Hindawi
Journal of Sensors
Volume 2022, Article ID 5969130, 10 pages
https://doi.org/10.1155/2022/5969130

Research Article

Semantic Interaction Strategy of Multiagent System in Large-Scale Intelligent Sensor Network Environment

Xi Chen,[1] **Zhaoyang Yin**,[2] **and Miaomiao Zhu**[3]

[1]*School of Art and Media, Hubei University of Business and Commerce, Hubei Wuhan, 430073, China*
[2]*School of Art, Hubei University, Hubei Wuhan, 430062, China*
[3]*School of Fine Arts and Design, Huainan Normal University, Anhui Huainan, 232038 Anhui, China*

Correspondence should be addressed to Miaomiao Zhu; zmm@hnnu.edu.cn

Received 20 November 2021; Revised 21 December 2021; Accepted 10 January 2022; Published 30 January 2022

Academic Editor: Gengxin Sun

In a multiagent system, the semantic interaction between agents is an important aspect affecting multi-intelligence. The purpose of interaction is to reasonably arrange task objectives and behaviors through information sharing and communication interaction, so as to maximize the overall performance of multiagent system. This paper analyzes the communication and interaction process between agents from the perspective of semantic layer and introduces the BDI (belief, desire, intention) model of agent's thinking state into the communication and interaction process. Furthermore, we propose a multiagent semantic interaction strategy model based on a large-scale intelligent sensor network, which supports various types of negotiation and interaction on the basis of basic interaction behavior to solve the problem of information operational conflicts. In addition, this paper limits the scale of historical information through the definition of equivalence and the merging theorem of history, and it uses reinforcement learning algorithm to detect possible conflicts and delay communication and makes rational use of limited resources to improve system revenue and coordination efficiency. The experimental results show that compared with the previous methods such as debate and negotiation, the strategy model can realize the flexible interaction based on scene and is more practical. At the same time, the existence of reinforcement learning improves the efficiency analysis and the convergence performance of semantic interaction strategy.

1. Introduction

With the development of artificial intelligence (AI), a multiagent system (MAS) has become a research hotspot. A multiagent system is composed of a group of independent and interactive autonomous agents [1, 2]. As an interactive autonomous learning paradigm, reinforcement learning provides an effective method to solve the distributed collaboration of multiagent systems [3]. Multiagent Reinforcement Learning (marl) has attracted extensive attention [4]. Drawing on the technologies and concepts of artificial intelligence, game theory, psychology, and sociology, marl provides a promising method to learn satisfactory agent behavior in complex environment, which is widely used in distributed control, multirobot system, resource allocation management, and automatic transaction [5].

In the past, the research on agent interaction in multiagent system (MAS) can be divided into two main parts: communication mechanism and negotiation method, but they lack connection and universality [6]. In the design of MAS system, in order to enable agents to obtain semantic information from exchange data, it is necessary to have a new understanding of the content and mode of communication interaction [7]. Firstly, communication should not be a passive behavior determined by the protocol, but the behavior that one agent wants another agent to accept some kind of belief or intention in the scene of communication [8]. The specific answer to each communication should be determined by the interactive target [9]. This interaction model can be applied to flexible interaction scenarios and provide means for communication based on target requirements [10]. Secondly, interaction is to share the information of

both sides, understand the intention of both sides, and adjust their plans in a certain order [11]. As long as the transmission and inquiry of information, intention and planning adjustment are expressed; a considerable number of interactive processes can be expressed [12]. And its scope of application is only limited to the planning ability of agent to take interactive action to achieve its purpose [13].

Due to environmental uncertainty, incomplete information, distributed learning, concurrent learning, and other problems, multirobot system (MRS) is widely used in UAV, spacecraft, autonomous underwater vehicle, ground mobile robot, and other practical problems [14]. As an interaction-oriented autonomous learning paradigm, Multiagent Reinforcement Learning (RL) allows robots to learn the mapping from state to action through the reward obtained by interaction with the environment, so as to cooperate with robot behavior and complete specific tasks, which is widely used in multirobot systems [15–17]. In reinforcement learning, each agent learns the optimal strategy by interacting with its dynamic environment [18]. When single agent reinforcement learning is applied to a multiagent system, reinforcement learning faces some challenges. The centralized learning method regards the multiagent system as a whole [19]. Through the observation of the global environmental information, the single agent reinforcement learning method is applied to learn the joint optimal behavior of the multiagent system [20]. Because it depends on the scale of real problems, centralized learning methods usually face scalability problems. Therefore, the centralized learning method can not be applied to multirobot systems. In a multirobot system, each robot needs to have complete control over the individual robot, that is, the distributed control of multirobot system. As a model-free reinforcement learning method, Q-learning has been widely used in multirobot systems such as soccer robot, chasing robot [21], chasing robot prey, and moving target observation robot [22]. The literature [23] applied the independent Q-learning algorithm to the soccer robot to realize the cooperation of robot behavior. The documenters [24] improved the learning efficiency through robot cooperative learning. This research work accelerated the learning process by sharing perceptual information and learning experience [24]. The distributed independent learning method models each agent, and each agent only observes its local environment. The distributed independent learning method does not rely on the observation of global environment information. It has the characteristics of high robustness and good scalability. At the same time, it can solve the dimensional disaster problem faced by centralized learning [25]. The contributions of this paper are summarized as follows: (1) this paper analyzes the communication and interaction process between agents from the perspective of semantic layer and introduces the BDI model of agent thinking state into the communication and interaction process; on the basis of basic interaction behavior, it supports various types of negotiation and interaction to solve the problem of information operation conflict. (2) This paper limits the scale of historical information through the definition of equivalence and historical merging theorem, uses reinforcement learning algorithm to detect possible conflicts and delayed communication, and makes rational use of limited resources to improve system revenue and coordination efficiency. (3) This paper constructs a large-scale intelligent sensor network system to verify the superiority and reliability of the algorithm.

