Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 15, No. 3, 2024

Application of Reinforcement Learning in Complex Environmental Decision-making Problems

Chenchen Zhao*

Boston University, Boston, Massachusetts, 02215, United States
* **Corresponding author Email:** zhao900805@gmail.com

Abstract: This paper provides an in-depth discussion on the application of reinforcement learning in decision-making problems in complex environments, focusing on the challenges it faces such as data sparsity and the balance between exploration and utilization, and proposes corresponding solutions. The study realizes an integrated process from data collection to decision support by developing a decision support system for sustainable enterprise management based on big data. Meanwhile, the exploration strategies and optimization algorithms in reinforcement learning are used to improve decision-making efficiency and accuracy. An application case in an intelligent transportation system demonstrates the potential of the multi-intelligence reinforcement learning framework in vehicle collaboration and optimization of traffic flow, effectively alleviating traffic congestion. Experiments show that the system significantly improves decision-making efficiency and accuracy, verifying the effectiveness of reinforcement learning in decision-making in complex environments.

Keywords: Reinforcement learning; Complex environment decision-making; Multi-intelligence reinforcement learning; Intelligent transportation system.

1. Introduction

Reinforcement learning, as a core branch of machine learning, shows great potential for solving complex decision-making problems by maximizing the cumulative rewards through trial-and-error of intelligences in the environment, especially when dealing with high-dimensional states and dynamic environments. Since Sutton and Barto laid the theoretical foundation, although reinforcement learning has achieved remarkable results in many fields, it still faces challenges such as data sparsity, balance between exploration and utilization, and the generalization ability and computational efficiency need to be further improved in practical applications [1]. This study deeply explores the application of reinforcement learning in decision-making in complex environments, focusing on environment modeling, state representation, exploration strategy and optimization algorithm innovation, aiming at constructing a reinforcement learning framework applicable to complex environments, and verifying its effectiveness through experiments to provide support for theory and practice. Meanwhile, for the Q-learning dimension disaster in path planning, an improved algorithm based on deep Q network (DQN) is proposed, which adopts end-to-end learning and combines deep convolutional neural network feature extraction with reinforcement learning decision-making ability, with a view to improving the feasibility and generalization ability of the algorithm, and to promote the application of reinforcement learning in a wider range of fields.

2. Theoretical foundations of reinforcement learning

2.1. Fundamentals of Reinforcement Learning

Reinforcement learning, as a strategic decision-making method, has its basic principles rooted in early psychological theories, aiming to simulate and imitate biological behavior. Under this framework, an intelligent body detects the environment based on its own conditioned responses, obtains the feedback value of the conditioned reflexes by simulating multiple behavioral strategies, and determines the optimal strategy through conditional filtering (see Figure 1). The reinforcement learning system consists of two elements: the environment and the intelligent body, and contains four core parameters: the strategy π, the value function, the reward signal R, and the environment model. The policyπ is concerned with the effectiveness of the intelligences' behavior and system performance; the value function reflects the cumulative weighted compensation expectation of the intelligences from the state to the goal under the policy; the reward signal R evaluates the value of the policy and guides the selection; and the environment model is used to simulate and predict the external state. Reinforcement learning uniquely balances the analysis of outcome and process data in the search for control strategies that maximize the performance of intelligences, demonstrating significant advantages and broad application perspectives that indicate its potential in multiple domains and novel algorithm development.

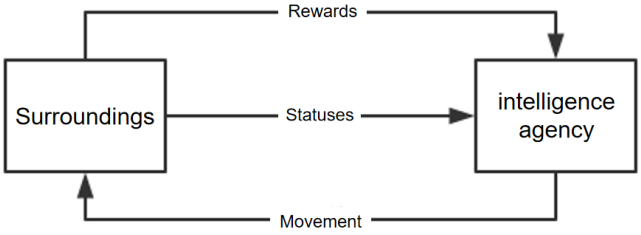


Figure 1. Diagrammatic representation of enhanced learning mechanisms

2.2. Main Algorithms of Reinforcement Learning

Reinforcement learning, as an important field of machine learning, has made remarkable achievements in algorithm research and application in recent years. Among them, Q-

