nodes, and supernodes, with their initial energy ranging from low to high. We set up each heterogeneous sensor node with GPS. When heterogeneous nodes are placed in the monitoring area, they will inform BS of their location, remaining energy, and heterogeneous energy type through wireless communication [12].

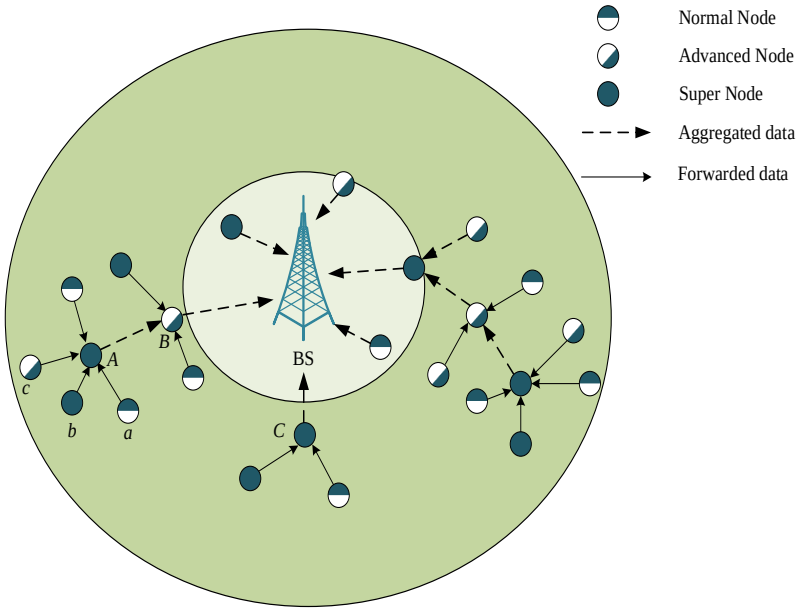

Figure 2. Heterogeneous wireless sensor layout area.

Given the continuous energy consumption of sensors during data transmission, it becomes vital to prioritize optimal energy utilization. Therefore, relay nodes are chosen based on their lower energy consumption and higher remaining energy. In this paper, we consider a typical first-order radio model to assess the energy consumption of a node transmitting b bits of data over a distance d. This can be expressed as

$$\begin{aligned} E_{tx}(l,d) &= E_{tx-elec}(l) + E_{tx-amp}(l,d) \\ &= \begin{cases} l \times E_{elec} + l \times \varepsilon_{fs} \times d^2, d < d_{th} \\ l \times E_{elec} + l \times \varepsilon_{mp} \times d^4, d \geq d_{th} \end{cases} \end{aligned} \quad (1)$$

where ε_{fs} and ε_{mp} represent the power amplification factors of the amplifier under the free space channel model and the multipath attenuation model, respectively; d is the physical distance between two nodes [13].

4. Proposed Approach

In this section, we first propose a DQN-based Deep reinforcement learning algorithm - HDQN to solve the routing problem of HWSNs. This algorithm uses important parameters such as distance between heterogeneous nodes, remaining energy, and number of relays to find the optimal routing problem for HWSNs. Specifically, we first take the coordinates and energy of each node and cluster head as the state. Then, based on the reward function, the DQN algorithm is used to learn and select the next hop route action until the information is transmitted to BS [14].

4.1. Design of the Reward in DQN Network

We assume that a super node acts as an Agent, with the goal of finding the op-

timal data route by taking a series of actions for different states and receiving corresponding rewards. The state space $\{S_t\}$, action spaces $\{A_t\}$, and reward function $\{R_t\}$ of the agent at time T are defined as follows:

1) State space $\{S_t\}$: coordinates of all heterogeneous nodes, the remaining energy of CH, Number of relays for CH.

2) Action space $\{A_t\}$: next hop routing.

