2025 International Journal of Management Science Research, Vol.8, Issue 9, (Sep)
ISSN 2536-605X
DOI: 10.53469/ijomsr.2025.08(09).06

Game AI Training Based on Reinforcement Learning and Deep Reinforcement Learning

Yushen Chen

Shenzhen University, Shenzhen, Guangdong 518060

Abstract: ***Game AI training is a combination of computer and artificial intelligence. It is a main carrier environment in the field of reinforcement learning. At this stage, in the environment of game AI training, we are faced with moral difficulties and technical innovation. It mainly focuses on the feedback analysis of coefficient and delay, the space environment of high- dimensional state action and the characteristics of unstable environment. At present, based on deep reinforcement learning, we need to put forward the basic framework of deep reinforcement learning based on attention mechanism through the progress of reinforcement learning and deep reinforcement learning, so as to solve the problem of cluster intelligence in complex environment.***

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

1. INTRODUCTION

The definition of game AI training is relatively broad, and the method can be generated in the existing working environment to produce an appropriate level of intelligence, Then, making the game more realistic, interesting, and challenging can be regarded as game AI. There are only two ways in the entire game AI framework, one is based on the characteristics of finite state machines or behavior trees, Its own form can be predicted and analyzed by relevant technologies, while the other is based on neural networks, which can obtain non-qualitative AI through genetic neural networks and data calculation analysis, and its own behavior is difficult to be predicted and analyzed. Gao and Gorinevsky (2018) initially explored probabilistic methods for grid balancing in this context [1]. This work was subsequently expanded by Gao, Tayal, and Gorinevsky (2019), who applied probabilistic planning to minigrids [2]. Further developing these concepts, Gao and Gorinevsky (2020) formulated a probabilistic modeling framework for optimizing resource mix, which was published in IEEE Transactions on Power Systems [3]. Concurrently, advancements in other domains leveraging complex models have emerged. Chen et al. (2023) proposed a generative, text-guided 3D vision-language pretraining method for unified medical image segmentation [4]. In finance, Su et al. (2025) introduced an anomaly detection and early warning system for financial time series using a WaveLST-Trans model [5], while Zhang et al. (2025) developed MamNet, a hybrid model for network traffic forecasting and frequency analysis [6]. The application of deep learning in green finance was demonstrated by Zhang, Li, and Li (2025), who focused on carbon market price forecasting and risk evaluation [7]. Research at the intersection of vision and language continues to be prolific, as seen in the work of Peng et al. (2025) on 3D Vision-Language Gaussian Splatting [8] and Zhang et al. (2025) on dynamic cross-attention for fine-grained image captioning in advertising [9]. The optimization and security of large language models (LLMs) are also critical areas of investigation. Wen et al. (2025) presented a dynamic data filtering framework for fine-tuning LLMs [10], and Weng et al. (2025) proposed SecureGen, a framework to enhance security in text-to-image models [11]. Furthermore, several studies have focused on leveraging LLMs for text assessment, including Zhao et al.'s (2025) LLaM-ScoreNet for automated text quality evaluation [12] and Zhang et al.'s (2025) work on maximizing scoring divergence in automated essay assessment [13]. Huang et al. (2025) enhanced document-level question answering using multi-hop retrieval-augmented generation with LLaMA 3 [14]. The broader innovative applications of large models in computer science were discussed by Zhang et al. (2025) [15]. Beyond AI core technologies, Fang (2025) designed a cloud-native microservice architecture for cross-border logistics [16], and Wang and Bi (2025) introduced a hierarchical adaptive fine-tuning framework for multi-task learning in large-scale models [17]. In the financial sector, Cheng et al. (2025) developed FinStack-Net for financial fraud detection using ensemble learning [18]. Finally, Yang (2024) applied computer-assisted methods to communicative competence training in cross-cultural English teaching, bridging technology and pedagogy [19].

2. ANALYSIS OF THE CURRENT SITUATION OF TRAINING GAME AI 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 present intelligent characters in the game throughout the game, Translated into a simple Markov model structure, the intelligent character utilizes his intelligent perception system or regional environment as well as his own form of work, Fully combine their own experience to effectively select the execution of operational behavior analysis, which mainly implements the game system and achieves the continuous progress of the game process, while the environmental situation often provides feedback analysis on the execution behavior of intelligent characters. For intelligent actors, how to obtain the best behavioral measures can make obtaining the best feedback state of the environment the primary task goal. An intelligent character that performs this form of action in a game environment can be structured as a foundational Markov chain. The best feedback result that can be obtained when the cycle is persistent is the main task objective of the current work. Traditional machine learning methods are mainly based on low-dimensional inputs, and the practical results achieved in this environment are relatively satisfactory. However, with the gradual improvement of the actual performance of different hardware facilities, and with the technological innovation of the neural network itself, the problems faced by game intelligence are also relatively numerous, and the technical problems gradually become apparent, mainly focusing on the following aspects at this stage.

