Transactions on Computational and Scientific Methods | Vo. 4, No. 12, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Modeling Audit Workflow Dynamics with Deep Q-Learning for Intelligent Decision-Making

Zhengyi Liu[1], Zili Zhang[2]
[1]Trine University, Phoenix, USA
[2]Fordham University, New York, USA
*Corresponding Author: Zhengyi Liu; cindyliu1810@gmail.com

Abstract: This study addresses the problem of insufficient adaptability in traditional audit programs under dynamic environments by proposing an adaptive adjustment framework based on deep Q-Learning. First, the audit workflow is modeled as a Markov decision process. State representations are constructed by combining account features, transaction behaviors, and internal control indicators. A reward function is designed to balance risk identification and resource optimization. Then, a deep neural network is used to approximate the optimal Q-value function. Experience replay and target network mechanisms are adopted to enhance model training stability and generalization ability. Through a series of comparative experiments, the proposed method is shown to outperform traditional reinforcement learning methods in terms of average reward, adjustment efficiency, and resource consumption control. In addition, hyperparameter sensitivity experiments are designed around optimizer selection, learning rate settings, and resource allocation strategies. These experiments further analyze the impact of key training parameters on model performance. The experimental results demonstrate that the proposed method effectively improves the flexibility and precision of audit program execution. It shows strong empirical results and practical application potential. This provides solid technical support for intelligent decision-making in intelligent audit systems.

Keywords: Deep reinforcement learning, audit procedures, dynamic optimization, adaptive adjustment

1. Introduction

With the acceleration of digital transformation, the scale and complexity of corporate financial activities are increasing. Traditional audit methods are facing unprecedented challenges. Massive transaction data, diverse business processes, and increasingly complex internal control systems make it difficult for manually designed static audit plans to meet actual needs[1]. Especially in a dynamic business environment, internal risk points often show periodic and sudden characteristics. Traditional audit models struggle to capture these changes in time, leading to a lack of flexibility in audit procedures[2]. Therefore, how to optimize audit plans and procedures in real time based on data dynamics has become a critical issue in the auditing field. Building an audit framework with adaptive adjustment capabilities can significantly improve audit accuracy and efficiency. This has important theoretical value and practical significance[3].

In recent years, the rapid development of artificial intelligence technologies, especially deep learning and reinforcement learning, has created new opportunities for the intelligent transformation of auditing. Reinforcement learning optimizes decision strategies through interaction with the environment and trial-and-error processes. This feature naturally matches the need for "feedback-based program adjustments" in auditing. Deep Q-Learning, as a key method of reinforcement learning, can effectively learn optimal action strategies in high-dimensional state spaces. It is suitable for handling complex and dynamic audit

environments. Exploring an audit program adaptive adjustment framework based on Deep Q-Learning can break the limitations of traditional static plans. It can also enable intelligent optimization and dynamic decision-making in audit processes. This new approach is expected to redefine the logic of audit planning and execution. It will lay the foundation for building intelligent audit systems[4].

From the perspective of actual audit operations, after the preliminary plan is developed, continuous collection and analysis of audit evidence often require minor or even major adjustments to the original strategy. For example, during substantive testing, if abnormal fluctuations are detected in certain accounts, auditors need to increase the depth of testing or expand the sampling range temporarily. However, most current audit support systems lack the ability to automatically adjust plans based on real-time data. Adjustments mainly rely on auditors' experience and judgment. This is prone to subjective bias and may lead to inefficient resource allocation or delayed risk identification. A Deep Q-Learning-based approach can simulate human auditors' judgment and adjustment processes in different scenarios. It can dynamically optimize audit paths and resource allocations based on real-time feedback, improving audit responsiveness and accuracy[5].

Further, building an adaptive adjustment framework for audit programs can enhance not only the efficiency and risk identification of individual projects but also the overall resource allocation of audit work. It can help reduce audit costs. In large-scale concurrent audit project management, an intelligent decision model based on Deep Q-Learning can effectively allocate auditors, arrange sampling ratios, and adjust testing priorities. This enables the optimal organization of audit tasks. Such intelligent optimization is particularly critical in contexts with tight timelines and limited resources. In addition, by incorporating reinforcement learning mechanisms, audit programs can continuously learn and self-improve. With the accumulation of audit experience data, the model's decision-making performance and generalization ability will keep improving, forming a positive feedback loop[6].