3) Reward function $\{R_t\}$: reward function.

$$R_N = \alpha D(n) - \beta E(n) - \omega C(n) \tag{2}$$

where α, β, and ω represent the weight ratios of $D(n)$, $E(n)$, and $C(n)$ respectively, aiming to achieve energy balance and prolong network survival time, as described in this paper, with $\alpha + \beta + \omega = 1$. $D(n)$ represents the distance between nodes and the step distance when transmitting data. The purpose is to select the node closest to the current node and closest to the destination as the next hop node. $E(n)$ reflects the comprehensive ratio of the remaining energy of the next hop to the average energy of adjacent nodes and the initial energy of the node [15]. Nodes with higher energy are given preference. $C(n)$ represents the number of cluster heads (CHs) in a route. A higher value of $C(n)$ indicates a lower probability of selecting that route as the best route. This selection criterion ensures that the optimal route chosen at each instance involves the fewest number of hops to reach the base station, while also balancing the participation of CHs in the transmission process to minimize energy consumption. Assuming N is the number of all CHs, the number of CH participants in a route can be expressed as $\sum_{i=1}^{N} CH\_part_i$.

In the automatic networking stage, a node transmits a portion of its effective information to adjacent nodes through broadcast. This information is also included in the packet. Simultaneously, it incorporates the node's absolute geographical location to calculate the energy consumption involved in transmitting and receiving information, while also considering the distance factor. In this paper, the proposed combined reward for distance and stepping is denoted as:

$$D(n) = -\mu d_n + (1-\mu) r_d \tag{3}$$

where d_n represents the distance between two adjacent nodes M and N. $d_{MB} = \sqrt{x_{MB}^2 + y_{MB}^2}$ is the distance from node M to the base station. $d_{NB} = \sqrt{x_{NB}^2 + y_{NB}^2}$ is the distance from node N to the base station.

We assume that the initial energy of all sensor nodes is equal, so the proposed energy synthesis $E(n)$ can be expressed as:

$$E(n) = \tau\left(\frac{E_{re}}{E_{ie}} + \frac{E_{re}}{\overline{E_{ae}}}\right) + (1-\tau)(E_{ts} + E_{rs}) \tag{4}$$

where E_{re} and represent the remaining energy and initial of node n, respectively. $\overline{E_{ae}}$ donates the average energy of the node group composed of adjacent nodes n [16].

Data forwarding adopts multi-hop routing, ultimately forwarding data packets

to the BS. The focus is on updating the routing scheme and processing data retransmission through reinforcement learning to achieve the lowest overall energy consumption and energy balance in the network. **Figure 3** illustrates the DQN deep reinforcement learning algorithm for solving routing problems. First, the state s_t of the environment is sent to the evaluation network in the DQN [17]. The evaluation network generates the Q-value for all actions and selects an action based on the Q-value, following the greedy strategy, which can be expressed as:

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha\left(r + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})\right) \tag{5}$$

Next, the reward r_t is obtained, and the experience $\{S_t, A_t, R_t, S_{t+1}\}$ is stored in the experience replay. M mini-batch samples are then randomly selected from the experience replay for training the evaluation network. The loss function used in the DQN algorithm is derived from the mean squared error (MSE) between the predicted Q-values and the target Q-values. The formula for the loss function is as follows:

$$\mathcal{L}(\delta_i) = E_{s_t, a_t}\left[\left(y_t - Q(s_t, a_t|\delta_i)\right)^2\right] \tag{6}$$

where δ is the parameter of DNN, and i is the index of iteration. y_t is the target Q-value for the state-action pair, calculated as the Bellman equation:

$$y_t = r(s_t, a_t) + \gamma \cdot \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}|\delta_i) \tag{7}$$

By the above method, the best route is determined.

