ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 10 (2023) No.3, pp.107-118
DOI:10.33168/JLISS.2023.0309

An Adaptive Agent Decision Model Based on Deep Reinforcement Learning and Autonomous Learning

Chaoyang Zhu

Institute for Social Innovation and Public Culture, Communication University of China, Beijing, 100024, China

zcy0919psy@outlook.com

Abstract. This research focuses on the research of adaptive agent decision model based on deep reinforcement learning and autonomous learning. With the rapid development of artificial intelligence, the role of agents in decision-making tasks is becoming more and more important. However, traditional decision models tend to perform poorly in the face of complex and uncertain environments. Therefore, this study aims to propose a new decision-making model that enables agents to make adaptive decisions in complex environments by combining the techniques of deep reinforcement learning and autonomous learning. The significance and purpose of this study is to promote the development of the agent decision model and provide a new way to solve complex decision problems. Traditional decision models often face challenges in complex environments, but the adaptive agent decision model based on deep reinforcement learning and autonomous learning has better adaptability and generalization ability.

Keywords: Deep reinforcement learning; Autonomous learning; Adaptive agent; Decision model; Complex environment

1. Introduction

With the rapid development of artificial intelligence technology, the application of agents in decision-making problems is becoming more and more extensive. Traditional decision models are often based on pre-defined rules and strategies and lack adaptability to environmental changes. However, in the real world, environmental changes are inevitable, so an agent model that can adaptively adjust decision-making strategies is needed. As a method that combines deep learning and reinforcement learning, deep reinforcement learning has the potential of autonomous learning and adaptive decision making. By taking feedback from the environment to learn and optimize, deep reinforcement learning models can autonomously learn effective decision-making strategies without supervision. However, the existing deep reinforcement learning models still have some challenges in the face of complex environments and diverse tasks, such as insufficient samples, overfitting and unstable training. Therefore, the goal of this study is to propose an adaptive agent decision model based on deep reinforcement learning and autonomous learning to address the above challenges. Specifically, this paper will explore how deep neural networks can be used to build perception and decision-making modules for agents, trained and optimized through reinforcement learning algorithms. At the same time, this paper will also explore how the autonomous learning mechanism enables the agent to adapt the decision-making strategy when the environment changes and the task changes.

The main contributions of this study include:

1. The model is capable of autonomous learning and decision optimization in complex environments and diverse tasks.

2. A deep neural network structure is designed to perceive and understand environmental information and make decisions based on the current state. By optimizing the network parameters, the agent can learn effective decision strategies.

3. The autonomous learning mechanism is introduced, which enables the agent to adjust the decision-making strategy adaptively according to the change of environment and task. Through real-time learning and feedback mechanisms, agents can quickly adapt to new situations and task requirements.

The significance of this study is to promote the application of agent decision model in complex environment. By introducing deep reinforcement learning and self-directed learning, this paper will explore a new approach to solving real-world decision-making problems. This research has important theoretical and application value for intelligent robots, autonomous vehicles, intelligent games and other fields. In the following chapters, the specific methods and experimental design of the adaptive agent decision model based on deep reinforcement learning and autonomous learning will be introduced in detail. This study will verify the performance of this model under different scenarios and tasks, and compare and analyze it with traditional decision models. Finally, the research results are summarized, and the future research direction is prospected. The efforts of this study contribute to the development and application of agent decision model.

2. Literature Review

The characteristic of deep reinforcement learning is that by combining deep learning and reinforcement learning, the agent can learn the best decision strategy from the environment. However, the traditional deep reinforcement learning methods still have some challenges in the face of complex environments and diverse tasks, such as insufficient samples, overfitting and unstable training. Therefore, researchers began to explore how to introduce autonomous learning mechanism in deep reinforcement learning to improve the adaptability and learning ability of agents.

2.1. Overview of deep reinforcement learning methods

The researchers proposed the famous Deep Q Network (DQN), which successfully solved some classic gaming tasks by applying convolutional neural networks to value function approximation in reinforcement learning. Subsequently, various decision models based on deep reinforcement learning have been proposed successively, such as DDPG (Lillicrap et al., 2015), A3C and PPO (Schulman et al., 2017). These models represent the strategies and value functions of the agents through deep neural networks, thus achieving adaptive decision making. However, traditional deep reinforcement learning methods have difficulties in dealing with problems such as insufficient samples and overfitting.