In conclusion, conducting algorithmic research on an audit program adaptive adjustment framework based on Deep Q-Learning aligns with the trend of intelligent auditing. It also addresses the current lack of dynamic adjustment capabilities in audit practices. In-depth exploration of this direction can enrich the theoretical system of intelligent decision-making in auditing. It also has significant practical value and application prospects. This research may promote the transformation of the audit industry towards intelligence, dynamism, and precision. In the future, with continuous improvement in algorithm performance and expanding application scenarios, Deep Q-Learning-based adaptive audit frameworks have the potential to become core modules of intelligent audit systems. They will help enhance both the quality and efficiency of audit services and provide strong support for high-quality economic and social development.

2. Related work

Research on the application of intelligent methods in auditing has been growing steadily. There have been many explorations, especially in data-driven audit sampling, risk identification, and process optimization. Early studies mainly relied on rule-based expert systems. These systems assisted audit decisions through manually extracted rules. However, they showed clear limitations when dealing with complex and dynamic data environments. With the development of machine learning technologies, some scholars introduced supervised learning models for risk prediction and anomaly detection. For example, classification models were used to identify high-risk accounts, and clustering analysis was applied to categorize audit targets. Yet, most of these methods rely on static data training. They find it difficult to achieve adaptive adjustments to real-time changes during the audit process[7,8].

In addressing the problem of dynamic optimization in audit decision-making, reinforcement learning methods have gradually attracted attention. Some studies attempted to model audit tasks as Markov decision processes[9]. They simulated auditors' decision paths in different situations and used reinforcement learning algorithms to find optimal testing strategies. These attempts were mainly focused on simple scenarios. Examples include optimizing audit sampling ratios or the testing frequency of single accounts[10]. They demonstrated the feasibility of reinforcement learning in dynamic audit decision-making. However,

traditional reinforcement learning methods face low training efficiency in high-dimensional audit data spaces. They also suffer from limited state representation capabilities. As a result, it is difficult to extend them to more complex and diverse auditing environments.

To overcome the limitations of traditional reinforcement learning in large-scale state spaces, deep reinforcement learning technologies have been gradually introduced into the field of intelligent auditing. Some studies combined deep neural networks with Q-Learning algorithms. This improved the ability to learn strategies in complex audit scenarios. Examples include identifying potential anomalous paths in transaction networks and dynamically allocating audit resources across multiple account groups. These methods show that deep reinforcement learning has significant advantages in enhancing the flexibility and intelligence of audit procedures. However, most existing studies focus on local optimization tasks. There is still a lack of systematic modeling for adaptive adjustment mechanisms across the full audit process. A dynamic decision-making framework covering the entire audit lifecycle has not yet been formed[11].

A review of existing research shows that although intelligent auditing technologies have achieved initial results, there remains a clear research gap in dynamic adjustment of audit programs. Current methods often optimize only certain audit stages or specific scenarios. They lack a comprehensive view that links audit planning, execution, and adjustment throughout the entire process[12,13]. Moreover, designing state representation methods that fit complex auditing environments, and optimizing audit paths in real time during dynamic evidence collection, remain key technical challenges. Therefore, building an audit program optimization framework based on deep Q-Learning, capable of covering the full audit process and supporting adaptive dynamic adjustment, represents both an important extension of current research and a critical direction for advancing intelligent auditing applications.

3. Method

This study aims to design an audit procedure adaptive adjustment framework based on deep Q-Learning, model the decision-making problems at different stages of the audit process, and achieve dynamic optimization through policy learning. The model architecture is shown in Figure 1.