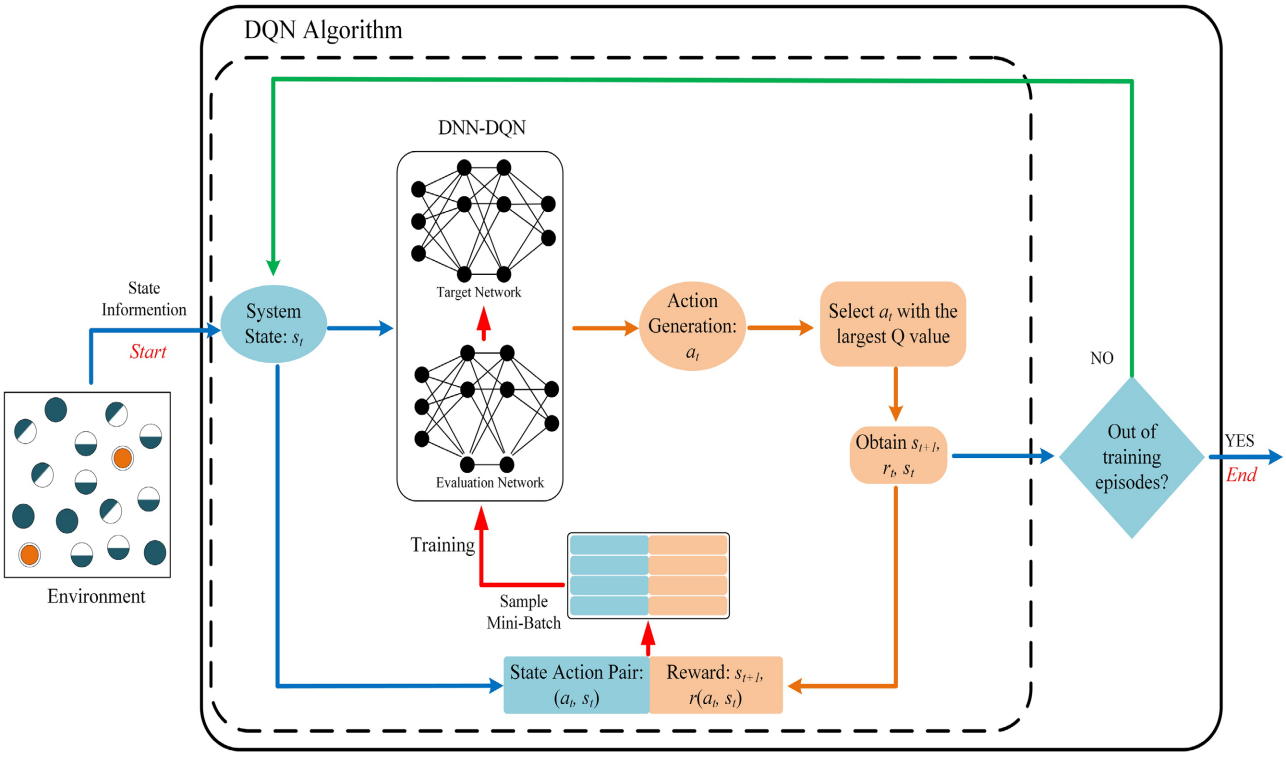

Figure 3. Routing protocol for HWSNs based on DQN.

4.2. DQN-Based Algorism for HWSNs Routing

As shown in Figure 3, the data forwarding mechanism in this study employs a multi-hop routing approach. Assuming that source node *a* possesses an ample cache of data packets locally, it forwards the data to CH A. Subsequently, CH A determines the subsequent routing hop by evaluating the Q-value. Additionally, the MAC address of CH A and node *a* is appended to the data header. The data packets are then broadcasted to neighboring CHs B and C. In cases where CH C does not receive the data, it discards the data content while retaining the header information. Once the corresponding CH B receives the data packet, it continues to propagate it to the next designated hop. Furthermore, there exists data transmission between CH A and CH B [18].