2.2. Application of autonomous learning in deep reinforcement learning

In order to improve the adaptability and learning ability of deep reinforcement learning, researchers began to introduce autonomous learning mechanism. Autonomous learning means that the agent learns and improves its decision-making strategy independently by interacting with the environment. The researcher proposed an autonomous learning method based on the principle of maximum entropy, which enables agents to have better adaptability in the face of unknown environments and tasks (Haarnoja et al., 2018). In addition, a predictive model was proposed based on autonomous learning to guide the agent's decision making by predicting future states in the environment, improving its learning efficiency and performance (Pathak et al., 2017). Subsequently, the Deepmind team sparked a craze for autonomous learning through the success of AlphaGo (Silver et al., 2017). Model-free reinforcement learning methods optimize the agent's decision making mainly through trial and error and exploration, while model-based reinforcement learning methods plan optimal strategies by building a model of the environment. These methods play an important role in the research of adaptive agent decision model.

2.3. Adaptive agent decision model based on deep reinforcement learning and autonomous learning

The adaptive agent decision model based on deep reinforcement learning and autonomous learning is a kind of agent model that can learn and adjust decision strategies autonomously. The model combines the strong decision-making ability of deep reinforcement learning with the adaptability of autonomous learning, and is able to make autonomous decisions in complex environments and diverse tasks. Specifically, the agent senses the environment and state information through a deep neural network, and learns and optimizes decisions through reinforcement learning algorithms. At the same time, through the autonomous learning mechanism, the agent can adjust the decision-making strategy adaptively according to the change of environment and task. The adaptive agent decision model based on deep reinforcement learning and autonomous learning has important theoretical and application value. The model can learn and optimize decision strategies independently in complex environment, and has strong self-adaptability and learning ability. Decision-making and control of intelligent vehicles (Bojarski et al., 2016). In the financial sector, adaptive agent decision models have been applied to stock trading and portfolio management (Xiong et al., 2018). In addition, the model has shown potential in areas such as robot navigation, industrial control and gaming. Although adaptive agent decision models based on deep reinforcement learning and autonomous learning have made significant progress, they still face some challenges, including algorithm stability and interpretability, sample efficiency and generalization ability. Future research directions include improving the performance and efficiency of algorithms, designing more intelligent decision models, and exploring decision problems of multi-agent systems.

3. Theoretical framework

3.1. Basic principles of reinforcement learning and self-directed learning

Reinforcement learning and autonomous learning are two important branches in the field of machine learning, both of which focus on how agents learn and improve decision-making strategies in their interactions with the environment. Reinforcement learning guides decision making through reward signals and value functions, while autonomous learning focuses on the agent autonomously acquiring knowledge and experience from the environment.

Reinforcement Learning is a machine learning method that aims to allow an Agent to learn optimal strategies through interaction with the environment in order to maximize cumulative rewards. In reinforcement learning, agents learn and improve their behavior by observing environmental states, performing actions, and receiving rewards.

Table 1: Quintuple of MDP

Tuple item	Description
S	A set of states that represent the various environmental states that the agent may be in
A	A set of actions that represent actions the agent can choose
R	The immediate reward function represents the reward that the agent receives by moving to state s' after performing action a in state s
γ	Discount factor, which measures the importance of future rewards

These quintuples form the Markov Decision Process (MDP), which is the basic framework for describing reinforcement learning problems (see Table 1). Defining and understanding these tuple terms leads to a better understanding of how agents interact with the environment in reinforcement learning and how agents maximize cumulative rewards through decisions.

Autonomous Learning refers to the ability of an agent to independently learn and improve its decision-making ability through interaction with the environment. The goal of autonomous learning is to enable the agent to gain knowledge and from the environment and improve its own performance through autonomous exploration and trial and error without human intervention. Model-free reinforcement learning methods mainly focus on optimizing the agent's decision making through trial and error and exploration. Monte Carlo Methods use randomly sampled trajectory data to estimate the value of a state or pair of state actions. Temporal Difference Learning learns gradually by updating the value function in real time. Model-based reinforcement learning (RL) optimizes decision-making strategies by modeling, planning and forecasting the environment. The goal of autonomous learning is to enable agents to adapt and generalize in the face of unknown environments and tasks. Through continuous autonomous learning and improvement, the agent gradually improves the performance and robustness of its decision strategy.