In this paper, for the behavior coordination problem in multiple environments, an improved reinforcement learning mechanism based on planning fusion is proposed. The history and belief information are expressed as a function of the state. On the premise of ensuring that there is no loss of effective information, the historical information is combined by the methods of possible conflict detection and delayed communication, and the limited resources are reasonably used to obtain more system benefits. The mechanism takes the belief pool as the basic way of inter coordination and uses the strategy merging theorem to losslessly merge the historical information, so as to improve the efficiency of solving the problem with large-scale historical information. At the same time, the mechanism of conflict detection and delayed communication is adopted to effectively use the limited communication resources to strengthen the resolution of behavior conflicts and the exchange of important information.

2. Architecture Design of Multiagent System

2.1. Multiagent System Architecture. A multiagent system is an important field in the application of multiagent technology. The multiagent system is a group organization with multiple independent abilities. Each has a certain thinking state, such as belief, knowledge, and intention, and they will perform some actions according to their thinking state. The necessity of coordination lies in the existence of other intentions. The purpose of coordination is to change individual intentions and enable all individuals in the system to work together in a consistent and harmonious way. The goal of multiagent system is to make several systems with simple intelligence but easy to manage and control realize complex intelligence through mutual cooperation, so as to reduce the complexity of system modeling and improve the robustness, reliability, and flexibility of the system. The main characteristics of multiagent system are as follows:

(1) Autonomy: in the multiagent system, each agent can manage its own behavior and achieve independent cooperation or competition.

(2) Fault tolerance: agents can jointly form a cooperative system to achieve independent or common goals. If some agents fail, other agents will independently adapt to the new environment and continue to work, and the whole system will not fall into a failure state.

(3) Flexibility and scalability: MAS system itself adopts distributed design, and the agent has the characteristics of high cohesion and low coupling, which makes the system show strong scalability.

(4) Ability to collaborate: the multiagent system is a distributed system. Agents can cooperate with each

other to achieve the global goal through appropriate strategies.

Deliberative type is also called knowledge type or cognitive type. Its biggest feature is to use symbols to realize the representation and reasoning of entities in the real world and make decisions according to the reasoning at a certain stage. There are only some simple actions in reactivity-perceptual behavior pattern. The above two are extreme representations of two ways of thinking. The cautious type requires strict theoretical background such as knowledge representation, behavior planning, and decision-making strategies. The real implementation process is too complex. Although the reactive type is simple, it only makes reasoning and decision-making according to local perceived information, and empirical knowledge can not be effectively used, so it is difficult to effectively solve practical problems. Therefore, there is a hybrid type. It integrates the characteristics of the above two types and can make up for each other to a certain extent. It is the most ideal structural model. The architecture diagram is shown in Figure 1.

As shown in Figure 1, the multiagent architecture mainly includes environment awareness module, information processing module, communication module, decision and control module, and execution module. In addition, when an agent predicts environmental changes, it should consider that the activities of other agents are generally not controlled by themselves and difficult to predict. In order to better predict environmental changes, enhance their own action ability, and realize their own needs, agents must communicate. The capability of a single agent is limited, but multiple agents can be organized through an appropriate architecture to make up for the shortcomings of each agent and make the capability of the whole system exceed that of any single agent. A multiagent system means that a problem needs multiple solving entities. This system has the advantages of traditional distributed and concurrent problem solving and has complex interaction mode. Communication ability is not a necessary characteristic of rational agents; it is the embodiment of agent sociality. Communication action is also a specific planning action, which is scheduled in the process of completing agent requirements. From the semantic level, communication interaction is the transmission of thinking state between agents.

2.2. Structure of Multiagent System Based on Large-Scale Sensors. A distributed cooperative control method based on multiagent system is shown in Figure 2, which is characterized by the following steps: (1) build a multiagent three-level control architecture, that is, a control architecture of "local droop control-secondary power optimization control-centralized optimization and regional autonomy." Each network using droop control installs agents to realize the semantic interaction of multiagents. (2) The dispatching decision-making function module is designed to coordinate the adjustable resources with different control response rates, which respond to the internal and external energy demand of the LAN and quickly stabilize the power fluctuation of the tie line in the process of power failure, parallel, and off network switching in the energy LAN. (3) A distributed sparse communication network based on the multiagent system is constructed. Furthermore, the generation and completion of communication must have certain objective conditions, such as the existence of communication carrier and other factors. At the same time, there must be explicit intention for information exchange in communication. No matter whether the communication medium is language or action, the sender knows that its intention will be received by the other agents; the receiver of communication must also have the need to receive information. Communication is also a group behavior between two agents, which cannot be fully represented by only one agent sending information. The occurrence and implementation of communication depend on the existence of agents similar to themselves in the world information of other agents following the same communication processing mode and thinking mode. In addition, a large-scale intelligent sensor network is an information collection platform. It is a multihop self-organizing network system formed by a large number of cheap sensor nodes. Sensor nodes can collect the information of the monitoring area in real time, transmit the collected information to multiagent through multihop routing, and realize the semantic interaction between multiagent. Therefore, the large-scale intelligent sensor network is cooperation an important part of multiagent cooperation.

Intention transfer has a direct impact on agent behavior and can be used for behavior coordination among agents. Intention is expressed as the expected world state in the thinking state of agent, which has the same expression as observation information and knowledge, and can be transmitted as information and knowledge. After the intention is transferred to an agent a, a should decide whether to take it as his intention and start planning and action Starting from the self-interest principle of autonomous agent, agent adopting the intention of other agents should help to improve its effectiveness. There are many choices in which criteria the agent chooses the acceptable intention. Therefore, the construction of large-scale wireless sensor networks is the basis of multiagent intention transmission.