First, space and motion space technology under high-dimensional models face certain bottlenecks, and at this stage, behavioral intelligence technology intelligently completes walking relatively simply, and cannot achieve different and complex behavioral models such as effective prediction, judgment and decision-making.

Second, the feedback in the game is sparse, and there is also a certain latency effect, and the entire game's patterns and processes are difficult to have a more direct impact on the environment under special circumstances, which makes the training analysis of intelligent objects more difficult. In response to the emergence of such problems, neural network pattern optimization has become a major measure to address the adjustment of high-dimensional motion space management technology, and reinforcement learning provides a potential response path to feedback coefficients and latency. An intelligent operating approach to an unstable environment can establish a communication-efficient management model or share relevant parameters to achieve the best response.

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

Through the screening analysis of the experimental vector, the relevant technical means were initially formulated, and the basic prototype of the experiment was formed while the experimental basic platform was built, thus achieving the pre-training analysis of traditional algorithms. The analysis of several intensive learning training platforms related to game intelligence reveals that We use first-person design classics as the carrier to speed up the analysis and sorting of the carrier structure, and the construction of a combat platform developed by a well-known game engine is the basis for ensuring the promotion of reinforcement learning and the orderly progress of game AI training. In the context of our current work, we are using a relevant operational management development model as an experimental content, using game engine crisis and giving the platform stable operational working standards, while building an experimental environment and version iterative analysis model with a high degree of confidence. The bottom layer adopts rational control optimization measures for analysis, and based on the realization method of software development, relevant researchers can familiarize themselves with the work content of game AI training as soon as possible, ensuring the best game effect.

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

Reinforcement learning mainly takes the whole intelligent body and environment interaction as the overall core problem, and builds the maximum cumulative reward method through selective behavior analysis, so as to continuously implement learning optimization measures, and then form the optimal work sequence from there. Through reinforcement learning in the game environment, we can obtain a strategy for AI behavior in a complex environment based on the reward-based training set. This describes the systematic reinforcement learning model in primary education.

The construction of deep reinforcement learning, which mainly combines an enhanced learning foundation formed by deep learning, can perform optimally in many environmental settings, but often shows poor performance when faced with length decision problems. To this end, we base our work model on deep reinforcement learning theory, multi-intelligent system theory, reinforcement hierarchical thought theory, and attention mechanisms to achieve innovation. Reinforcement learning - in the context of a comprehensive observational system, the decision-making of a single intelligence is defined primarily as the process of constructing a Kolmaev chain. The implementation of this process is mainly constructed in a modular structure. In a particular practical environment, the intelligent body is always in a stable state environment, executes the time limit of the action by strategy, receives environmental feedback reward harvest, and enters the next state mode according to the structural pattern of the transfer equation. For a Markov chain constructed with intelligent individuals at its core, define the reward feedback effect with a loss coefficient in the environment and the definition of the action value. By maximizing the numerical response, we can obtain an optimal decision policy outcome, so the optimal policy function value to be adopted is a certain one. In reinforcement learning, there is no fixed Markov process, and the intelligence needs to learn the best strategy through interaction with the environment. The most widely used measures in current methodologies are those used in conjunction with deep learning, which estimates core numerical functions by backing up iterations.

Research in functional aspects has developed more rapidly, which is also due to the research analysis and breakthrough presentation of deep neural network data, which in combination with deep neural networks can directly transmit and methodize some high-dimensional data. The way in which these constructed functions approach equations operate is within this. Deep Q-Learning (DQN), as we speak at this stage, is a widely accepted approach and method that is currently used in different domains, including Go, Atari and so on. The most concrete way of presenting this is that in the first iteration environment, a certain empirical element is extracted from memory to implement the update transmission of its parameter information, and the minimized loss function equation is used in the update process to ensure the best working effect. Experience memory uses the most advanced queue manifestation, and the intelligent exploration strategy and experience element content data stored in it support the steady advancement of later operations. The speed and frequency of updating the relevant network parameters of the target in this form are relatively low. This combination of experiential playback learning mechanism constructs a stable Deep Q-Learning relationship.