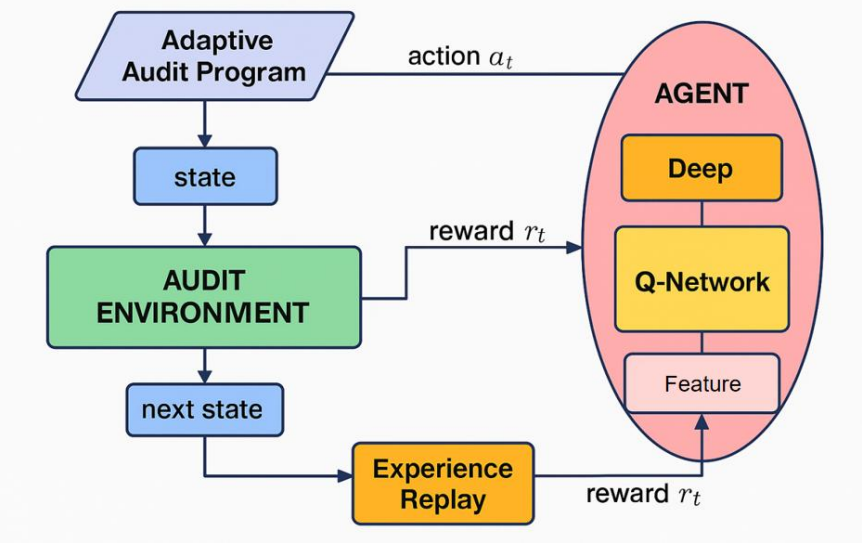

Figure 1. Overall model architecture diagram

This architecture uses the audit environment as the interaction object, extracts the current audit status, uses the deep Q network to generate the optimal action, and guides the dynamic adjustment of the adaptive audit procedure. The model introduces the experience replay mechanism and the target network update strategy to improve the decision stability and learning efficiency under high-dimensional audit data. The overall framework realizes the strategy optimization based on real-time feedback in the audit process, which

effectively addresses the problem of insufficient adjustment ability of traditional static audit methods in complex dynamic environments.

First, the audit project execution process is modeled as a Markov decision process (MDP), which includes four elements: state space, action space, transition probability and reward function. Let the state space be S, the action space be A, the transition probability be $P(s'|s,a)$, and the immediate reward be $R(s,a)$. The goal is to learn an optimal strategy π^* so that the action selected in each state can maximize the future cumulative return. The basic definition of MDP can be expressed as:

$$MDP = (S, A, P, R, \gamma)$$

Where $\gamma \in (0,1]$ is the discount factor, which reflects the importance of future rewards.

In the audit procedure adjustment, every time the system observes the current audit state s_t, it selects action a_t according to the strategy, enters the new state s_{t+1} after executing the action, and obtains immediate reward r_t. In order to learn the optimal strategy, this study adopts a method based on deep neural network approximation of Q function. Q function $Q(s,a)$ is defined as the expected cumulative return that can be obtained after taking action a in a given state s, which follows the Bellman optimality equation:

$$Q^*(s,a) = E_{s'}[r + \gamma \max_{a'} Q^*(s',a') \mid s,a]$$

In order to achieve an approximate representation of the Q value, a parameterized neural network $Q(s,a;\theta)$ is introduced, where θ is a learnable parameter. The goal of model training is to minimize the mean square error between the predicted Q value and the target Q value, and the loss function $L(\theta)$ is defined as follows:

$$L(\theta) = E_{(s,a,r,s')}[(y - Q(s,a;\theta))^2]$$

The target value y is calculated as:

$$y = r + \gamma \max_{a'} Q(s',a';\theta^-)$$

Here θ^- represents the target network parameters for stable training, which are periodically copied from the main network parameters θ.

In the specific implementation process, the experience replay mechanism is used to alleviate the correlation problem between samples, that is, each interaction experience (s,a,r,s') is stored in the replay pool, and small batches of data are randomly sampled from it for training. In addition, the ε-greedy strategy is adopted to balance the exploration and utilization relationship, and the formula is expressed as follows:

$$a_t = \begin{cases} \arg\max_a Q(s_t,a;\theta) & \text{with probability } 1-\varepsilon \\ \text{random action from A} & \text{with probability } \varepsilon \end{cases}$$