learning, as a classical model-free algorithm, optimizes the strategy by iteratively updating the Q-value, which performs well in solving small-scale discrete problems, but has limitations when facing high-dimensional continuous state space, which promotes the development of Deep Q-Network (DQN) algorithms (see Figure. 2).DQN integrates deep learning, uses neural networks to approximate the Q-function, successfully copes with dimensionality disasters and made breakthroughs in areas such as game AI. Meanwhile, Policy Gradient method, as another important algorithm, directly optimizes the policy to maximize the expected return, in which the REINFORCE algorithm updates the policy parameters by gradient ascent, but suffers from high variance problem. The Proximal Policy Optimization (PPO) algorithm, on the other hand, improves on this foundation by controlling the policy update amplitude to effectively reduce the variance, which improves the stability and convergence speed of the algorithm. These algorithms have excellent performance in real-world applications, showing a wide range of application potential and continuing to promote the progress of the reinforcement learning field.

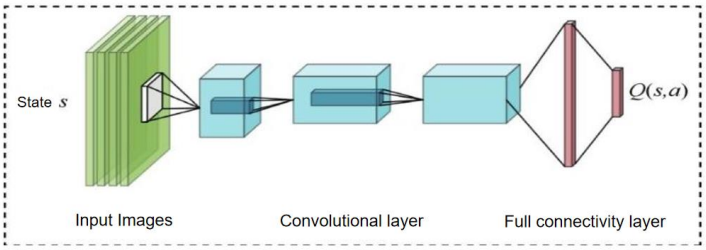


Figure 2. Network structure of DQN

2.3. Extensions and Variants of Reinforcement Learning

Reinforcement learning, as an important branch of machine learning, continues to expand at the theoretical and algorithmic levels, forming a variety of variants including hierarchical reinforcement learning and multi-intelligence body reinforcement learning [2]. Hierarchical reinforcement learning improves learning and decision-making efficiency through task decomposition and decision-making at different levels to cope with complex tasks, as shown in the study of autonomous decision-making for search and rescue robots. Multi-intelligence reinforcement learning, on the other hand, focuses on the interaction of multiple intelligences in a common environment, optimizes individual strategies and achieves inter-intelligence synergy and competition to achieve the global optimum, which provides a new solution for the fields of intelligent transportation and multi-robot collaboration. These extensions and variants have made remarkable achievements in theory and practice, which not only promote the expansion of reinforcement learning to a wider field, but also indicate that it is expected to play an important role in providing new ideas and methods for solving complex decision-making problems in the future.

3. Characteristics and Challenges of Decision Making Problems in Complex Environments

3.1. Definition and Characterization of Complex Environments

In the field of deep reinforcement learning, the decision-making strategies of intelligences in complex environments are crucial. The characteristics of complex environments include high-dimensional state and action spaces, which require the intelligences to recognize the key elements and make choices from the rich information (see Fig. 3). This study utilizes deep learning models, especially convolutional neural networks, to learn control strategies directly from raw sensory inputs, which is effective in the Atari 2600 game. To address partial observability and uncertainty, an empirical playback mechanism is introduced to smooth the training distribution by randomly sampling historical transformations to alleviate the data correlation and non-smoothness problems, while effectively managing the dynamics and non-smoothness of the environment. The intelligences are able to learn from raw video data, adapt to the time-dependence of state transitions and changing transition probabilities, and achieve stable learning. These methods highlight the potential of deep reinforcement learning to deal with complex environments, providing a solid foundation for intelligences to make effective decisions and cope with environmental dynamics in high-dimensional decision spaces.