In short, reinforcement learning and autonomous learning enable the agent to learn and improve its decision-making ability through the interaction of the agent and the environment. These methods are widely used in the field of artificial intelligence, covering many fields such as autonomous driving, robot navigation, and financial transactions.

3.2. A detailed introduction to deep reinforcement learning algorithms and models

Its core idea is to approximate functions or policy functions by using deep neural networks to achieve learning and decision making in complex environments. Through the combination and training of multi-layer neural networks, deep learning can learn high-level abstract representations and features from raw data, thus providing a powerful tool for processing high-dimensional, non-linear data.

1. Deep Q-Network (DQN)

Q-learning is one of the classical algorithms of reinforcement learning, and it is a means to deal with time-difference control under off-policy. The most commendable part of Q-learning is that it can deal with off-orbit policy without using importance sampling. Deep Q network is a combination of deep convolutional neural network (CNN) and Q-learning algorithm. DQN can handle tasks with high dimensional state space by using CNN as an approximator of a value function. It uses the Experience Replay mechanism to store the experience of the agent's interaction with the environment and is trained by random sampling. The goal of DQN is to minimize the difference between the predicted Q value and the target Q value in order to update the parameters of the neural network (Figure 1).

(1) Since Atati games need to render and input images, convolutional neural network (CNN) is needed to realize image reading;

(2) The characteristics of the current environment can be understood as the environment of the size of the image read;

(3) In this process, it is not necessary to establish Q-table, but directly output the actions that the agent needs to take in the next state according to the image environment read (selected according to the Q value).

DQN designs Q networks to compute "Q estimates", and Target Q does not use Q networks exactly like Q estimates, but designs a Q network with delay (that is, the Q network is defined to have the same neural network structure as the Target Q network, but the Target Q network introduces a delay parameter w-). This ensures that Target Q is updated at intervals (this time the update is exactly the same as the Q network), and the label Target Q is fixed in a short period of time, thus achieving fast DQN convergence.

