Reinforcement Learning and Deep Reinforcement Learning for Game AI Training: Methods and Applications

Chenyu Shen

Shenzhen University, Shenzhen, Guangdong 518060

Abstract: ***Game AI training represents an interdisciplinary integration of computer science and artificial intelligence, serving as a primary testbed environment within the field of reinforcement learning. At the current stage, game AI training environments present challenges related to both ethical considerations and technical innovation. These challenges primarily center on feedback analysis of coefficients and delays, the high-dimensional state-action space environment, and the characteristics of unstable environments. Building upon recent advances in deep reinforcement learning, we propose a foundational deep reinforcement learning framework incorporating an attention mechanism. This framework is designed to address the problem of collective intelligence in complex environments.***

Keywords: Game AI training; Intensive learning; Deep reinforcement learning.

1. INTRODUCTION

The definition of game AI training is relatively broad, and it can generate an appropriate level of intelligence in the existing work environment, making games more realistic, interesting, and challenging. It can be regarded as game AI. In the entire framework of game AI, there are only two ways: one is AI based on the characteristics of finite state machines or behavior trees, whose form can be predicted and analyzed by relevant technologies, and the other is mainly based on neural networks, which can obtain non qualitative AI through genetic neural networks and data calculation analysis, and whose behavior is difficult to predict and analyze. Zheng and Jiang [1] developed a new methodology for Chinese term extraction from scientific publications, enhancing the precision of domain-specific terminology identification in scholarly literature. Wu et al. [2] addressed structural health monitoring by proposing a small-sample object detection method for surface cracks in concrete structures of high-rise buildings via multi-level transfer learning, overcoming the practical challenge of limited labeled data in critical infrastructure inspection. In computer vision, Peng et al. [3] introduced a dual-augmentor framework for domain generalization in 3D human pose estimation, improving model robustness when deployed to previously unseen environments. Addressing challenges in dynamic data processing, Ukey et al. [4] developed an efficient method for continuous kNN join over dynamic high-dimensional data, providing scalable solutions for real-time similarity search in evolving datasets. Extending the frontier of 3D vision, Peng et al. [5] proposed 3D Vision-Language Gaussian Splatting, a novel approach that bridges vision and language modalities for high-quality 3D scene representation. In the area of graph-based computation, Yang et al. [6] introduced HGMatch, a match-by-hyperedge approach for subgraph matching on hypergraphs, enhancing the efficiency of complex graph analytical queries. Ding et al. [7] achieved substantial progress in visual surveillance through multi-scale adaptive clustering and local consistency learning for unsupervised clothing-changing person re-identification, tackling one of the most challenging variants of person retrieval. Within the automotive industry, Ziren [8] conducted dynamic optimization and multi-regional performance validation of automotive sales strategies in the United States, offering data-driven insights for market-specific decision-making. Lian and Chen [9] contributed to foundational AI methodologies through research on complex data mining analysis and pattern recognition based on deep learning, advancing capabilities for large-scale knowledge discovery. In photonic device engineering, Tang et al. [10] presented work on the design and optimization of shallow-angle grating couplers for vertical emission from Indium Phosphide devices, contributing to integrated optics development. Sun [11] explored AI-assisted UI design, demonstrating how generative tools can enhance both efficiency and creativity in the user interface design process. In the financial security domain, Yang et al. [12] constructed multi-dimensional network credit-related transaction risk maps by integrating graph neural networks, enabling early warning capabilities for financial anomalies. Complementing this, Shen et al. [13] applied the Whale Optimization Algorithm to financial payment fraud detection, demonstrating the effectiveness of bio-inspired optimization techniques in identifying fraudulent transactions.

2. ANALYSIS OF THE CURRENT STATUS OF GAME AI TRAINING BASED ON REINFORCEMENT LEARNING AND DEEP LEARNING

In the development of computer learning technology, we can use reinforcement learning theory to design game AI that can transform the process of presenting intelligent characters in the entire game into a simple Markov model structure. Intelligent characters use their intelligent sensing system or regional environment and their own work form, fully combining their own experience, to effectively select and analyze the operational behavior to be executed. The main implementation of this behavior affects the game system and realizes the continuous progress of the game process. At the same time, feedback analysis and evaluation of the intelligent character's execution behavior are often conducted under environmental conditions. For intelligent characters, the main task objective is to obtain the optimal feedback state of the environment by obtaining the best behavioral measures. Intelligent characters executing this form of operation in the game environment can be structured into a basic Markov chain. The best feedback effect obtained after a persistent loop is the main task objective of the current work. The traditional machine learning methods are mainly based on low dimensional inputs, and the actual convergence effect in this environment is relatively satisfactory. However, with the gradual enhancement of the actual performance of different hardware facilities and the innovation of neural network technology itself, there are relatively more problems faced by game intelligence, and technical issues are gradually becoming more prominent. At present, they mainly focus on the following aspects.