3. Research on Semantic Interaction Strategy of Multiagent System Based on Planning Fusion

When an agent predicts environmental changes, it should consider that the activities of other agents are generally not controlled by themselves and difficult to predict. In order to better predict environmental changes, enhance their own action ability, and realize their own needs, agents must communicate. Communication ability is not a necessary characteristic of rational agents. It is the embodiment of agent sociality. Communication action is also a specific planning action, which is scheduled in the process of completing agent requirements. From the semantic level, communication interaction is the transmission of thinking state between agents. Reinforcement learning is based on the premise that the interaction between agent and environment is regarded as a Markov decision-making process; that is, the next system state is determined only by each current state and

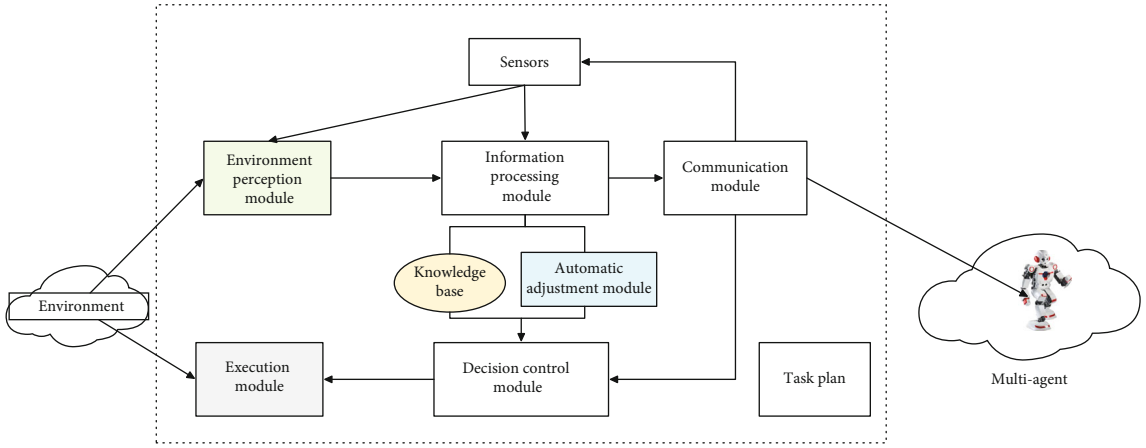


Figure 1: Multiagent architecture diagram.

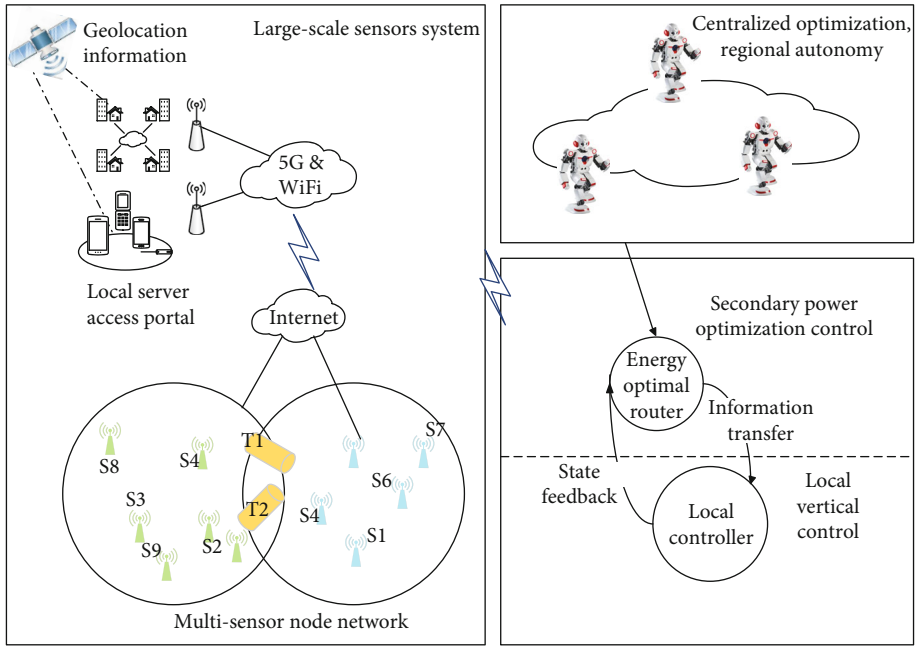


Figure 2: A distributed cooperative control framework based on multiagent system.

selected action, and a fixed state transition probability distribution is determined, which is independent of the previous historical state. The goal of learning is to find a strategy to maximize future reward by sampling the environment. Experience is an important basis for future behavior selection. Learning accumulated experience is an effective way to solve the problem of semantic interaction. The overall framework of the algorithm is shown in Figure 3.

As shown in Figure 3, the algorithm framework mainly includes reinforcement learning units, BP agent decision, and candidate model. The whole multisystem has gone through a stage of reinforcement learning process. The system stores the corresponding knowledge system and has the ability to adapt to the changes of the external environment. When it is determined that the system enters the emergency state, the management compares the historical data stored in the database with the real-time data sent by the guidance office, finds out the similarities, and assists in the decision-making according to the optimal decision made when the historical data occurs. If there is no similar historical data, the management will combine other reinforcement learning processes, make tentative action attempts, obtain the feedback of the environment, and then modify the decision judgment and cycle to obtain the optimal solution. Since it is set in this section that the system has been running for a long time and has corresponding knowledge base for data and decision support, it is assumed that the management can directly make the optimal decision for this accident.