Given the above assumptions, the data packet header structure is obtained, and its overall is divided into two parts, including DRL-related data and transmission-related data, as shown in Figure 3. To ensure the complete and seamless transmission of data to the base station, the following data fields need to be incorporated into the header:

1) Data Packet ID: This field serves as a unique identifier for the data packet.

2) Source Address: Indicates the initial sending address of the data packet.

3) CH Address: Indicates the address of the CH node to which the source node belongs

3) Destination Address: Refers to the address of the base station where the data packet is intended to reach.

4) MAC Address of the Next Hop: Determines the receiving node for the subsequent hop.

Furthermore, to guarantee the timeliness of reinforcement learning, nodes should include their pertinent information in the data packet header. This information should encompass:

1) V-Value: Represents the V value associated with the current node. The V-value is a measure of the expected cumulative reward for the node in the reinforcement learning framework.

2) Remaining Energy: Indicates the remaining energy level of the current node.

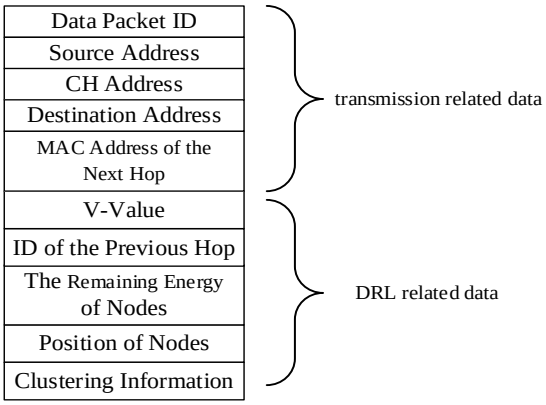

Figure 4. Data packet header structure of HWSNs.

3) ID or MAC Address of the Previous Hop Node: Includes either the ID or MAC address of the previous Hop Node.

As shown in **Figure 4**, by including these additional fields in the header, the data packet incorporates important node-specific information that aids in the efficient execution of reinforcement learning algorithms while ensuring effective communication among nodes [19].

Algorithm 1 provides a detailed process of the HDQN routing algorithm.

HDQN-based Algorithm for routing

Input: status of the node,
Output: routing policy
Initialize evaluation, target networks with parameters δ;
Initialize experience replay memory D;
for Episode = 1, 2, ..., N^{eps} do
  Initialize state s_t;
  for TS $t = 1, 2, ..., T$ do
  Obtain s_t;
  Select $a_t = \arg\max Q(s_t, a_t)$ with probability ε;
  Randomly select a_t with probability $1-\varepsilon$;
  Forward the data to the next node, obtain the corresponding reward from formula and s_{t+1};
  Update the current state to the next latest state to get new network input;
  Store transition $\{s_t, a_t, r_t, s_{t+1}\}$ into experience replay memory;
    if the learning process starts then
      Randomly sample M transitions from experience replay memory;
      Update evaluation network from formula;
      Calculate the target Q-value for the current state:

$$y_i = \begin{cases} r_j, & \text{if data is successfully forwarded to BS} \\ r_j + \gamma \max_{a_{t+1}} \hat{Q}\left(\phi_{j+1}, a_{t+1}; \theta^{-}\right), & \text{otherwise} \end{cases}$$

      Update target network periodically;
      end if
  end for
end for

5. Simulation Analysis of HDQN Algorithm

In this section, we use numerical simulation to evaluate the performance of our algorithm HDQN. Assume that the network coverage is 200 m × 200 m, where the BS coordinates are located at (100, 100), and 100 to 300 heterogeneous sensor nodes are randomly distributed. The parameters required for the experiment are shown in **Table 1**.

We ran the LEACH, SEP, DEEC, and HDQN protocols and found that when comparing the changes in the number of live nodes from 100 to 0 in each protocol, the HDQN protocol outperformed the other three protocols. It is observed from **Figure 5** that the HDQN protocol has a later death time for the first and last nodes compared to other protocols. This indicates that the protocol can effectively extend the lifecycle of HWSNs [20].