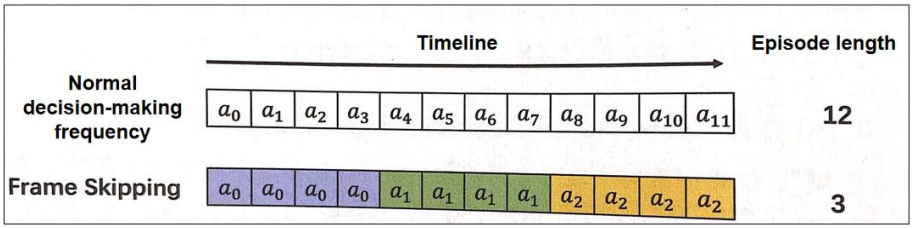


Figure 3. Optimization of intelligent body action space

3.2. Challenges in Complex Environmental Decision-Making Problems

In the study of complex environmental decision-making problems, reinforcement learning models need to find a balance between data sparsity and the exploration-exploitation dilemma, while ensuring that the models have the ability to generalize to unknown situations. Intelligents make

decisions within limited computational resources and tight time frames, which together form the core of the research. Exploratory analysis of data, as a starting point for modeling, requires a deep understanding of the business logic behind the data set. In the case of the Jane Street Market Trading Prediction Competition, the dataset was designed to train models with historical trading and return information to predict future trading returns and maximize profits. Initial analysis reveals extra features in the training set compared to the test set that are closely correlated with returns, indicating the impact of returns on trading decisions over different time horizons. Further correlation analysis (see Table 1) revealed significant correlations between return features and stronger correlations between neighboring columns, providing important data support for the intelligences' decision making in complex environments.

Table 1. Heat map of the correlation between the columns of the rate of return

	resp_1	resp_2	resp_3	resp_4	resp_5
resp_1	1.000000	0.890214	0.678381	0.358949	0.452159
resp_2	0.890214	1.000000	0.823383	0.469230	0.595274
resp_3	0.678381	0.823383	1.000000	0.805952	0.815972
resp_4	0.358949	0.469230	0.805952	1.000000	0.956197
resp_5	0.452159	0.595274	0.815972	0.956197	1.000000

(Source: https://cloud.tencent.com/developer/article/1800714)

This analysis is critical to understanding the relationships between data features and helps guide feature engineering and model optimization (see Figure 4), thus laying the foundation for building effective predictive models [3].

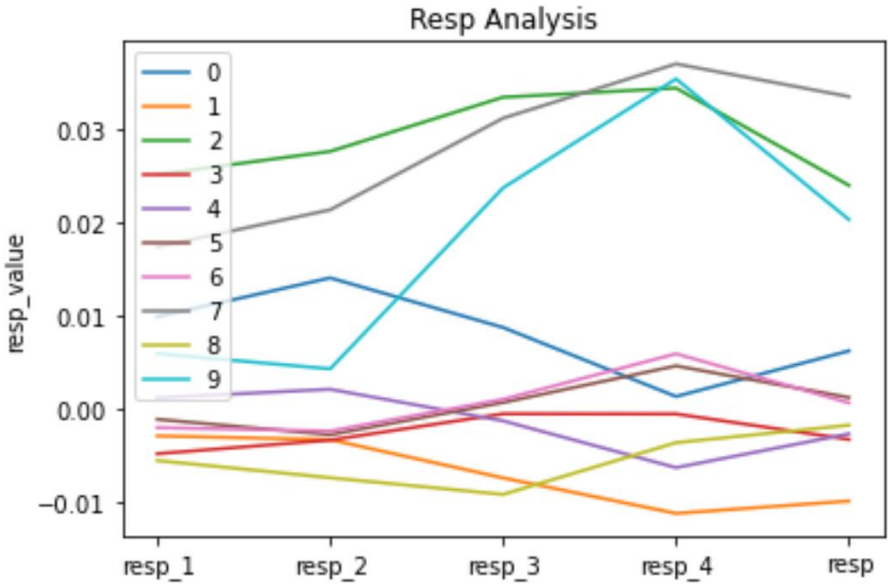


Figure 4. Line graph of return columns

(Data source: https://cloud.tencent.com/developer/article/1800714)

4. Strategies for reinforcement learning in complex environmental decision-making problems

4.1. Environment modeling, exploration strategy and optimization algorithm in reinforcement learning

Environment modeling is the basis for the application of reinforcement learning in complex environments, and its accurate state representation plays a key role in efficient decision-making. State abstraction and feature extraction is the core part of this process, deep learning technology shows significant advantages here, such as convolutional neural network can efficiently process image data and extract key features in automatic driving; recurrent neural network is good at capturing the time change law of environmental state. Meanwhile, exploration strategies and optimization algorithms occupy a central position in reinforcement learning. Curiosity-driven and hierarchical exploration strategies motivate intelligent bodies to explore the unknown and optimize the decision-making process. For specific problems, suitable optimization algorithms should be selected and parameters should be adjusted according to training feedback to improve learning efficiency and decision-making performance. In the training of automatic driving system, reasonable algorithm selection and parameter adjustment can significantly improve the safety and accuracy of driving decisions, which fully demonstrates the potential of reinforcement learning in complex environments.