3.1. Improved Reinforcement Learning Mechanism Based on Planning Fusion. The perceptron can sense the changes of the external environment and other actions and states. When encountering a learned situation, take action directly

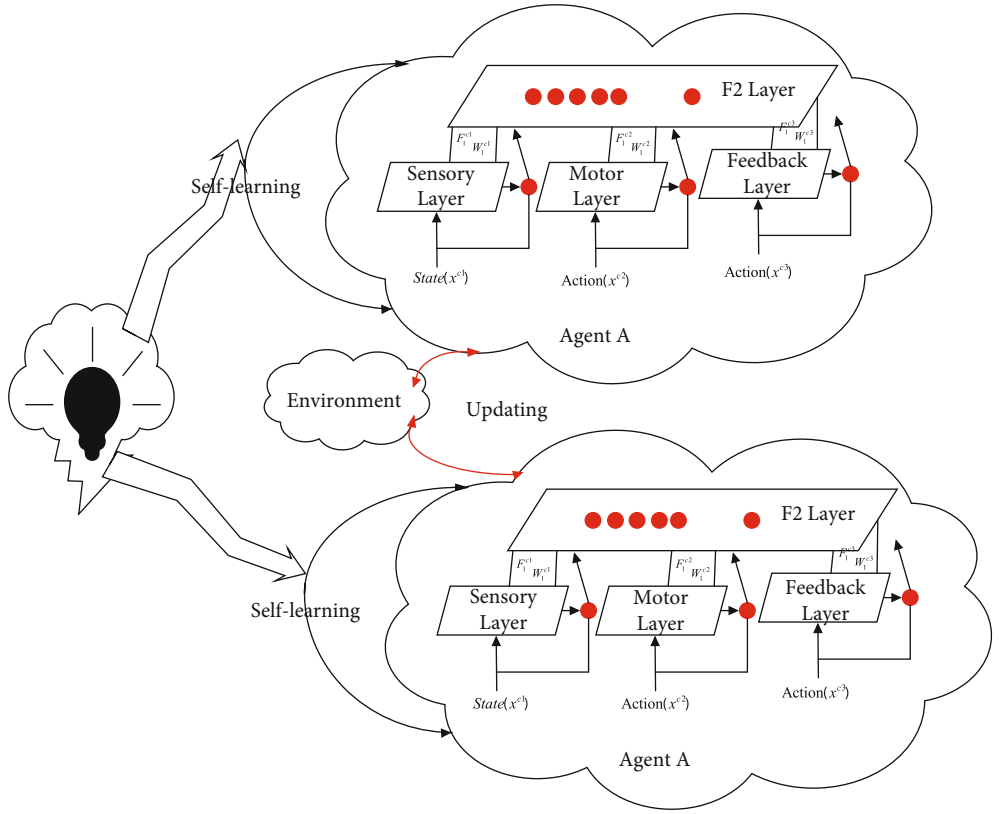


Figure 3: Improved reinforcement learning structure based on planning fusion.

through reflection. When encountering a new, learned, and more complex situation, combined with other behaviors and states, make behavioral decisions and take optimal actions through cooperative reinforcement learning strategy.

In some cases, the behavior selection of multiagent depends partly on the past behavior history, and h is defined as the action sequence executed and its observation sequence. At the time when the time step is t, the history of agent i can be expressed as

$$h_i^t = \left(a_i^0, o_i^1, a_i^0, \cdots, o_i^{t-1}, a_i^{t-1}, o_i^t\right), \tag{1}$$

where $h' = [h_1, h_2, \cdots, h_n]$ is the joint matrix of multiagent semantic interaction.

The joint belief b of multiagent is a function of joint history h', that is, $b(h) \in \Delta(S)$. Its essence is a probability distribution of environmental state, which is composed of initial belief state and sufficient statistics of joint history. If the joint history h^{t-1} before time t is known, the method of calculating the joint belief b^t at the current time can be obtained by using Bayesian rules:

$$\forall s' \in S, b^t\left(s' \mid h'\right) = \frac{O\left(o^t, s', a^{t-1}\right)\sum_{s\in S} P\left(s' \mid s, a^{t-1}\right) b^{-1}\left(s \mid h^{-1}\right)}{\sum_{s\in S} O\left(o^t, s'', a^{t-1}\right)\sum_{s\in S} P\left(s'' \mid s, a^{t-1}\right) b^{-1}\left(s \mid h^{-1}\right)}. \tag{2}$$

3.1.1. Local Joint Strategy δ. It is set that δ_i is a mapping from the history set h to the action set A. This mapping is determined and becomes the local determination strategy of agent i. It is easy to understand that $\delta(h) = \langle \delta_1(h_1), \delta_2(h_2), \cdots, \delta_n(h_n)\rangle$ means multiple joint determination strategies at a certain time and $\delta(h) = \vec{a}$. The local random strategy $\pi_i(a_i|h_i)$ represents the mapping from the historical set to the action probability distribution. It is different from the local determination strategy, because the uniqueness of the selected action cannot be determined according to the historical information but can only make the action selection process obey a certain probability distribution $\pi(h) = \langle \pi_1(h_1), \pi_2(h_2), \cdots, \pi_n(h_n)\rangle$ representing the local joint random strategy of multiagent.

3.1.2. Belief Pool. The belief pool at time t is represented as a binary array $\langle \{H_i^t(i \in I)\}, B^t\rangle$, where H_i^t is represented as the history of agent I at time t, and B^t is represented as the joint belief of multiagent at time t. The purpose of setting belief pool is to provide a medium for information sharing and coordination among multiagents.

In a MAS environment, due to the multifunction and heterogeneous structure, it is impossible to determine the behavior rules of other agents in many cases, so agents must interact and coordinate to jointly complete the overall goal. Therefore, agents need to be able to understand the strategies and knowledge of other agents through online learning, so as to determine the optimal behavior strategy and adapt to the changes of system environment. In this case, the state