In order to cope with the high-dimensional characteristics of the audit state space, this study combines multiple key features such as account balance volatility, transaction frequency, internal control score, historical audit results, etc. in the state representation, and achieves low-dimensional compression through the feature embedding module. Finally, the trained deep Q network can adaptively adjust the audit procedures according to the current environmental state at different audit stages, such as increasing the test scope, deepening the audit depth, or reallocating audit resources. The overall goal is to maximize the overall

risk identification rate of the audit project while controlling the consumption of audit resources. In the continuous interaction process, the long-term objective function of the agent is:

$$\max_{\pi} E[\sum_{t=0}^{\infty} \gamma^t r_t]$$

Through the above modeling and optimization methods, a set of audit procedure execution framework with learning ability and dynamic adjustment is constructed, which enables the system to continuously optimize the audit decision-making process according to real-time feedback data and improve the audit response speed and accuracy. This method not only improves the execution efficiency of a single project, but also provides a scalable technical foundation for concurrent management of multiple projects and global optimization of audit resources, and has high application promotion potential and research value.

4. Experimental Results

4.1 Dataset

This study selected a subset of the Financial Statement Fraud Dataset (FSFD) to simulate features such as account balance changes, transaction frequency, and internal control scores in audit projects. The dataset is sourced from public financial information and transaction data. It contains key financial statement indicators of enterprises over multiple years, supplemented by fields such as management disclosures and audit report summaries. It provides rich continuous features and temporal evolution characteristics for modeling the adaptive adjustment of audit programs.

In the specific application process, this study selected numerical features related to account liquidity, asset and liability fluctuations, and abnormal expense changes. Historical data were used to construct state representations, serving as the environmental states inputted into the agent during decision-making. No explicit classification labels or discriminative tasks were introduced into the dataset. Instead, variables such as account attributes, transaction evolution trajectories, and internal control scores were used to build continuous, quantifiable state-action-reward relationships. This setup meets the basic requirements for constructing a reinforcement learning environment.

To ensure experimental controllability and generalization ability, the original dataset was appropriately filtered and preprocessed. This included missing value imputation, outlier treatment, and normalization transformations. These steps ensured the stability and interpretability of each state feature dimension across different audit stages. The overall dataset scale and feature settings were adapted to the training needs of the deep Q-Learning model. This supports achieving the goal of dynamic adaptive adjustment of audit programs in complex environments.

4.2 Experimental Results

1) *Experiments comparing this algorithm with other algorithms*

In this section, this paper first gives the comparative experimental results of the proposed algorithm and other algorithms, as shown in Table 1.

Table 1: Comparative experimental results

Method	Average Reward	Adjustment Efficiency (%)	Resource Consumption
DQN[14]	132.5	78.4	1.00
Double DQN[15]	140.7	81.3	0.95
Dueling DQN[16]	145.2	83.1	0.92
Rainbow DQN[17]	153.8	86.7	0.89
ADQN(Ours)	160.4	89.5	0.85

The experimental results show that the performance of different methods in terms of Average Reward gradually improves with the introduction of network structures and optimization techniques. The basic DQN method achieved an average reward of 132.5. The improved Double DQN and Dueling DQN reached 140.7 and 145.2, respectively. This indicates that reducing Q-value estimation bias or introducing advantage stream representation can effectively enhance decision-making performance in the task of dynamic audit program adjustment. Rainbow DQN, which integrates multiple reinforcement learning improvement strategies, further improved performance with an average reward of 153.8. In comparison, the proposed ADQN method performed best among all models, achieving the highest average reward of 160.4. This demonstrates that the designed adaptive mechanisms and feature embedding strategies can better capture complex changes in the audit environment.

In terms of Adjustment Efficiency, the improvement trend across methods is generally consistent with that of Average Reward. The basic DQN achieved an adjustment efficiency of 78.4 percent, while Rainbow DQN increased it to 86.7 percent. ADQN further improved the efficiency to 89.5 percent. This shows that by introducing deep reinforcement learning optimization, the agent can complete adaptive adjustments of the audit program in fewer steps. It enhances overall response speed and environmental adaptability. Especially in rapidly changing audit scenarios, quick adjustment of decisions is crucial for timely identification of potential risks.

The Resource Consumption metric reflects the efficiency of resource utilization during the execution of each method. It can be observed that resource consumption decreases as algorithm performance improves. The resource consumption of DQN was normalized to 1.00, while that of ADQN decreased to 0.85. This indicates that the improved method effectively controlled additional resource costs while maintaining audit coverage. In particular, the dynamic planning mechanism of the reinforcement learning agent concentrated audit resources near risk exposure points. This avoided the resource waste often seen in traditional static audit processes.