Table 1. Simulation parameters.

Simulation Parameters	Values
Network Area Size (m^2)	(200 × 200)
BS location	(100, 100)
Number of sensors	100, 200, 300, 400
Energy heterogeneity Node Type	3-level; normal, advanced and super nodes
Proportion of various heterogeneous nodes (super nodes: advanced nodes: normal nodes)	1:2:7
Initial energy of the normal node (J)	300
Initial energy of the advanced node (J)	600
Initial energy of the super node (J)	1200
Number of rounds	1000
Learning rate	0.01
Discount factor	0.9
E_{elec} (nJ/bit)	50
E_{amp} (pJ/bit)	0.0012
E_{fs} (pJ/bit)	10
d_{th} (m)	30

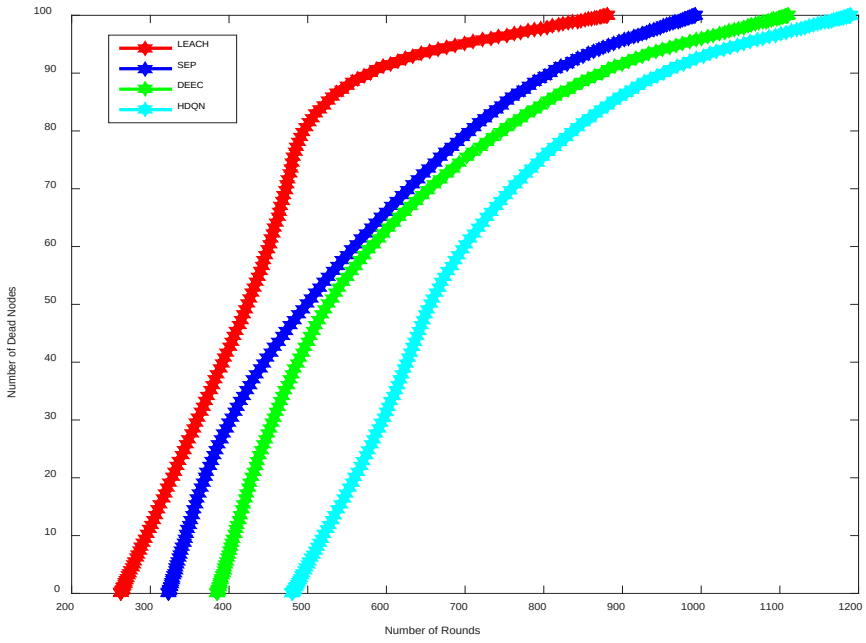

Figure 5. Dead nodes in the network for various algorithms.

Due to the specific working environment of HWSNs, the HDQN algorithm can intelligently allocate routine tasks, enabling high-energy sensor nodes to undertake more data transmission tasks, thereby extending the lifespan of the entire network. This personalized routing decision can effectively balance energy

consumption, thereby improving the overall energy efficiency of the system. As shown in Figure 5, the HDQN algorithm has a better node survival time than other algorithms.

Compared to some traditional routing algorithms, once the DQN algorithm establishes a Q-value function, sensor nodes can directly choose the optimal routing behavior based on their current state, without the need for complex path calculation and maintenance. This decentralized decision-making approach can reduce the complexity of the algorithm. Under the same monitoring environment, we gradually increased the number of nodes in HWSNs from 100 to 300, with the ratio of super nodes, advanced nodes, and ordinary nodes unchanged. As shown in Figure 6 the simulation experiment results show that the HDQN algorithm with different node numbers has more residual energy than LEACH, SEP, and DEEC.