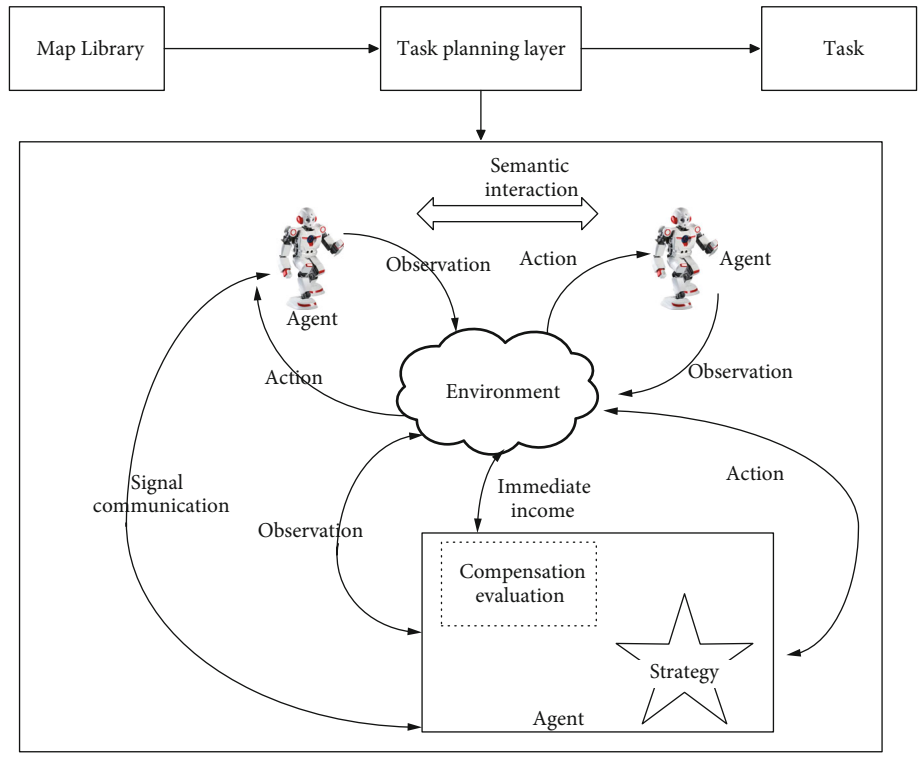


Figure 4: Improved Q-reinforcement learning model of multiagent.

change of a single agent is affected by the joint action of agents, so the traditional Q-learning formula needs to be extended. The q-reinforcement learning model under MAS is shown in Figure 4.

The immediate reward function under the defined environment is $R(s, \vec{a})$, where $\vec{a}$ is the joint action of system. The state transition function is $P(s_i, \vec{a}, s_j)$. The corresponding modification of the value function by the action of state i is as follows:

$$Q^{\pi}(s,a) = R\left(s, \vec{a}\right) + \gamma \sum_{s \in S} P\left(s' \mid s, \vec{a}\right) \max\left(s', \vec{a}'\right). \quad (3)$$

Equation (3) represents the discount income obtained by executing the joint action in the state and iteratively executing it according to the principle of optimal reward value. Therefore, Q function update formula is

$$Q_{t+1}\left(s, \vec{a}\right) = (1-\alpha_i) Q_t\left(s, \vec{a}\right) + \alpha_i \left[r_i + \gamma \max Q_t\left(s', \vec{a}'\right)\right], \quad (4)$$

where α_i is the dynamic learning rate or discount factor and t represents the number of iterations.

3.2. Research on Communication of Semantic Interaction. In the planning fusion framework, agents learn each other's knowledge by sharing belief pool, so as to maintain the coordination between agents. However, there is such a problem that we need to focus on considering that the belief pool contains all the historical information, but in some cases, there may be conflicts between these historical information and the local observation of the agent. At this time, communication can be used to deal with these conflicts more effectively and improve the efficiency of coordination.

If the agent understands the current system state, the detection will become easy, but in the planning fusion environment discussed above, these states cannot be known. Each agent can understand in the execution stage which is its local observation of the environment, and the local observation can only provide part of the information about the current system state. However, we can determine whether there is conflict by detecting the relationship between these local observations and belief pool. Equation (5) formally defines the conflict between the two agents.

When the belief pool B^t satisfies the following formula, we call the conflict degree ε between B^t and local observation o_i^t.

$$\max\left\{\sum_{s' \in S} O\left(\vec{o} \mid s', \vec{a}\right) \sum_{s \in S} P\left(s' \mid s, \vec{a}\right) b(s)\right\} \le \varepsilon. \quad (5)$$

In essence, it is to test the conflict between the local history of agent i and its observation. The value of ε is determined by the observation function. If the observation uncertainty is very small, the value of ε is correspondingly small. Nevertheless, the above method cannot detect all conflicts in the belief pool, but only conflicts based on current observations. The number of communication times is determined by two factors: the observed structure and the heuristic