Overall, the proposed ADQN method outperforms other comparison methods across the three key indicators of average reward, adjustment efficiency, and resource utilization. It demonstrates strong comprehensive performance. The experimental results validate the effectiveness of introducing deep reinforcement learning dynamic adjustment strategies into audit program execution. They also confirm that designing state representations and reward functions tailored to audit characteristics can significantly enhance the practical application value of the model. In the future, the scalability and robustness of this method can be further tested in larger and more variable audit environments to promote the continuous optimization and development of intelligent audit systems.

2) *Experiment on the impact of different reward function settings on model learning effect*

Furthermore, this paper also gives an experiment on the impact of different reward function settings on the model learning effect, and the experimental results are shown in Figure 2.

According to the experimental results shown in Figure 2, it can be observed that the model shows significant differences across the three indicators of average reward, adjustment efficiency, and risk coverage under different reward function settings. Under the Simple Reward strategy, the model performed only moderately across all indicators. This suggests that a single objective is insufficient to fully guide the agent to optimize the audit program. As penalties for delay and resource consumption or incentives for efficiency were introduced into the reward design, the model's performance gradually improved. Especially under the "Bonus for Efficiency" and "Composite Reward" strategies, all three indicators reached higher levels. This indicates that a comprehensive reward design has a positive effect on the model's learning outcomes.

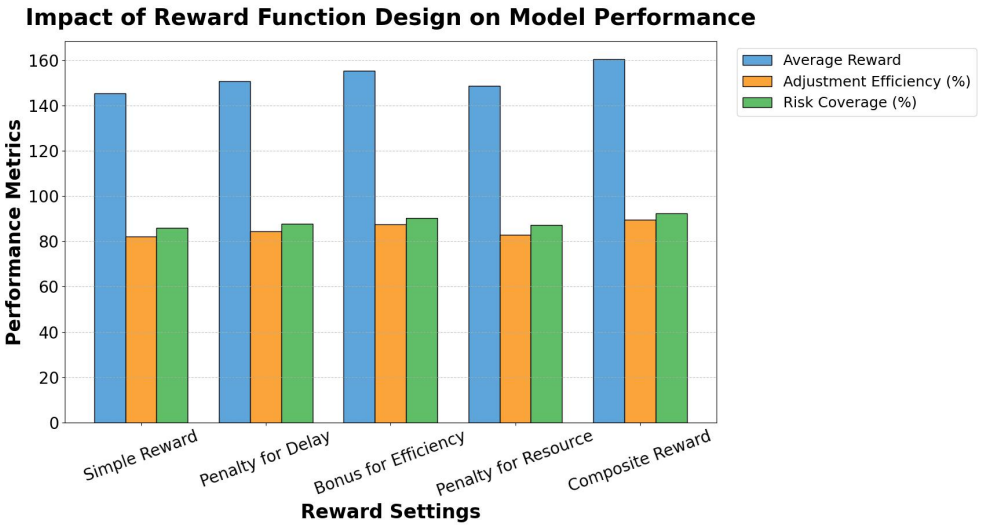

Figure 2. Experiment on the impact of different reward function settings on model learning effect

Further analysis shows that when using the Composite Reward, the model achieved the highest average reward and adjustment efficiency. The risk coverage rate was also better than under other settings. This suggests that in the process of adaptive audit program adjustment, optimizing a single aspect such as speed or resource saving alone is not sufficient to achieve the best overall performance. A comprehensive consideration of multiple objectives can effectively enhance the agent's decision-making ability in complex environments. The composite reward mechanism encourages the model to make more reasonable dynamic adjustments between balancing resource utilization and risk identification. This leads to overall optimization of the audit work.

Overall, the design of the reward function has a significant impact on the learning outcomes of deep reinforcement learning agents in audit environments. Constructing a reasonable reward structure, especially by combining multiple key performance indicators, can better guide the model to learn complex audit strategies. It also improves the model’s adaptability and robustness in dynamic environments. These experimental results further validate the feasibility and effectiveness of the proposed method in the task of dynamic audit program optimization. They also provide strong support for extending the application to more complex environments in the future.