Compared to some traditional routing algorithms, once the DQN algorithm establishes a Q-value function, sensor nodes can directly choose the optimal routing behavior based on their current state, without the need for complex path calculation and maintenance. This decentralized decision-making approach can reduce the complexity of the algorithm. As The DQN algorithm can adjust routing strategies in real-time based on the energy status and environmental changes of sensor nodes in the network. This enables the algorithm to flexibly respond to changes in energy consumption rates of different nodes, as well as dynamic adjustments to network topology. The algorithm can adapt to changes in the energy distribution of the network, thereby improving the stability and robustness of the network. As shown in Figure 7, the running time of HWSNs

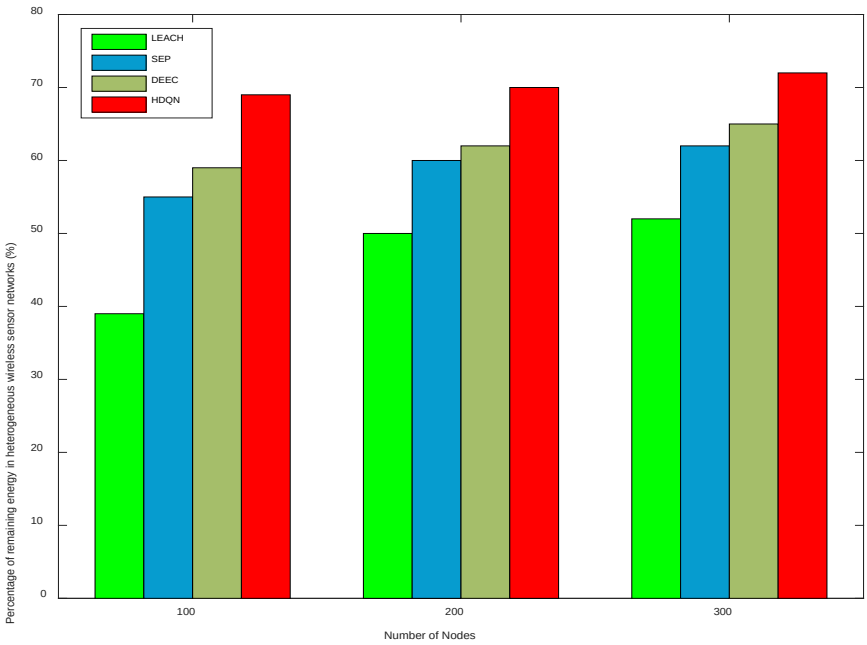

Figure 6. Comparison of changes in energy consumption when the number of HWSNs nodes increases.

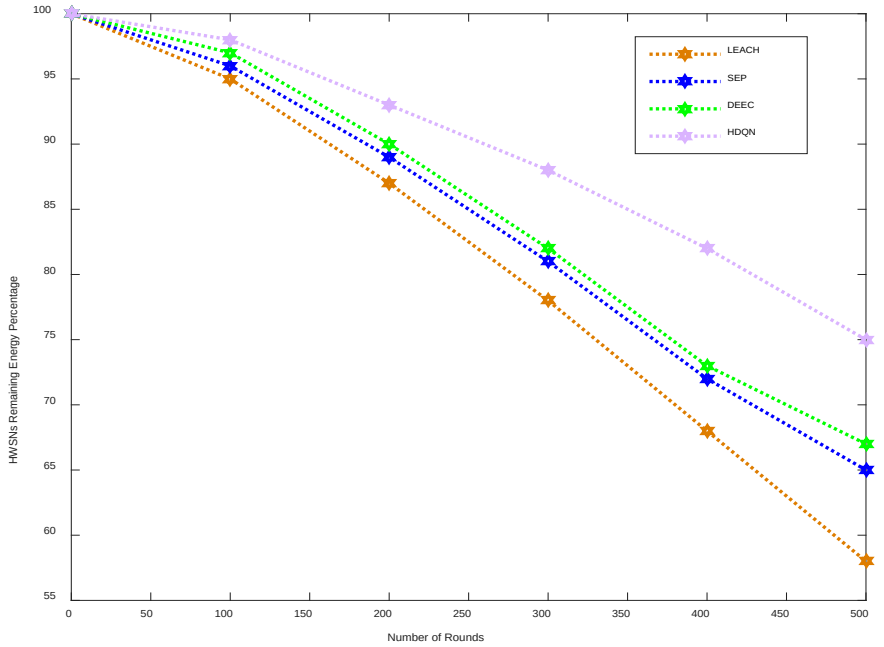

Figure 7. The percentage of remaining energy of HWSNs after each protocol runs the same cycle.