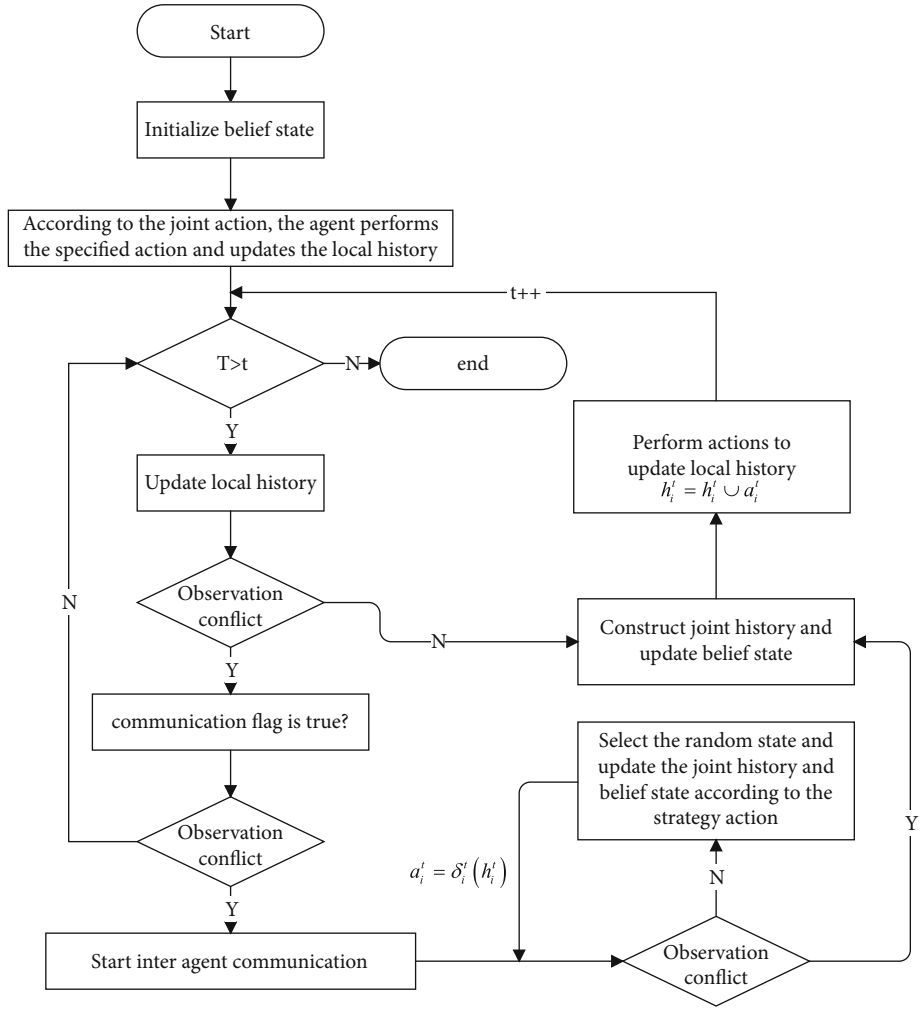


Figure 5: Implementation flow chart of algorithm.

function. The advantage of adopting this communication coordination method is that when the communication conditions are not available, it is allowed to delay until it meets the communication conditions. In many previous communication coordination methods, when the number of failures exceeds a certain limit, they tend to adopt extreme methods for coordination processing or ignore local observation information and completely rely on history, alternatively relying entirely on the current local observation to derive adverse results. To sum up, the implementation flow chart of the algorithm is shown in Figure 5.

As is shown in Figure 5, the main work is carried out in the first two stages. In the behavior planning stage, the local history of each agent in the belief pool is used to calculate the joint strategy δ'. In the execution stage, each agent updates the local historical information and first updates the historical information of the previous step according to its latest observation o. Then, according to the joint strategy δ' calculated in the planning stage and the current local history h pair, the corresponding actions are performed through $h_i^t = h_i^t \cup a_i^t$ calculation, and $a_i^t = \delta_i^t(h_i^t)$ is used at the same time. Update the current local history for the last update stage, and update the old local history of each in the belief pool with the foot of the execution stage.

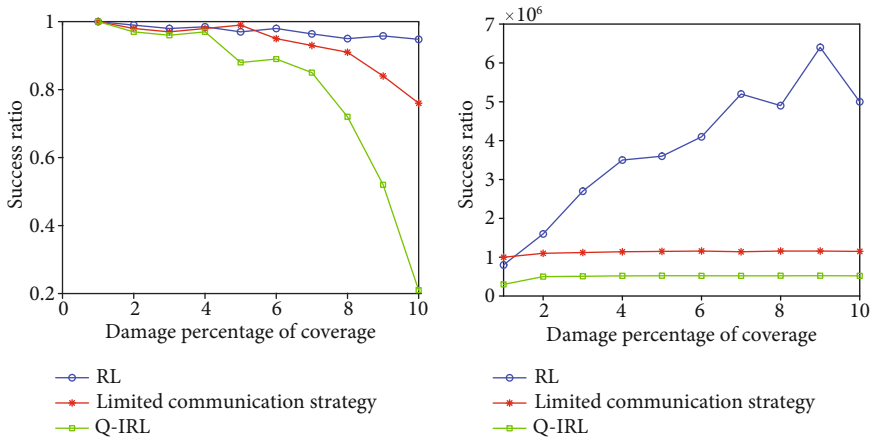
4. Experiment and Result Analysis

4.1. Experimental Setup and Experimental Environment

4.1.1. Environment State. Each agent may be in any grid other than the obstacle grid. The state of the machining center can be either idle or working, so the environment state space of a single agent is $13 * 2$ and the joint state space of two agents is $13 * 13 * 2$.

4.1.2. Action Space. Each action has a kind of move up, move down, turn left, turn right, and maintain the original position. Therefore, the size of action space is $5 * 5$.

4.1.3. Observation Set. Bit binary characters are used to represent each observation. The first bit indicates whether there are obstacles in the upper, left, lower, and right directions. The obstacles here include obstacles and surrounding walls. The second bit indicates whether it is currently in the top grid. For two agent systems, the observed set size is $2^{5\times2}$.



(a) Fault tolerance success ratio (b) Communication loss

Figure 6: Efficiency comparison of three semantic interaction strategies.

4.1.4. Historical Information. In MAS, the joint observation of agents in some observable environments can not accurately reflect the current state information of the group, and it is also necessary to express the trust of each observation through belief. The popular understanding of belief is a subjective view of what actions should be performed in a certain state according to the laws of experience or historical statistics.

4.1.5. Revenue Function. The immediate return value of each step is $r = -1$ when hitting an obstacle. Cooperation to achieve the goal is$r = 10$. Other return values are $r = 0$.

Assuming that the processing time of the product is small enough, it can be ignored that each agent does not know the return value of its other agents at the beginning of learning, obtains the return through learning, and guides the next step. At the end of a round of experiment, agent 1 sends the goods from the processing center to the processing center, and agent 2 takes the goods back from the processing center to u.

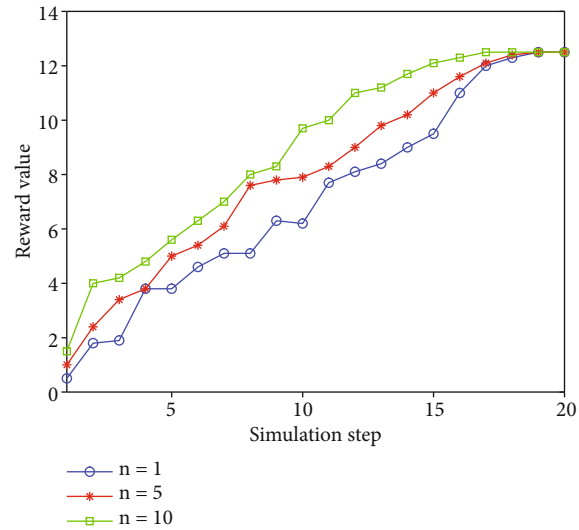


Figure 7: Relationship between discount reward and simulation step size.