One is that the spatial and motion space technology in high-dimensional modes faces certain bottleneck problems. At present, behavioral intelligence technology can only perform simple walking and cannot achieve effective prediction, judgment, decision-making, and other complex behavioral patterns.

Secondly, the feedback in the game is relatively sparse, and there is also a certain degree of delay effect. The entire game mode and process are difficult to have a direct impact on the environment in special circumstances, making the training and analysis of intelligent objects even more difficult. In response to the emergence of such problems, neural network pattern optimization has become the main measure for adjusting high-dimensional action space management technology, and reinforcement learning provides a potential response path for feedback coefficients and delays. The intelligent operation mode can establish a communication efficient management mode or share relevant parameters for unstable environments, thereby achieving the best response.

3. BASIC ANALYSIS OF GAME AI TRAINING EXPERIMENTS USING REINFORCEMENT LEARNING AND DEEP REINFORCEMENT LEARNING

Through the screening and analysis of experimental carriers, relevant technical means were preliminarily formulated, and a basic prototype of the experiment was formed while constructing the experimental platform, thus achieving pre training analysis of traditional algorithms. Through research on multiple reinforcement learning training platforms related to game intelligence, it can be concluded that we use first person design classic games as carriers to accelerate the analysis and organization of carrier structures. The construction of a battle platform developed by a well-known game engine is the foundation for ensuring the orderly advancement of reinforcement learning and game AI training. In the context of our current work, we use the relevant operation management development model as the basis and use it as the experimental carrier content. We focus on the crisis of game engines and endow this platform with stable operation standards. At the same time, we can highly construct the experimental environment and version iteration analysis mode. The bottom layer adopts rational control optimization measures for analysis. Based on the implementation method of software development, it can enable relevant researchers to quickly familiarize themselves with the work content of game AI training and ensure the best game effect.

4. CONSTRUCTION OF GAME AI TRAINING DESIGN SCHEME BASED ON REINFORCEMENT LEARNING

Reinforcement learning mainly regards the interaction between the entire intelligent agent and the environment as the overall core problem, and constructs the maximum cumulative reward method through selective behavior analysis, so as to continuously implement learning optimization measures and form the best work sequence from it. It is possible to use reinforcement learning in a game environment to obtain a strategic approach for AI behavior in the entire complex environment based on reward based training. This introduces the reinforcement learning mode of the primary system.

The construction of deep reinforcement learning mainly combines the reinforcement learning ability foundation formed by deep learning, which can demonstrate the best working effect in many environmental situations. However, in the process of facing length decision problems, it often presents poor performance. Based on the theories of deep reinforcement learning, multi-agent systems, reinforcement layering, and attention mechanisms, we aim to innovate the working mode. Reinforcement learning is the process of constructing a Kolmogorov chain for the decision definition of a single agent based on a comprehensive observation system. The implementation of this process is mainly formed by constructing a tuple structure. In a special practical environment, the intelligent agent is always in a stable state environment, and through the time limit of policy execution actions, it obtains environmental feedback rewards and gains. At the same time, according to the structural form of the transition equation, it enters the next state mode. For Markov chains built around intelligent individuals, define reward feedback effects with loss coefficients in the environment and define action values. By maximizing the numerical response, we can obtain an optimal decision strategy result, so the optimal strategy function value used is constant. In the process of reinforcement learning, there is no fixed Markov process, and the agent needs to learn the optimal strategy through interaction with the environment. The widely used measure in various current environments is the combination of deep learning, which estimates the core numerical function through backup iteration.