increases, the HDQN algorithm outperforms other algorithms in the simulation experiment of remaining energy.

6. Conclusion

Deep reinforcement learning technology is currently one of the most effective and reliable ways to solve intelligent decision-making problems. In order to demonstrate the advantages of deep reinforcement learning technology in HWSNs and further improve the level of intelligence, this paper proposes an energy-balanced heterogeneous routing algorithm HDQN based on deep reinforcement learning to address the shortcomings of classic routing algorithms in HWSNs, in order to intelligently achieve routing selection and energy consumption balance in HWSNs. This article designs the state space, action space, and reward function of the intelligent agent based on the characteristics of HWSNs, allowing the agent to intelligently select the best data routing for nodes in the network based on the current network state. The simulation experimental results show that the HDQN algorithm can significantly reduce the transmission time of data packets, have good routing performance, effectively improve the energy efficiency of HWSNs, avoid monitoring holes caused by premature "death" of HWSNs nodes, and extend the lifespan of HWSNs.

Funding

This research was funded by the Opening Project of Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things (2022WYY01), by the 2022 network ideological and political

education research project of Sichuan University of Science & Engineering (SZ2022-21), by the Zigong Key Science and Technology Plan Project (Collaborative Innovation Class of Zigong Medical Big Data and Artificial Intelligence Research Institute) (2022ZD16).

Conflicts of Interest

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

References

[1] Ghaderi, M.R., Tabataba Vakili, V. and Sheikhan, M. (2020) FGAF-CDG: Fuzzy Geographic Routing Protocol Based on Compressive Data Gathering in Wireless Sensor Networks. *Journal of Ambient Intelligence and Humanized Computing*, **11**, 2567-2589. https://doi.org/10.1007/s12652-019-01314-1

[2] Al-Sulaifanie, A.I., Al-Sulaifanie, B.K. and Biswas, S. (2022) Recent Trends in Clustering Algorithms for Wireless Sensor Networks: A Comprehensive Review. *Computer Communications*, **191**, 395-424. https://doi.org/10.1016/j.comcom.2022.05.006

[3] Pal, V., Singh, G. and Yadav, R.P. (2015) Effect of Heterogeneous Nodes Location on the Performance of Clustering Algorithms for Wireless Sensor Networks. *Procedia Computer Science*, **57**, 1042-1048. https://doi.org/10.1016/j.procs.2015.07.376

[4] Xu, S.W. (2016) Optimal Cluster Head Selection Mechanism for Heterogeneous Wireless Sensor Networks. *Computer System & Applications*, **25**, 187-191.

[5] Hatzivasilis, G., Papaefstathiou, I. and Manifavas, C. (2017) SCOTRES: Secure Routing for IoT and CPS. *IEEE Internet of Things Journal*, **4**, 2129-2141. https://doi.org/10.1109/JIOT.2017.2752801

[6] Tanwar, S., Kumar, N. and Rodrigues, J.J. (2015) A Systematic Review on Heterogeneous Routing Protocols for Wireless Sensor Network. *Journal of Network and Computer Applications*, **53**, 39-56. https://doi.org/10.1016/j.jnca.2015.03.004