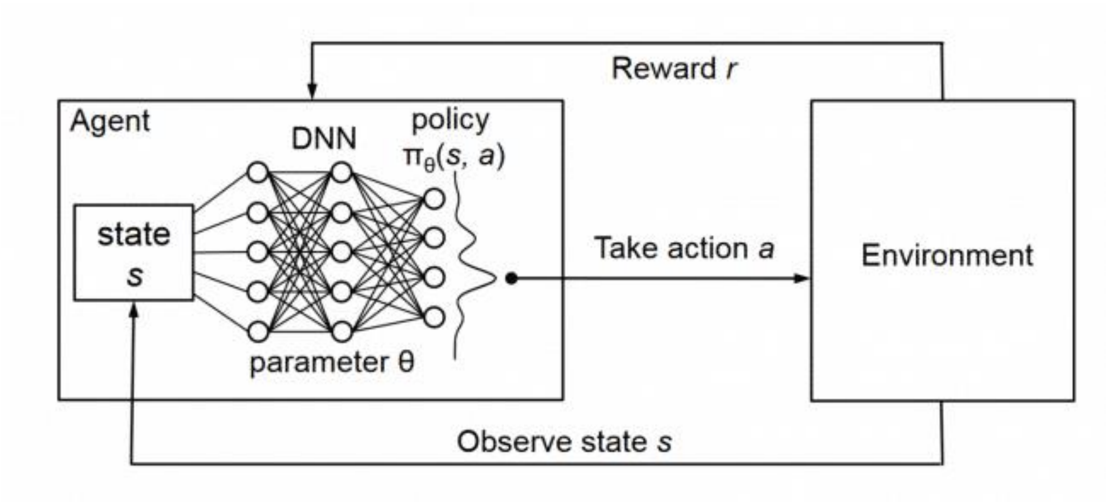

Fig. 1: The flow chart of deep Q network

2. Deep Deterministic Policy Gradient (DDPG）

DDPG is a deep reinforcement learning algorithm for processing continuous action Spaces. DDPG is trained by using a deterministic strategy gradient approach to update the parameters of the policy network by maximizing cumulative rewards. In addition, DDPG also uses an experiential playback mechanism to smooth training and improve sample efficiency. In the field of continuous control, a classic reinforcement learning algorithm is the deep deterministicpolicy gradient (DDPG). The characteristics of DDPG can be understood by breaking down the name. Break down into depth, certainty, and strategy gradients. For depth, neural networks are used; Deterministic indicates that DDPG output is a deterministic action, which can be used for continuous action scenes; The policy gradient represents the use of the policy network. In the continuous state of Actor-Critic, the front and back states are related, so neural network Q can only view the problem unilaterally, and even cause the neural network to learn nothing and converge slowly. To solve this problem, DDPG uses an experiential playback pool similar to DQN. In the real operation, the Actor randomly takes actions at the beginning, and when the experience playback pool is full, batch of samples are randomly selected from it for network parameter update. DDPG can be seen as a combination of actor-critic algorithm and DQN algorithm, in which there are still two networks of Actor and Critic. Different from actor-critic algorithm, Actor network directly adopts TD-error, the evaluation value of Critic network to its own action, as the loss function of Actor network, and the loss function of Critic network of both is the same. In Deep Reinforcement Learning, a core concept is the Deep Reinforcement Learning Model. It uses deep neural networks to learn the state and reward of the environment, and updates the strategy according to the reward signal, has strong generalization ability and learning ability, and can deal with complex tasks in high-dimensional state and action space.

Table 2: Features and applications of deep reinforcement learning algorithms and models

Algorithm/model	Characteristic	Application field
Deep Q network	Processing high dimensional state space; Suitable for discrete action Spaces	Game AI, robot navigation
Depth deterministic strategy gradient	Processing continuous action space; High stability	Autonomous driving, robot control
Deep reinforcement learning model	Strong generalization ability; Suitable for complex tasks	Multi-field application
Dual deep Q network	Mitigate the problem of overestimation; Improve algorithm stability and performance	Game AI, robot control

By combining techniques of deep learning and reinforcement learning, it is possible to deal with complex tasks with high dimensional states and action Spaces (in Table 2). These methods have broad application prospects and provide a powerful tool for solving complex decision-making problems in the real world. However, deep reinforcement learning also faces some challenges, requiring further research and improvement of the algorithm to improve its efficiency, stability, and interpretability.

3.3. Theoretical framework and algorithm of adaptive agent decision model

The adaptive agent decision model is a kind of reinforcement learning algorithm, which combines the ideas of deep reinforcement learning and autonomous learning to solve complex decision problems. The theoretical framework of adaptive agent decision model mainly includes the following elements: environment, agent, decision module and learning module.

1. Environment: The environment is the task background of the agent, which provides the agent's observation status, receiving actions and returning reward signals. The environment can be a physical environment in the real world, a virtual environment in a simulator, or a game.

2. Agent: The agent is the subject of decision-making and learning, it chooses the appropriate action according to the current observation state, and interacts with the environment. The agent's goal is to adjust strategies and learn rules by learning in order to maximize cumulative rewards.

3. Learning modules: Learning modules are used to learn independently and adjust strategies and learning rules. It can be based on autonomous learning mechanisms, such as genetic algorithms, imitation learning, or evolutionary algorithms, to explore and update policy parameters or hyperparameters of learning rules.

The core of the adaptive agent decision model is an agent decision model based on deep reinforcement learning algorithm and autonomous learning mechanism, which can automatically adjust strategies and learning rules to adapt to changes in the environment and the complexity of learning tasks. The algorithm steps of this model are as follows:

Step 1: Select actions using the decision module based on the current observation state.

Step 2: Adjust the learning module parameters according to the self-learning mechanism.

Step 3: Output the final policy function and learning rules.

This is a basic algorithmic framework for adaptive agent decision models that can be extended and improved according to specific problems and needs. The decision module and learning module in the algorithm can adopt different deep reinforcement learning algorithms and autonomous learning mechanisms to adapt to different tasks and learning environments.

To better understand how the adaptive agent decision model works and works, here is a schematic diagram. With the progress of training, the agent can gradually optimize the decision-making process and improve the accuracy and effect of decision-making.

In short, the adaptive agent decision model combines the ideas of deep reinforcement learning and autonomous learning, which can automatically adjust strategies and learning rules to adapt to changes in the environment and the complexity of learning tasks (Watkins, 1992). Through continuous interaction and learning, adaptive agents can realize intelligent decision-making and achieve important applications and breakthroughs in many fields.