4.1.6. Simulation Environment and Parameters. The simulation process is realized by MATLAB. With the help of POMDP solver open source software package, the time step of the problem is set to $t = 200$, and the discount return is calculated every 10 steps. The simulation parameters are as follows: $\alpha = 0.8$; $\varepsilon = 0.74$; and $l = 8$. The simulation parameter n is a parameter that can only be determined by experiment.

4.2. Efficiency Analysis of Semantic Interaction Strategy. In this section, Java language is still used to design multiagent sensor system simulation platform in Eclipse: within the rectangular range of $800 * 600$, 50-38, the fourth plane is randomly distributed to activate 300 sensor agents in the cluster fault-tolerant method. The number of agents in different experiments is different and increases with the step value of 50. Agents have unique identification IDS, but their performance is the same. The sensing and communication range is set to a circle with a radius of 100. At the beginning of the experiment, the system is initialized, and each agent generates NT table information and saves it separately. The abscissa in Figure 6 shows that the number of agents gradually increases from 50 to 300, and the density of node distribution in the experiment is briefly explained. Figure 6(a) describes the difference in fault tolerance success rate between the three methods. Figure 6(b) describes the different communication losses of the three methods and takes all the communication times during the experiment as the reference standard of communication loss.

It is obvious from Figure 6(a) that there is no obvious difference in the fault tolerance of the three methods at the beginning. However, as the error model is closer and closer to large-scale centralized errors, this paper proposes that the error tolerance success rate of the semantic interaction strategy of multiagent system in the environment of large-scale intelligent sensor network is getting lower and lower, and there is no obvious advantage in communication loss. Because the conventional reinforcement learning and the activation cluster under the semantic interaction mechanism

Table 1: Convergence performance.

Case number	Strategy	Minimum iterations	Maximum iterations	Average iterations	Success rate
Case 1	Our	11	46	22.6	20/20
	Traditional	26	103	62.8	20/20
Case 2	Our	22	51	29.5	20/20
	Traditional	51	286	104.4	17/20
Case 3	Our	8	28	18.7	20/20
	Traditional	12	200	56.8	19/20

do not care about the size of errors and the number of failed nodes. In addition, the range of errors in the experiment is fixed, the range of activation cluster is relatively fixed and the change of communication loss is not obvious. Under the semantic interaction strategy mechanism of multiagent system in a large-scale intelligent sensor network environment, the activated clusters are divided into each other by the ID of failed nodes. Although it can prevent the overflow of information between different clusters, such activated clusters cannot fully reflect the scale of errors. Therefore, the communication loss is greatly increased, but the success rate of fault tolerance is greatly reduced. According to Figures 6(a) and 6(b), it can also be found that although the conventional algorithm is designed for large-scale centralized errors, it can also deal with a single centralized error well and better take into account the fault-tolerant success rate and communication loss.

The word reinforcement learning comes from behavioral science. It imitates the natural learning process of human and animals and establishes the mapping from environmental state to behavior through repeated exploration of the environment Therefore, simulation step size is one of the most important parameters of reinforcement learning algorithm. In order to verify the effect of the strategy in this paper, this paper verifies the variation law between the reward values harvested. The relationship between the discount reward and the simulation step n is shown in Figure 7.

It can be seen from Figure 7 that when the number of experiments is divided into 1, 5, and 10, the change trend of the reward value was obtained through reinforcement learning. Since the last belief state is calculated as the next initial belief state value, the learning effect will gradually become ideal after multiple rounds of experiments. In this paper, the simultaneous interpreting strategy is adopted in the improved learning mechanism of planning and integration. Compared with the traditional distributed communication strategy, the most important feature of the scheme is to make timely and appropriate use of communication resources with limited resources and large unit communication costs. The coordination between them is mostly carried out by means of information sharing. In addition, in reality, we often encounter a series of optimization problems under different parameters In the case of a specific structure, the optimization problem under all parameters is solved by training a model for different parameters. Different from the traditional method, we do not train our model by multiple independent sampling of different parameters but use reinforcement learning to accelerate the training process. In a reinforcement learning algorithm, the strategy network is used to obtain the optimization results and the value network is used to evaluate the strategy. The two networks are trained iteratively to optimize the strategy.

4.3. Comparison of Convergence Performance of Strategies. In order to verify the performance indexes of the improved algorithm, reinforcement learning algorithm and improved Q-reinforcement learning algorithm are selected as reference in the experiment. If the algorithm reaches the set accuracy within the specified number of iterations, the convergence of the algorithm is recognized. If the number of iterations exceeds and the set accuracy is not reached, the algorithm terminates, and it is considered that the algorithm does not converge. The test results of 20 times are shown in Table 1.

As can be seen from Table 1, the improved Q-reinforcement learning algorithm achieves better experimental results than the basic reinforcement learning algorithm. Under the same precision, for the three test functions, although the two algorithms can successfully complete the optimization task, the difference in optimization speed is obvious. The proposed algorithm has more advantages in the number of iterations required for algorithm convergence.

5. Conclusion

In the multiagent system, the coordination degree between agents has an important impact on the overall intelligence of multiagent system. The purpose of coordination is to reasonably arrange task objectives and behaviors through information sharing and communication interaction, so as to maximize the overall performance of multiagent system. The communication of agent is to change the information carrier and send the carrier to the observable environment receiving Ag NT. This communication view can expand the form of communication, not limited to language communication. The transmission of intention has a direct impact on the behavior of agents and can be used for behavior coordination among agents. Intention is expressed as the expected world state in the thinking state of agents. After the intention is transmitted to agent a, a should decide whether to take it as his intention and start planning and action. An improved learning mechanism based on planning fusion is proposed to express the history and information as a function of the state. On the premise of ensuring no loss of effective information, the method of possible conflict detection and delayed communication is adopted for historical information merging, and the limited resources are reasonably

used to obtain more system benefits. Through experiments, the effectiveness of above strategies is analyzed and compared. In addition, how to use reinforcement learning and the reinforcement learning process before cooperation to deal with a virtual event to illustrate how the agent system determines the accident, makes decisions, and solves the accident after the accident and how the agents cooperate with each other is the focus of the next research.