[7] Chatap, A. and Sirsikar, S. (2017) Review on Various Routing Protocols for Heterogeneous Wireless Sensor Network. 2017 *International Conference on I-SMAC* (*IoT in Social*, *Mobile*, *Analytics and Cloud*) (*I-SMAC*), Palladam, 10-11 February 2017, 440-444. https://doi.org/10.1109/I-SMAC.2017.8058388

[8] Bdul-Qawy, A.S.H., Almurisi, N.M.S. and Tadisetty, S. (2020) Classification of Energy Saving Techniques for IoT-Based Heterogeneous Wireless Nodes. *Procedia Computer Science*, **171**, 2590-2599. https://doi.org/10.1016/j.procs.2020.04.281

[9] Botvinick, M., Ritter, S., Wang, J.X., *et al.* (2019) Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences*, **23**, 408-422. https://doi.org/10.1016/j.tics.2019.02.006

[10] Zhang, Y., Lan, J.L., Wang, P. and Hu, Y.X. (2015) A Polymorphic Routing Derivation Mechanism Based on Markov Decision Process. *Telecommunications Science*, **31**, 54-60.

[11] Sun, P., Hu, Y., Lan, J., *et al.* (2019) TIDE: Time-Relevant Deep Reinforcement Learning for Routing Optimization. *Future Generation Computer Systems*, **99**, 401-409. https://doi.org/10.1016/j.future.2019.04.014

[12] Aljohani, T.M., Ebrahim, A. and Mohammed, O. (2021) Real-Time Metadata-Driven Routing Optimization for Electric Vehicle Energy Consumption Minimization

Using Deep Reinforcement Learning and Markov Chain Model. *Electric Power Systems Research*, **192**, Article ID: 106962. https://doi.org/10.1016/j.epsr.2020.106962

[13] Feng, W., Xu, D., Xia, X.W., *et al.* (2022) Design and Simulation of Secure Routing Algorithm for Wireless Body Area Networks. *Research and Exploration in Laboratory*, **41**, 138-142, 147.

[14] Huang, L., Ye, M., Xue, X., *et al.* (2022) Intelligent Routing Method Based on Dueling DQN Reinforcement Learning and Network Traffic State Prediction in SDN. *Wireless Networks*. https://doi.org/10.1007/s11276-022-03066-x

[15] Huang, X.-Q., Liu, A.-J., Liang, X.-H. and Wang, H. (2022) Load-Balanced Geographic Routing Protocol in Aerial Sensor Network. *Computer Science*, **49**, 342-352.

[16] Roy, M., Biswas, D., Aslam, N., *et al.* (2022) Reinforcement Learning Based Effective Communication Strategies for Energy Harvested WBAN. *Ad Hoc Networks*, **132**, Article ID: 102880. https://doi.org/10.1016/j.adhoc.2022.102880

[17] Wang, Z.S., Ding, H.W., Li, B., *et al.* (2021) Energy-Efficient WSNs Routing Protocol Based on Clustering. *Computer Engineering and Design*, **42**, 324-330.

[18] Zhou, R.Y., Chen, M., Feng, G.F., *et al.* (2010) Genetic Clustering Route Algorithm in WSN. 2009 6*th International Conference on Natural Computation*, **8**, 4023-4026. https://doi.org/10.1109/ICNC.2010.5584826

[19] Li, H., Ou, J., Cui, H., *et al.* (2022) GKFCR: An Improved Clustering Routing Algorithm for Wireless Sensor Networks. 2022 *IEEE International Conference on Sensing*, *Diagnostics*, *Prognostics*, *and Control* (*SDPC*), Chongqing, 5-7 August 2022, 222-227. https://doi.org/10.1109/SDPC55702.2022.9915965

[20] Muthanna, M.S.A., Muthanna, A., Rafiq, A., *et al.* (2022) Deep Reinforcement Learning Based Transmission Policy Enforcement and Multi-Hop Routing in QoS Aware LoRa IoT Networks. *Computer Communications*, **183**, 33-50. https://doi.org/10.1016/j.comcom.2021.11.010