In order to better address feedback reward factors or the problem of not achieving better work results in a latency environment, A reinforcement learning model is built by using specialized framework structures such as a system that defines that, in each time context, the intelligence selects a primitive action or a multistep measure to analyze. Each policy implementation covers different primitive actions or other policy means, while advancing the work task based on random function information. To this end, we expanded the traditional Markov methodology into a semi-Markov methodology for decision-making, dealing with sparse feedback and delayed feedback to problems.

Combined with multi-intelligence training analysis, it is possible to build a standalone network baseline training method in the existing environment, construct a network environment for each intelligent object separately, and place multiple intelligent objects together to complete the training operation. In this context, through the communication protocol analysis of intelligent objects, it is possible for intelligent objects to share some of the observation information data to achieve the purpose of accurate decision-making. Analysis of features is performed in convolutional neural networks, after which intelligent objects are investigated, intelligence data is shared with communication center analysis, and attention mechanisms are efficiently processed. At this stage, in order to analyze the intelligent object research analysis of action games under deep reinforcement learning method, we can synthesize the working characteristics of the small input of the superstructure, use it as the intelligent object view field for image transmission feedback, and perform feature extraction analysis of convolutional networks. The extracted feature data information completes the shared analysis to the intelligent object through the construction of communication center, then the entire data information is transmitted to the main operation structure for weighting and tracking processing analysis, and then transferred to the network training mode, there are multiple practice planning target structures. Most of the inputs in the lower layer structure are vision based on intelligent objects and planning objectives formed by the upper layer structure. It is mainly through network training analysis after input of relevant execution actions, and then handed over to the game's intelligent object to execute, achieve integration with the environment, and then achieve the construction of the task goal. This process continues to complete the training operation, with the ultimate expectation that the analysis will lead to optimal collaboration processing and optimal working mode strategy for policy parameters.

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

In order to accomplish the work goals of each subtask, it is necessary to speed up the analytical processing of different subtasks so that the relevant task content can be truly completed. The goal of a reinforcement learning task based on a stochastic neural network provides a means of generalizing skills in a pre-learning environment using stochatic neural networks, At the same time, different skill instruments are adjusted for the implementation of training for each task and for separate strategic measures, so the proposed reinforcement learning model designs workers, The manager's two-tier structure forms, in which staff follow the relevant requirements and migration results of each step as the manager's migration results task goals. Because the manager"s practice is more prominent, the intelligent body's exploration direction and capabilities have been comprehensively improved and optimized.

Reinforcement learning has a significant advantage in complex game design environments. For example, different platform environments are built because they are affected by characteristics such as the length of the environment, the reward factor, and so on. Traditional reinforcement learning methods themselves do not perform optimally in game players, and other games such as maze and ant search require special design analysis to achieve optimal results because of the high complexity of different scene environments. But achieving reinforcement learning through layers requires separate subgoal construction in the context of task decomposition and based on a layer of reward analysis. This can greatly optimize the learning effect, ensure the implementation of control measures, and achieve the optimal task level. To enhance the learning method, it is through the method of learning to design complex games, thereby forming the potential of optimized coping methods.

While reinforcement learning can effectively address the core issues of decision-making under current working methods, most of these forms of structure require a strategic training approach, with relatively low sample utilization efficiency and a high practical difficulty in training. In this context, reinforcement learning is combined with heterogeneous strategy training, and work is implemented in combination with heterogenous strategy correction methods to replace subtarget structures in the sample through maximum estimate analysis. Analysis of the results of the trial shows that the working effect of the basic algorithm is better than that of the traditional algorithm structure. But traditional algorithms use random sampling methods to seek an approximate alternative to the computation of subtargets, This pattern of practice does not adequately limit and influence the subtarget space, and the subtarget can still choose a method that is not meaningful or directly achievable, allowing the problem of underlying learning to have an unstable working effect.

With the development and optimization of the industrial structure, the working mechanism and behavioral environment of the game have become more real and complex and diverse. In many environments, the interaction analysis between AI games needs to be realized through the complex behavior of design, which to a large extent affects the quality of the game product itself and the game experience of the user. The design of this kind of AI game is often time-consuming and labor-intensive, which is a huge challenge for game developers. The invention is based on reinforcement learning methods, which break down the old behavioral means of AI and environment interaction into a series of small task target themes. In the process of completing the construction of Q table, accelerate the creation of task objectives and ensure the orderly completion of various tasks.