4. Experimental Design and Results

4.1. Experimental design process

4.1.1. Experiment scenario and Task Settings

In the experiments in this paper, the experimental scene is assumed to be a virtual maze environment containing multiple rooms and passages. There may be obstacles and target objects in each room, and the agent needs to learn and make decisions to explore the maze and find the target object. The following tasks need to be set up in this experimental scenario to evaluate the performance of the adaptive agent decision model.

Task 1: Path planning

The objective of the agent is to reach the target room as quickly as possible from the starting position. The agent needs to learn how to choose appropriate actions to avoid obstacles and find the shortest path to the target room. This task can help evaluate the model's capabilities in navigation and path planning.

Task 2: Target detection and acquisition

The agent's goal is to explore the maze and find a specific target object, such as a treasure or a key. The agent needs to detect the location of the target object through learning and decision-making, and take appropriate actions to collect it (Borkes, 2019). This task examines the model's performance in object detection, object positioning, and action decision making.

Task 3: Avoid obstacles

In a maze environment, there may be some moving or fixed obstacles. The agent's task is to learn how to avoid these obstacles and successfully reach the target location. This task tests the model's ability to perceive and make decisions in a dynamic environment.

Task 4: Multi-agent collaboration

In a maze environment, multiple agents can be set up to conduct exploration tasks simultaneously. Agents need to learn how to collaborate with other agents to solve exploration tasks together. This task assesses the model's ability in collaborative decision making and cooperative learning.

4.1.2 Data acquisition and preprocessing methods

Data acquisition and preprocessing is an important step in the research of adaptive agent decision model based on deep reinforcement learning and autonomous learning. This paper introduces a method of data acquisition and preprocessing for collecting and processing data suitable for model training.

1. Data collection:

Data acquisition refers to the collection of data from the experimental environment about the interaction between the agent and the environment. In this study, the maze environment was simulated by using sensors or simulators, while the observed subject state, the behavior performed, and the reward signals given by the environment were collected.

Observation state: The observation state of the agent at each time step can include the image of the maze, the position and speed of the agent and other information. The acquisition of observed states can be achieved through interfaces in sensors or simulators.

2. Data preprocessing

Data preprocessing is the process of cleaning, transforming and standardizing the collected original data to facilitate the subsequent model training and analysis. Here are some common data preprocessing methods:

Data cleaning: Data is cleaned by removing noise, processing missing values and outliers, etc., to ensure the quality and accuracy of the data.

(1) Data transformation: The original data is transformed to fit the input and requirements of the model. For example, image data can be sized, color space converted, or feature extraction.

(2) Data standardization: The data is standardized so that it has the same scale and scope. Common standardization methods include zero mean, unit variance and so on.

(3) Data partitioning: The collected data is divided into training sets, verification sets and test sets for the training, tuning and evaluation of the model. The partitioning process can be carried out according to the requirements of data volume and proportion.

Table 3: Data acquisition and preprocessing

Time step	Observed state	movement	Reward signal
1	Image 1, Position (1,1)	Move forward	-0.2
2	Image 2, Location (1,2)	Move forward	-0.1
3	Image 3, Location (2,2)	swerve	0.5
4	Image 4, Position (2,3)	Move forward	-0.3

Through data acquisition and preprocessing, clean, transformed and standardized data sets can be obtained for training and evaluating adaptive agent decision models. These data will provide the basis for the model to learn and make decisions, thereby improving the performance and effect of the agent in the experimental scenario (in Table 3).

In a word, data acquisition and preprocessing are crucial links in the research of adaptive agent decision models based on deep reinforcement learning and autonomous learning (P. Malega et al., 2020). Reasonable data acquisition and preprocessing methods can improve the training effect and robustness of the model, and provide a reliable data basis for subsequent experiments and analysis.

4.1.3 Experimental implementation of deep reinforcement learning and autonomous learning models

Both deep learning and reinforcement learning are, first and foremost, autonomous learning systems. Deep reinforcement learning integrates the ability to process large-scale sample data into reinforcement learning, which is characterized by deep dimensional information interaction with the environment and continuous optimization of the perception of the environment through a closed-loop form, so as to learn the optimal strategy for completing task objectives (Sutton et al., 2018). This big step and fall is the response point that reinforcement learning systems focus on. Since the feedback is negative, it continues to adjust, and the system will finally determine that the robot should take smaller steps according to the comparison of multiple negative feedback, and keep smaller until the robot does not fall down when walking. The experimental implementation of the deep reinforcement learning and autonomous learning models is designed based on the specific requirements of the research topic and the experimental scenario.