Data Availability

Data are available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Hubei University of Business and Commerce.

References

[1] S. Listopad, "Architecture of the hybrid intelligent multi-agent system of heterogeneous thinking for planning of distribution grid restoration," *Baltic Journal of Modern Computing*, vol. 7, no. 4, pp. 487–499, 2019.

[2] F. Yang, Y. Qiao, S. Wang, X. Wang, and X. Wang, "Blockchain and multi-agent system for meme discovery and prediction in social network," *Knowledge-Based Systems*, vol. 229, 2021.

[3] Q. Wu, J. Wu, J. Shen, B. Yong, and Q. Zhou, "An edge based multi-agent auto communication method for traffic light control," *Sensors*, vol. 20, no. 15, 2020.

[4] A. Benayache, A. Bilami, K. Benaggoune, and H. Mouss, "Industrial IoT middleware using a multi-agent system for consistency-based diagnostic in cement factory," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 14, no. 3, pp. 291–310, 2021.

[5] D. Wu, Q. Liu, H. Wang, D. Wu, and R. Wang, "Socially aware energy-efficient mobile edge collaboration for video distribution," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2197–2209, 2017.

[6] E. Ahmed and H. Gharavi, "Cooperative vehicular networking: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 996–1014, 2018.

[7] A. Belhadi, Y. Djenouri, G. Srivastava, and J. C. W. Lin, "Reinforcement learning multi-agent system for faults diagnosis of mircoservices in industrial settings," *Computer Communications*, vol. 177, pp. 213–219, 2021.

[8] Z. Bouattou, H. Belbachir, and R. Laurini, "Multi-agent system approach for improved real-time visual summaries of geographical data streams," *International Journal of Intelligent Systems Technologies and Applications*, vol. 17, no. 3, pp. 255–271, 2018.

[9] Z. Yahouni, A. Ladj, F. Belkadi, O. Meski, and M. Ritou, "A smart reporting framework as an application of multi-agent system in machining industry," *International Journal of Computer Integrated Manufacturing*, vol. 34, no. 5, pp. 470–486, 2021.

[10] Z. Ma, M. J. Schultz, K. Christensen, M. Værbak, Y. Demazeau, and B. N. Jørgensen, "The application of ontologies in multi-agent systems in the energy sector: a scoping review," *Energies*, vol. 12, no. 16, 2019.

[11] K. Tazi, F. M. Abbou, and F. Abdi, "Multi-agent system for microgrids: design, optimization and performance," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 1233–1292, 2020.

[12] B. Teixeira, T. Pinto, F. Silva, G. Santos, I. Praça, and Z. Vale, "Multi-agent decision support tool to enable interoperability among heterogeneous energy systems," *Applied Sciences*, vol. 8, no. 3, 2018.

[13] E. Serrano and J. Bajo, "Discovering hidden mental states in open multi-agent systems by leveraging multi-protocol regularities with machine learning," *Sensors*, vol. 20, no. 18, p. 5198, 2020.

[14] S. Howell, Y. Rezgui, J. L. Hippolyte, B. Jayan, and H. Li, "Towards the next generation of smart grids: semantic and holonic multi-agent management of distributed energy resources," *Renewable and Sustainable Energy Reviews*, vol. 77, pp. 193–214, 2017.

[15] R. Kamdar, P. Paliwal, and Y. Kumar, "A state of art review on various aspects of multi-agent system," *Journal of Circuits, Systems and Computers*, vol. 27, no. 11, 2018.

[16] B. Okreša Đurić, J. Rincon, C. Carrascosa, M. Schatten, and V. Julian, "MAMbO5: a new ontology approach for modelling and managing intelligent virtual environments based on multi-agent systems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 9, pp. 3629–3641, 2019.

[17] V. Mascardi, D. Weyns, A. Ricci et al., "Engineering multi-agent systems," *ACM SIGSOFT Software Engineering Notes*, vol. 44, no. 1, pp. 18–28, 2019.

[18] L. S. Melo, R. F. Sampaio, R. P. S. Leão, G. C. Barroso, and J. R. Bezerra, "Python-based multi-agent platform for application on power grids," *International Transactions on Electrical Energy Systems*, vol. 29, no. 6, 2019.

[19] W. Du and S. Ding, "A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications," *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3215–3238, 2021.

[20] B. S. Sami, "Intelligent energy management for off-grid renewable hybrid system using multi-agent approach," *IEEE Access*, vol. 8, pp. 8681–8696, 2020.

[21] S. H. Choi, M. Kim, and J. Y. Lee, "Situation-dependent remote AR collaborations: image-based collaboration using a 3D perspective map and live video-based collaboration with a synchronized VR mode," *Computers in Industry*, vol. 101, pp. 51–66, 2018.

[22] S. Mariani, A. Omicini, R. Calegari, and E. Denti, "Logic programming as a service in multi-agent systems for the Internet of Things," *International Journal of Grid and Utility Computing*, vol. 10, no. 4, pp. 344–360, 2019.

[23] X. Zhang, S. Tang, X. Liu, R. Malekian, and Z. Li, "A novel multi-agent-based collaborative virtual manufacturing environment integrated with edge computing technique," *Energies*, vol. 12, no. 14, 2019.

[24] S. Munawar, S. Khalil Toor, M. Aslam, and E. Aimeur, "Pacaits: a multi-agent system for intelligent virtual laboratory courses," *Applied Sciences*, vol. 9, no. 23, p. 5084, 2019.

[25] E. Amador-Domínguez, E. Serrano, and D. Manrique, "A hierarchical multi-agent architecture based on virtual identities to explain black-box personalization policies," *Expert Systems with Applications*, vol. 186, 2021.