6. CONCLUSION

In order to achieve the analysis of complex environments and effectively judge complex behavior decision-making problems, it is necessary to use relevant equipment engines and plug-in information to realize the development and construction of AI counter-intelligent systems. Against this background, the framework structure body constructed by this paper has quite distinct characteristics and advantages in itself. Only by strengthening the training of game AI with reinforcement learning and deep reinforcement learning can we achieve the best work processing effect, realize the overall promotion of the game, reduce the direct impact of different hidden dangers on the realization of the game, and improve the game experience.

REFERENCES

[1] W. Gao and D. Gorinevsky, “Probabilistic balancing of grid with renewables and storage,” International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2018.
[2] Gao, W.; Tayal, D., Gorinevsky, D.: Probabilistic planning of minigrid with renewables and storage in Western Australia. In: 2019 IEEE Power & Energy Society General Meeting (PESGM) (2019). https://doi.org/10.1109/pesgm40551.2019.8973483
[3] Gao W and Gorinevsky D 2020 Probabilistic modeling for optimization of resource mix with variable generation and storage IEEE Trans. Power Syst. 35 4036–45
[4] Chen, Yinda, et al. "Generative text-guided 3d vision-language pretraining for unified medical image segmentation." arXiv preprint arXiv:2306.04811 (2023).

[5] Su, Tian, et al. "Anomaly Detection and Risk Early Warning System for Financial Time Series Based on the WaveLST-Trans Model." (2025).
[6] Zhang, Yujun, et al. "MamNet: A Novel Hybrid Model for Time-Series Forecasting and Frequency Pattern Analysis in Network Traffic." arXiv preprint arXiv:2507.00304 (2025).
[7] Zhang, Zongzhen, Qianwei Li, and Runlong Li. "Leveraging Deep Learning for Carbon Market Price Forecasting and Risk Evaluation in Green Finance Under Climate Change." Journal of Organizational and End User Computing (JOEUC) 37.1 (2025): 1-27.
[8] 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.
[9] Wen, Hairu, et al. "A Dynamic Adaptive Data Filtering and Sampling Framework for Optimized Fine-Tuning of Large Language Models." 2025 5th International Conference on Consumer Electronics and Computer Engineering (ICCECE). IEEE, 2025.
[10] Weng, Yijie, et al. "SecureGen: A Robust Framework for Enhancing Security and Resilience in Text-to-Image Generation Models." 2025 8th International Symposium on Big Data and Applied Statistics (ISBDAS). IEEE, 2025.
[11] Zhang, Wenqing, et al. "Dynamic Cross-Attention and Multi-Level Feature Fusion for Fine-Grained Image Captioning in Advertising." 2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE). IEEE, 2025.
[12] Zhao, Shihao, et al. "LLaM-ScoreNet: An Advanced Architecture for Automated Text Quality Assessment Based on LLaMA Language Model." 2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA). IEEE, 2025.
[13] Zhang, Danyang, et al. "Maximizing Scoring Divergence in Automated Essay Assessment with LLaMA-Based Meta-Attention Networks." (2025).
[14] Huang, Xinyue, et al. "Enhancing Document-Level Question Answering via Multi-Hop Retrieval-Augmented Generation with LLaMA 3." arXiv preprint arXiv:2506.16037 (2025).
[15] Zhang, Zheyu, et al. "Innovative Applications of Large Models in Computer Science: Technological Breakthroughs and Future Prospects." 2025 6th International Conference on Computer Engineering and Application (ICCEA). IEEE, 2025.
[16] Fang, Zhiwen. "Cloud-Native Microservice Architecture for Inclusive Cross-Border Logistics: Real-Time Tracking and Automated Customs Clearance for SMEs." Frontiers in Artificial Intelligence Research 2.2 (2025): 221-236.
[17] Wang, Yang, and Xiaowei Bi. "Hierarchical Adaptive Fine-Tuning Framework for Enhancing Multi-Task Learning in Large-Scale Models." 2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE). IEEE, 2025.
[18] Cheng, Zhang, et al. "FinStack-Net: Hierarchical Feature Crossing and Stacked Ensemble Learning for Financial Fraud Detection." (2025).
[19] Yang, C. (2024). A Study of Computer-Assisted Communicative Competence Training Methods in Cross-Cultural English Teaching. Applied Mathematics and Nonlinear Sciences, 9(1). Scopus. https://doi.org/10.2478/amns-2024-2895

Author Profile

Chenqi Shen (2000-), male, undergraduate, Han ethnic origin, student at Shenzhen University in Shanghai, research interests: computer science.