4.2. Multi-intelligence collaboration and competition in complex environments

The multi-intelligence reinforcement learning framework provides a theoretical foundation for multi-intelligence collaboration and competition in complex environments (see Table 2). In this framework, multiple intelligences learn and optimize strategies together by interacting with each other to

achieve global goals [4]. Collaboration strategies are designed to facilitate information sharing and collaborative work between intelligences to improve the overall system performance.

Table 2. A framework for multi-intelligent body reinforcement learning

Point	AI 1.0	AI 2.0	AI 3.0
Features	Symbolic System	Deep Learning, Reinforcement Learning	Interpretable artificial intelligence (AI)
Protagonist	Simon, Holland.	Hinton, Sutton	Cheung Cymbal, Robertsbury
Ideology	Rationalism	Empiricism	Rationalism + Empiricism
Driver	Knowledge Driven	Data Driven	Knowledge Driven + Data Driven
Rely on sth.	Knowledge + Algorithms + Computing Power	Data + Algorithms + Computing Power	Knowledge + Data + Algorithms + Computing Power
Require a solution	Search for	Study	Search + Study
Arithmetic	DFS, BFS, A*	CNN, RNN, GAN, PPO	
Vantage	Interpretable, Reasonable	End-to-end, Pattern Recognition, Intelligent Decision Making	Robust, Interpretable, Reasonable
Limitations	Incomplete rules, inelegant methods.	Unexplainable, poor robustness	

Reinforcement learning, especially the multi-intelligence body reinforcement learning framework, shows significant application value in intelligent transportation systems. The framework considers each vehicle as an intelligent body and realizes efficient collaboration among vehicles through mutual interaction and information sharing. In order to optimize traffic flow, a collaboration strategy combining a rule-based approach with deep reinforcement learning (deep RL) is designed so that the vehicles can adjust their driving speeds and routes according to the real-time traffic conditions, which can effectively alleviate the congestion and improve the access efficiency. Meanwhile, higher mobility and energy efficiency in hybrid connected transportation is achieved by integrating visual perception methods with vehicle-to-infrastructure (V2I) communication. Hidden features are extracted using a multi-stream neural network and combined with a deep policy network to generate eco-driving behaviors, which further improves the system performance. Unity-based simulator with mixed traffic intersection scenario design and comparison experiments with multiple baseline approaches validate the superiority of the multi-intelligence reinforcement learning framework in reducing energy consumption and saving travel time see (Figure 5). This research provides innovative ideas for the development of intelligent transportation systems and highlights the potential application of multi-intelligent body reinforcement learning in decision-making problems in complex environments.