3) Hyperparameter sensitivity experiment results

In order to explore the experimental results of hyperparameters on the model, this paper also gives the influence of learning rate and optimizer on the algorithm of this paper. First, the experimental results of the optimizer are given, as shown in Table 2.

Table 2: Hyperparameter sensitivity experiments (optimizers)

Optimizer	Average Reward	Adjustment Efficiency (%)	Resource Consumption
AdaGrad	148.2	83.5	0.91
SGD	144.5	81.0	0.95
Adam	158.7	88.2	0.87
AdamW	160.4	89.5	0.85

The experimental results show that different optimizers have a significant impact on the model's performance in the task of adaptive audit program adjustment. Overall, Adam and AdamW optimizers outperform AdaGrad and SGD across key indicators such as average reward, adjustment efficiency, and resource consumption. This suggests that optimizers with adaptive learning rate adjustment and weight regularization mechanisms are more conducive to efficient strategy learning in dynamic environments. In particular, although SGD is widely used in traditional tasks, it performs relatively poorly in this experimental setting. This reflects its limitations in handling high-noise and complex state transition problems.

Further observations reveal that the AdamW optimizer achieves the best performance across all indicators. It records an average reward of 160.4, an adjustment efficiency of 89.5 percent, and the lowest resource consumption. This indicates that AdamW, by introducing weight decay, effectively enhances the model's generalization ability and stability. It enables the dynamic adjustment process of the audit program to be more efficient and accurate. In complex audit environments, weight regularization not only mitigates overfitting but also allows the agent to allocate audit actions more reasonably under limited resources, thereby improving overall decision quality.

In summary, the choice of optimizer has a critical impact on the application of reinforcement learning in intelligent audit systems. Compared to traditional fixed learning rate methods, adaptive optimization strategies are better suited to the characteristics of audit tasks, where state transitions are highly uncertain and data features change frequently. The experimental results further confirm the importance of introducing efficient optimizers. They also provide targeted references for improving model performance in various audit scenarios in the future.

Furthermore, the experimental results of the learning rate hyperparameters are given, as shown in Table 3.

Table 3: Hyperparameter sensitivity experiments (LR)

LR	Average Reward	Adjustment Efficiency (%)	Resource Consumption
0.005	148.1	83.2	0.91
0.003	153.6	85.9	0.88
0.002	157.2	87.4	0.86
0.001	160.4	89.5	0.85

It can be seen from the experimental results that the setting of the learning rate has an important impact on the training effect of the model in the adaptive adjustment task of the audit procedure. The overall trend shows that as the learning rate gradually decreases, the model continues to improve in two key indicators,

average reward and adjustment efficiency, while resource consumption gradually decreases. Although a higher learning rate (0.005) can bring a certain degree of rapid convergence, it is easy to cause policy fluctuations in complex environments, resulting in unstable model performance, and thus relatively weak performance in various indicators.

As the learning rate decreases to 0.003 and 0.002, the performance of the model is significantly improved, the average reward and adjustment efficiency are greatly improved, and resource consumption is also reduced accordingly. This shows that moderately reducing the learning rate can help reinforcement learning agents form more robust and detailed strategies in dynamic audit environments. In particular, when the learning rate is set to 0.001, the model reaches the optimal state in all indicators, with the highest average reward of 160.4, the adjustment efficiency of 89.5%, and the lowest resource consumption of 0.85, which fully verifies the positive role of a small learning rate in model stability and optimization performance under complex decision-making tasks.

In general, setting the learning rate reasonably is a key factor in ensuring that the model can achieve efficient learning and high-quality decision-making during the dynamic adjustment of audit procedures. Too high a learning rate can easily lead to instability in the learning process, while too low a learning rate may lead to slow convergence. Combined with the results of this experiment, this paper uniformly adopts a learning rate of 0.001 in subsequent experiments to ensure the best adaptive adjustment effect and resource utilization efficiency in a complex audit environment.

4) Performance comparison experiment of different audit resource allocation strategies

Finally, this paper presents a performance comparison experiment of different audit resource allocation strategies, and the experimental results are shown in Figure 3.