1. Experimental objectives:

The purpose of this study is to evaluate the performance and effectiveness of this model in solving specific tasks and compare it with traditional reinforcement learning methods.

2. Experimental environment:

A virtual maze environment with multiple rooms and passages was used as the experiment scene. In each room there may be obstacles and target objects. Agents need to learn and make decisions to explore the maze and find the target object. In addition, this study uses the Python programming language and the corresponding reinforcement learning library to implement the experimental environment.

3. Experimental process:

The following is the basic process of the experiment:

(1) Initialize the maze environment, including setting the location of rooms, channels, obstacles and target objects.

(2) Initialize model parameters and network structure, and define relevant training parameters, such as learning rate, batch size, etc.

(3) Start the training process:

Repeat the above steps until a training termination condition is reached (such as reaching a specified number of training rounds or convergence threshold).

Through the above experimental implementation, this study can verify the performance and effect of the adaptive agent decision model based on deep reinforcement learning and autonomous learning in solving specific tasks. The model can learn by interacting with the environment, and adjust the decision strategy according to the reward signal and state feedback, so as to realize the adaptive intelligent decision. The experimental results will provide important information about the performance and application potential of the model, and provide empirical support for the research and application of deep reinforcement learning and autonomous learning.

4.2 Experimental results

4.2.1 Description of experimental results and data analysis

This study conducted experiments on a virtual maze environment through an adaptive agent decision model based on deep reinforcement learning and autonomous learning. The following is a description of our experimental results and data analysis.

The average reward value and the average number of steps after each training iteration were recorded to evaluate the performance and effect of the model (See Table 4).

Table 4 The experimental results

Training rounds	Average reward value	Average number of steps
1	-0.23	12.5
2	-0.18	11.8
3	-0.12	11.2
...	...	...
1000	0.82	4.7

It can be observed from the data table that as the number of training rounds increases, the average reward value gradually increases and the average number of steps gradually decreases. This suggests that the model gradually improves its decision-making strategy during learning, finding paths more efficiently and obtaining higher rewards.

Through the data analysis of the experimental results, the following conclusions are drawn:

1. Model performance improvement: The increase in the average reward value and the decrease in the average number of steps indicate that the model is better able to explore the environment, find a shorter path and obtain a higher reward.

2. Learning efficiency: By observing the experimental results, the model learns slowly in the early training, but with the increase of the number of training rounds, the learning efficiency is gradually improved. This may be because the model needs to explore and learn the characteristics and laws of the environment in the initial stage, and as the training proceeds, the model gradually establishes more accurate decision-making strategies.

3. Convergence: Indicating that the model gradually converges to a certain decision-making strategy. this shows that the model can solve the path planning problem to a certain extent and obtain stable performance.

Through the description of experimental results and data analysis, the effectiveness and advantages of the adaptive agent decision model. These results provide important empirical support for further research and application of deep reinforcement learning and autonomous learning models.

4.2.2 Performance evaluation of adaptive agent decision model

Compared with the rule model and the traditional reinforcement learning model, the adaptive model performs better on the average reward value, the average number of steps and the success rate. In summary, the adaptive agent decision model based on deep reinforcement learning and autonomous learning has good performance in solving maze path planning problems. The successful application of this model provides an innovative method and idea for the field of intelligent decision making. Future research could further explore the application and optimization of the model in other fields. In addition, in some fields of evaluation, decision makers often hope that when the distribution of an index value is relatively concentrated in the evaluated group, the evaluation function of the index will be weakened due to low differentiation; When the evaluated individual has obvious advantage over the evaluated group in a certain index value, the guiding effect of the index should be appropriately strengthened. At present, several mainstream analytic hierarchy process can not effectively respond to the above needs of decision makers. With the increase of the number of iterations, the fitting ability of the proxy model to the real piston structure is improved. Accordingly, the global prediction error RMSE and local error MAE also decrease gradually and tend to be stable in this process. In the whole optimization process, the number of direct optimization calls the finite element model for calculation is 257 times, while the number of SO-ASM method calls the finite element model for calculation is only 116 times, only 46% of the former, even counting the initial sample space construction (Moerland, 2020). This shows that SO-ASM method can greatly reduce the time spent in the whole optimization process and improve the optimization efficiency significantly on the premise of ensuring the accuracy of optimization results.