The research and development in the field of functions has been relatively rapid, thanks to the research, analysis, and breakthrough presentation of deep neural network data. Combined with deep neural networks, some high-dimensional data can be directly transmitted and processed. The operation method of approximating equations with these constructed functions is included in it. At present, the Deep Q-Learning (DQN) we mentioned is a widely accepted processing method and approach, which has been widely used in different domain environments, including Go, Atari games, and so on. The most specific presentation method is to extract certain empirical elements from memory in the first iteration environment to update and transmit their parameter information. During the update process, the minimum loss function equation is used to ensure the best working effect. The experience memory adopts the most advanced queue representation form, in which the intelligent exploration strategy and experience tuple content data stored can support the stable progress of later operations. The speed and frequency of updating the relevant network parameters of the target in this form are relatively low, and the experience replay learning mechanism combined with it is constructed into a stable Deep Q-Learning relationship.

In order to better address the problem of not achieving optimal work results in feedback reward coefficients or delayed environments, a reinforcement learning mode is constructed by using a specialized framework structure such as a system. This framework is defined in each time environment, and the intelligent experience selects a primitive action or measures covering multiple steps for analysis. The implementation of each strategy involves different original actions or other strategic means, while completing the progress of work tasks based on random function information. Therefore, we will extend the traditional Markov policy process into a semi Markov decision-making process to address sparse feedback and delayed feedback to various problems.

By combining the training analysis of multiple agents, an independent network baseline training method can be constructed in the existing environment, and a network environment can be built separately for each intelligent object, placing multiple intelligent objects together to complete the training operation. In this context, by analyzing the communication protocol of intelligent objects, it is possible to enable them to share some of the observation information data, achieving the goal of accurate decision-making. Perform feature analysis and extraction in convolutional neural networks, then analyze and judge intelligent objects, achieve intelligence data sharing through communication center analysis and judgment, and achieve efficient processing of attention mechanisms. At present, in order to analyze the research and analysis of intelligent objects in action games under deep reinforcement learning methods, we can integrate the working characteristics of small input in the upper layer structure, use it as the visual field of intelligent objects for image transmission feedback, and then perform feature extraction analysis of convolutional networks. The extracted feature data information is shared and analyzed towards intelligent objects through the construction of communication centers, and then the entire data information is transmitted to the main operation structure for weighting and tracking analysis. After transferring to network training mode, multiple practical planning target structures will appear. The input of the lower level structure is mostly based on the perspective of intelligent objects and the planning objectives formed by the upper level structure. It mainly involves training and analyzing the network to input relevant execution actions, which are then handed over to the intelligent objects of the game for execution, achieving integration with the environment and ultimately building task objectives. This process continuously completes training operations, with the ultimate goal of achieving optimal collaborative processing and optimal working mode strategies under policy parameters through analysis.

5. IMPLEMENTATION OF GAME AI TRAINING BASED ON REINFORCEMENT LEARNING AND DEEP REINFORCEMENT LEARNING

In order to achieve the work objectives of each subtask, it is necessary to accelerate the analysis and processing of different subtasks, so as to truly complete the relevant task content. In the reinforcement learning task objective based on random neural networks, a general skill approach in the environment is pre learned using random neural networks. At the same time, different skill approaches are adjusted for the implementation and promotion of each task training, as well as individual strategy measures. In this proposed reinforcement learning model, a dual layer structure of workers and managers is designed. Workers use the relevant requirements and transfer results of each step as the transfer result task objective of managers. Due to the prominent practical nature of managers, the exploration direction and ability of agents are comprehensively improved and optimized.

Reinforcement learning has significant advantages in complex game design environments. Due to the influence of environment specific long sequences, reward coefficients, and other characteristics in the construction of different platform environments, traditional reinforcement learning methods themselves cannot achieve the best results among game subjects. In addition, some games such as mazes and ant search for items also require special design analysis to achieve the best results due to the high complexity of different scene environments. However, the method of implementing reinforcement learning through hierarchy requires the construction of separate sub objectives in the context of task decomposition, and based on a hierarchical reward analysis. This can greatly optimize and improve learning effectiveness, implement control measures, and achieve the best task level. Therefore, reinforcement learning method is to design complex games through learning, thereby forming potential optimization response methods.

Although reinforcement learning can effectively address the core problem of decision-making in the current working mode, this structural form mostly requires strategy training methods, with relatively low sample utilization efficiency and high practical difficulty in training. In this context, combining reinforcement learning with alternative strategy training and implementing alternative strategy correction methods to replace sub target structures in the sample through maximum estimation analysis. The analysis of the experimental results shows that the basic algorithm performs better than traditional algorithm structures. However, traditional algorithms use random sampling methods to seek an approximate replacement for the calculation of sub targets. This approach cannot effectively restrict and affect the sub target space, and sub targets can still choose a meaningless or non directly implementable method, resulting in unstable working effects due to the problems existing in lower level learning.