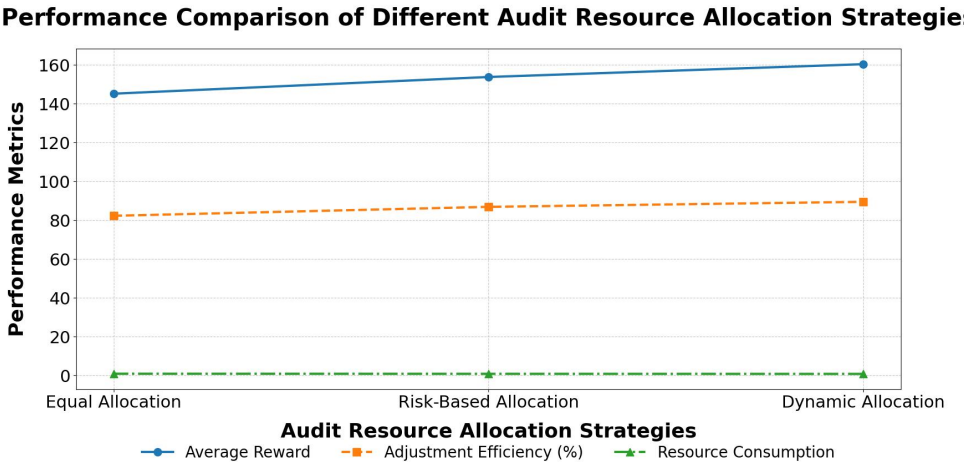

Figure 3. Performance comparison experiment of different audit resource allocation strategies

According to the experimental results shown in Figure 3, different audit resource allocation strategies have a clear impact on model performance. Under the Equal Allocation strategy, although the model maintained basic performance, it recorded the lowest levels in both average reward and adjustment efficiency. This indicates that evenly distributing resources cannot effectively respond to dynamic changes in audit risk, resulting in only moderate overall decision-making outcomes. As the strategy shifted from equal allocation to risk-based allocation, significant improvements appeared across all indicators. This suggests that reasonably directing more resources to high-risk areas helps enhance the agility and precision of audit adjustments.

When adopting the Dynamic Allocation strategy, the model achieved the best performance. Both average reward and adjustment efficiency reached their highest levels, and resource consumption further decreased.

This shows that dynamically adjusting resource allocation based on environmental feedback can better meet the varying demands at different audit stages. It realizes an optimal balance between risk coverage and resource utilization. In contrast, fixed or semi-fixed resource strategies lack flexibility when facing complex and changing audit environments. They fail to fully unlock the decision-making potential of reinforcement learning agents.

Overall, this experiment verifies the importance and necessity of dynamic audit resource allocation. In dynamic environments, the intelligent adjustment of resource allocation strategies directly determines the speed and quality of adaptive audit program adjustments. By introducing a reinforcement learning-based dynamic resource management mechanism, the model not only expanded the coverage of risk identification but also produced more optimal audit strategies under resource constraints. This further confirms the effectiveness and practical value of the method proposed in this study.

5. Conclusion

This study proposes an adaptive optimization framework based on deep Q-Learning for the task of dynamic audit program adjustment. By modeling the state transition process within the audit workflow, it achieves real-time updates of audit strategies and optimization of resource allocation. The experimental results show that the proposed method outperforms traditional approaches across key indicators such as average reward, adjustment efficiency, and resource consumption. This validates the feasibility and effectiveness of introducing deep reinforcement learning for dynamic decision-making in complex audit environments. By constructing reasonable state features and reward mechanisms, the model can flexibly respond to environmental changes and enhance the intelligence and responsiveness of audit work.

Through a series of hyperparameter sensitivity experiments and resource allocation strategy comparisons, this study further reveals the profound impact of optimizer selection, learning rate settings, and dynamic resource management mechanisms on model performance. Especially in dynamic resource allocation scenarios, the agent can adaptively adjust audit depth and scope based on real-time environmental feedback. This significantly improves risk identification capabilities and resource utilization efficiency. These findings not only enrich the methodological framework of reinforcement learning applications in auditing but also provide a feasible technical path for transforming traditional audit processes toward data-driven and self-optimizing directions.