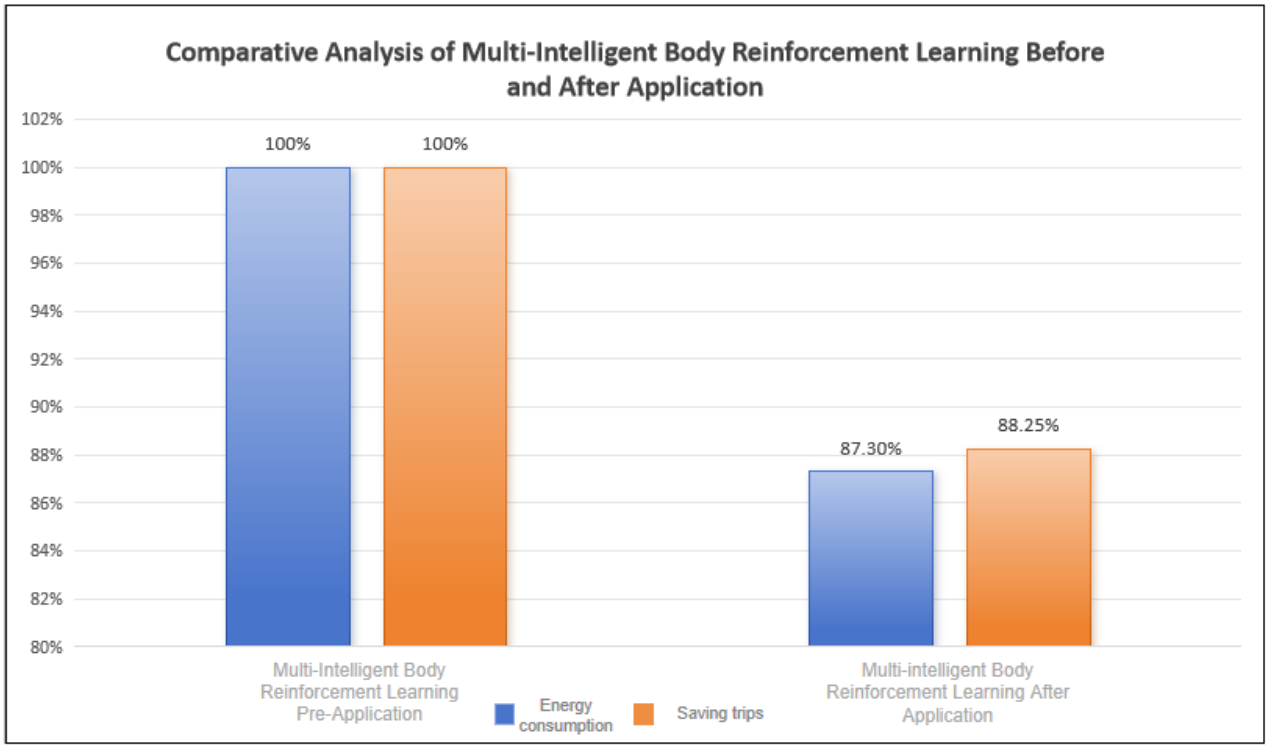


Figure 5. Comparative analysis before and after the application of multi-intelligence body reinforcement learning

5. Conclusion

This study systematically explores the application of reinforcement learning in decision-making in complex environments, and empirically verifies the effectiveness and strong adaptability of its algorithms. The research focuses on

key techniques such as state space representation and reward function design, which lays a theoretical foundation for the successful application of reinforcement learning in complex environments. However, facing the challenges of high-dimensional state space and sparse reward signals, the algorithm suffers from the problems of slow convergence, easy to fall into local optimization, and needs to take into account the safety, stability and interpretability in practical applications. Future research should explore algorithm improvement paths, such as combining deep learning to enhance the feature extraction ability, or using hierarchical reinforcement learning to simplify the complexity of the problem. At the same time, the integration of reinforcement learning with other technologies, such as machine learning and computer vision, is expected to improve the performance of the algorithm and broaden the application areas. Reinforcement learning shows great potential in cutting-edge fields such as intelligent manufacturing, autonomous driving and financial transactions, and its development is promising and full of challenges.

References

[1] Li Hui,Qi Yuming. A deep reinforcement learning-based path planning method for robots in complex environments[J]. Computer Application Research, 2020,37(S1):129-131.

[2] Yin Chenkun,Ji Hongxuan,Zhang Yanxin. Autonomous decision making based on heterogeneous policy hierarchical reinforcement learning for search and rescue robots in complex interactive scenarios[J]. Journal of Beijing Institute of Technology, 2023,49(04):403-414.

[3] ZHANG Wang,SHAO Xuehui,TANG Huilong,et al. A reinforcement learning interference decision-making method for exploring rate adaptive settings[J/OL]. Journal of Military Engineering,1-10[2024-09-03]

.http://kns.cnki.net/kcms/detail/11.2176.tj.20240827.1543.004.html.

[4] Luo Qing,Li Zhijun,Lv Tiansheng. Multi-intelligent reinforcement learning in complex environments [J]. Journal of Shanghai Jiao Tong University, 2002, (03):302-305.