With the development and optimization of industrial structure, the working mechanism and behavioral environment of games have become more realistic and diverse. In many environments, complex behaviors need to be designed to achieve interactive analysis between AI games. This level of technology greatly affects the quality of game products themselves and the gaming experience of users. The design of such AI games is often time-consuming and labor-intensive, posing a huge challenge for game developers. This invention is based on reinforcement learning methods and decomposes the existing behavioral means of artificial intelligence and environmental interaction into a series of small task target themes. In the process of constructing the Q-table, accelerate the innovation of task goal construction to ensure the orderly completion of various tasks.

6. CONCLUSION

In order to achieve analysis of complex environments and effective judgment of complex behavioral decision-making problems, it is necessary to utilize relevant device engines and plugin information to develop and construct AI adversarial intelligent systems. In this context, the framework structure constructed in this article itself has distinct advantages. Only by doing a good job in game AI training with reinforcement learning and deep reinforcement learning can the best work processing effect be achieved to fully promote the game, reduce the direct impact of different hidden problems on game implementation, and improve the game experience.

REFERENCES

[1] Zheng, H., & Jiang, T. (2025). A New Methodology for Chinese Term Extraction from Scientific Publications. Innovation & Technology Advances, 3(2), 19–45. https://doi.org/10.61187/ita.v3i2.222

[2] Wu, J., Luo, L., & Liao, N. (2025). Small-Sample Object Detection of Surface Cracks in Concrete Structures of High-Rise Buildings via Multi-Level Transfer Learning. Innovation & Technology Advances, 3(2), 57–72. https://doi.org/10.61187/ita.v3i2.262

[3] Peng, Qucheng, Ce Zheng, and Chen Chen. "A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[4] Ukey, N., Zhang, G., Yang, Z., Li, B., Li, W., & Zhang, W. (2023). Efficient continuous kNN join over dynamic high-dimensional data. World Wide Web, 26(6), 3759-3794.

[5] Peng, Q., Planche, B., Gao, Z., Zheng, M., Choudhuri, A., Chen, T., Chen, C. and Wu, Z., 3D Vision-Language Gaussian Splatting. In The Thirteenth International Conference on Learning Representations.

[6] Yang, Z., Zhang, W., Lin, X., Zhang, Y., & Li, S. (2023, April). HGMatch: A Match-by-Hyperedge Approach for Subgraph Matching on Hypergraphs. In 2023 IEEE 39th International Conference on Data Engineering (ICDE) (pp. 2063-2076). IEEE.

[7] Y. Ding, Z. Ye, I. Xu, S. Lyu and L. Zhang, "Multi-Scale Adaptive Clustering and Local Consistency Learning for Unsupervised Clothing-Changing Person Re-Identification," in IEEE Transactions on Information Forensics and Security, vol. 21, pp. 2889-2904, 2026, doi: 10.1109/TIFS.2026.3671089.

[8] Ziren, Z. (2026). Dynamic Optimization and Multi-Regional Performance Validation of Automotive Sales Strategies in the United States. Academic Journal of Natural Science, 3(1), 1-7.

[9] Lian, J., & Chen, T. (2024). Research on Complex Data Mining Analysis and Pattern Recognition Based on Deep Learning. Journal of Computing and Electronic Information Management, 12(3), 37-41.

[10] Tang, Yingheng, et al. "Design and Optimization of Shallow-Angle Grating Coupler for Vertical Emission from Indium Phosphide Devices." (2020).

[11] Sun, Lingxin. "AI-Assisted UI Design: Enhancing Efficiency and Creativity through Generative Tools." Journal of Computer Technology and Applied Mathematics 3.1 (2026): 19-27.

[12] Yang, X., Zheng, X., & Lu, Q. (2025, October). Construction and early warning of multi-dimensional network credit-related transaction risk maps by integrating graph neural network (GNN). In Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (pp. 919-923).

[13] Shen, Zepeng, et al. "Research on Application of Whale Optimization Algorithm in Financial Payment Fraud Detection." 2025 4th International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID). IEEE, 2025.

Author Profile

Chendi Shen (2000-), male, undergraduate, Han ethnicity, native place: Shanghai, student at Shenzhen University, research direction: computer science.