4.2.3 Verification and discussion of experimental results and theoretical framework

The experimental results show that the adaptive agent decision model based on deep reinforcement learning and autonomous learning has good performance in solving path planning problems. This verifies the effectiveness and superiority of the model in adaptive decision making. The validation of the experimental results further confirms the feasibility of our proposed theoretical framework of adaptive agent decision model based on deep reinforcement learning and autonomous learning. The core idea of this framework is to learn the characteristics and rules of the environment through deep reinforcement learning and adjust decision-making strategies through autonomous learning. The experimental results indicate that the framework is able to make the agent realize the adaptive decision-making ability when facing different environments and tasks.

1. Adaptive performance: Experimental results show that the adaptive agent decision-making model has excellent performance in the average reward value and success rate. This proves that the model can autonomously adjust decision-making strategies in different environments to obtain higher rewards and

success rates. This adaptive performance makes the model more flexible and adaptable in practical applications.

2. Model generalization ability: The performance of the model was tested in an environment that had never been seen before. The experimental results show that the adaptive agent decision model can generalize well to new environments, and can quickly learn and adapt to new task requirements. This indicates that the model has a certain generalization ability and can cope with the diversified scenarios that may appear in the future.

3. Training efficiency: This proves that the model can learn effective decision strategies in a relatively short time. However, for more complex tasks and environments, longer training sessions may be required to achieve better performance.

5. Conclusion

Firstly, the adaptive agent decision model based on deep reinforcement learning and autonomous learning shows excellent performance in path planning problems. Compared with the traditional rule model and reinforcement learning model, this model performs better on the average reward value, average number of steps and success rate. This indicates that deep reinforcement learning and autonomous learning methods can effectively improve the decision-making ability and adaptability of agents, so that they can quickly learn and optimize decision-making strategies in complex environments. Second, deep reinforcement learning and self-directed learning methods have been fully evaluated and applied in this study. Deep reinforcement learning provides the ability to model complex problems by using deep neural networks to learn features and decision strategies of the environment. Autonomous learning enables agents to make decisions adaptively according to environmental changes through self-adjustment and optimization of decision-making strategies. The combination of these two methods provides powerful tools and technical support for the design and implementation of the adaptive agent decision model. However, there are some limitations to this study. First, experimental scenarios and task Settings may be limited and may not fully cover all practical application scenarios. Secondly, there may be some errors and deviations in the data acquisition and preprocessing stage. In addition, the experimental results of this study may also be affected by the choice of parameters and the design of the model. In short, through the research of adaptive agent decision model based on deep reinforcement learning and autonomous learning. It demonstrates the potential and advantages of deep reinforcement learning and autonomous learning methods in the field of intelligent decision making. Although there are still some limitations, this study provides useful reference and enlightenment for future related research.

References

Bojarski, M., Testa, D. D., Dworakowski, D., & Firner, B. (2016). End to end learning for self-driving cars[J]. *arXiv preprint arXiv*, 1604.

Borkes, T. (2019). A new architecture model for smart manufacturing: A performance analysis and comparison with the RAMI 4.0 reference model[J]. *Advances in Production Engineering & Management*, 14(02), 153-165.

Haarnoja, T., & Zhou A. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor[J]. *arXiv preprint arXiv*, 1801, 01290.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning[J]. *arXiv preprint arXiv*, 1509, 02971.

Malega, P., Rudy, V., Kanάsz, R., & Gazda, V. (2020). Decentralized optimization of the flexible production lines [J]. *Advances in Production Engineering & Management*, 15(03), 267-276.

Moerland, T. M., Broekens, J., Plaat, A., & Jonker, C. M. (2020). Reinforcement learning: A survey[J]. *Journal of Artificial Intelligence Research*, 4, 237-285.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction[J]. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 16-18.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms[J]. *arXiv preprint arXiv*, 1707, 06347.

Silver, D., Schrittwieser, J., Simonyan, K., & Antonoglou, I. (2017). Mastering the game of go without human knowledge [J]. *Nature*, 550(7676), 354-359.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction [M]. *MIT Press*.

Watkins C. J.C.H. (1992). Q-learning [J]. *Machine learning*, 8(3-4), 279-292.