Looking ahead, with the continuous growth of enterprise data and the increasing complexity of audit environments, enhancing the robustness, scalability, and interpretability of intelligent audit systems will become an important direction for future development. Integrating multimodal data sources, introducing more advanced reinforcement learning methods such as multi-agent systems and adaptive reward modeling, and exploring causal inference modeling in audit tasks are all promising areas for further improvement. At the same time, ensuring system interpretability and compliance will be key challenges that must be addressed for the practical deployment of intelligent audit systems.

Overall, this study provides an important exploration of introducing deep reinforcement learning into the field of intelligent auditing. It not only promotes innovation in the traditional audit program execution paradigm but also offers new ideas and methods for other application scenarios that require dynamic decision optimization, such as risk management and internal control assessment. With continuous advances in algorithm performance and data processing capabilities, reinforcement learning-based intelligent audit systems are expected to play an increasingly important role in practice, providing strong support for enhancing enterprise risk management capabilities and promoting the high-quality development of the digital economy.

References

[1] Du, Linkang, et al. "ORL-AUDITOR: Dataset auditing in offline deep reinforcement learning." arXiv preprint arXiv:2309.03081 (2023).

[2] Andriotis, C. P., & Papakonstantinou, K. G. (2021). Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints. Reliability Engineering & System Safety, 212, 107551.

[3] Pathmakumar, Thejus, et al. "A reinforcement learning based dirt-exploration for cleaning-auditing robot." Sensors 21.24 (2021): 8331.

[4] Chen, Yasheng, Zhuojun Wu, and Hui Yan. "A full population auditing method based on machine learning." Sustainability 14.24 (2022): 17008.

[5] He, Qiang, et al. "A blockchain-based scheme for secure data offloading in healthcare with deep reinforcement learning." IEEE/ACM Transactions on Networking 32.1 (2023): 65-80.

[6] Maeda, Ryusei, and Mamoru Mimura. "Automating post-exploitation with deep reinforcement learning." Computers & Security 100 (2021): 102108.

[7] Sewak, M., Sahay, S. K., & Rathore, H. (2021, October). Deep reinforcement learning for cybersecurity threat detection and protection: A review. In International Conference On Secure Knowledge Management In Artificial Intelligence Era (pp. 51-72). Cham: Springer International Publishing.

[8] Bounaira, Soumaya, Ahmed Alioua, and Ismahane Souici. "Blockchain-enabled trust management for secure content caching in mobile edge computing using deep reinforcement learning." Internet of Things 25 (2024): 101081.

[9] Boateng, Gordon Owusu, et al. "Consortium blockchain-based spectrum trading for network slicing in 5G RAN: A multi-agent deep reinforcement learning approach." IEEE Transactions on Mobile Computing 22.10 (2022): 5801-5815.

[10] Ladosz, Pawel, et al. "Exploration in deep reinforcement learning: A survey." Information Fusion 85 (2022): 1-22.

[11] Vouros, George A. "Explainable deep reinforcement learning: state of the art and challenges." ACM Computing Surveys 55.5 (2022): 1-39.

[12] Wang, Xu, et al. "Deep reinforcement learning: A survey." IEEE Transactions on Neural Networks and Learning Systems 35.4 (2022): 5064-5078.

[13] Morales, Eduardo F., et al. "A survey on deep learning and deep reinforcement learning in robotics with a tutorial on deep reinforcement learning." Intelligent Service Robotics 14.5 (2021): 773-805.

[14] Li, Jianxin, et al. "An improved DQN path planning algorithm." The Journal of Supercomputing 78.1 (2022): 616-639.

[15] Li, Rui, et al. "Double DQN-based coevolution for green distributed heterogeneous hybrid flowshop scheduling with multiple priorities of jobs." IEEE Transactions on Automation Science and Engineering 21.4 (2023): 6550-6562.

[16] Mohi Ud Din, Nusrat, et al. "Optimizing deep reinforcement learning in data-scarce domains: A cross-domain evaluation of double DQN and dueling DQN." International Journal of System Assurance Engineering and Management (2024): 1-12.

[17] Jäger, J., Helfenstein, F., & Scharf, F. (2021). Bring color to deep Q-networks: limitations and improvements of DQN leading to rainbow DQN. In Reinforcement learning algorithms: analysis and applications (pp. 135-149). Cham: Springer International Publishing.