Artificial Intelligence and Computing Innovations

Intelligence Science Publishing

https://www.ispublishing.com

ISSN: 3070-4944

Vo.4, No.1, 2024

Game-Theoretic Reinforcement Learning for Stable Competitive Resource Allocation in Dynamic Cloud Environments

Chun-Yao Hsieh[1]*

[1]*Columbia University, New York, USA*

**Corresponding author: ch3884@columbia.edu*

Abstract: With the increasing prevalence of multi-tenant concurrent access, heterogeneous task hybrid operation, and dynamic fluctuations in resource demand in cloud computing platforms, the stability problem of competitive allocation under limited resource conditions is becoming increasingly prominent. Addressing the shortcomings of traditional methods in simultaneously considering resource utilization, fairness constraints, and system stability in complex competitive environments, this paper proposes a cloud resource competitive allocation method based on the fusion of game theory and reinforcement learning. This method first starts from the multi-agent resource competition relationship in the cloud environment, modeling resource requesters and the scheduling process as a dynamic interaction process. State vectors are used to uniformly characterize key information such as remaining resources, queue pressure, task priority, and competition intensity, thereby enhancing the model's ability to describe real-world cloud resource competition scenarios. Based on this, a game-theoretic interaction mechanism and a hybrid strategy selection mechanism are introduced to model the strategy coupling relationship between different agents under limited resource constraints, enabling resource allocation decisions to better reflect the mutual influence and equilibrium characteristics in the multi-agent competition process. Furthermore, by combining reinforcement learning strategy optimization mechanisms, a reward-shaping guidance model simultaneously focuses on resource utilization efficiency, fair allocation effectiveness, and conflict suppression capabilities during long-term interactions. Stability regularization constraints are used to reduce fluctuations and imbalances in resource allocation, thereby improving the continuity, coordination, and robustness of the scheduling strategy in dynamic environments. After constructing a resource competition allocation scenario using public cloud cluster data, this paper systematically validates the proposed method. Results show that this method effectively improves the overall performance of resource competition allocation, demonstrating good application potential in resource utilization, fairness assurance, system stability, and overall coordination. The research indicates that the synergistic integration of game theory mechanisms and reinforcement learning strategy

optimization provides a feasible and interpretable technical path for studying the stability of resource competition allocation in complex cloud environments.

Keywords: Cloud resource scheduling; multi-agent decision-making; hybrid strategy selection; stability regularization

1. Introduction

With the continued popularization of cloud computing infrastructure, computing, storage, and network resources are gradually being opened to various types of users in a service-oriented and elastic manner[1]. Cloud platforms are also rapidly shifting from the traditional static resource supply model to a dynamic collaborative scheduling model oriented towards complex business scenarios. Against the backdrop of multi-tenant concurrency, task heterogeneity, and significantly increased demand fluctuations, cloud resource allocation is no longer a simple capacity allocation problem, but a comprehensive decision-making problem involving multiple objectives such as performance assurance, service fairness, resource utilization, and system stability. Especially in high-concurrency and highly competitive environments, different users, tasks, and services often engage in continuous games over limited resources, making resource contention, scheduling oscillations, and revenue imbalances more prominent, which places higher demands on the continuous and stable operation of cloud platforms[2].

Currently, the cloud resource competitive allocation problem exhibits significant dynamism, uncertainty, and coupling. On the one hand, user requests arrive randomly, task loads change frequently, and service quality requirements vary significantly, leading to continuous evolution of resource status. On the other hand, different entities exhibit clear strategies in the resource application, occupation, and release processes, and their behavior is influenced by price mechanisms, priority rules, historical allocation results, and environmental feedback. In this context, traditional resource allocation methods relying on static rules or single-step optimality often fail to accurately depict the long-term interactions between competing entities and struggle to achieve continuous and stable global optimization in complex environments. Therefore, studying cloud resource competition and allocation from the perspective of interactive decision-making and dynamic learning has become a crucial direction for improving the intelligent governance capabilities of cloud platforms[3,4].

Game theory can characterize the conflict of interests, strategy selection, and equilibrium formation mechanisms in the resource competition process from the perspective of multi-agent interaction, providing strong theoretical support for analyzing the interactive relationships between different participants in the cloud environment[5]. Reinforcement learning, through continuous interaction with the environment, autonomously learns resource decision-making strategies adapted to dynamic scenarios, thereby improving the system's real-time response capabilities and long-term performance under complex and uncertain conditions. Integrating game theory mechanisms with reinforcement learning methods helps to simultaneously consider competitive relationship modeling and adaptive strategy optimization, enabling the resource allocation process to not only reflect the characteristics of rational decision-making by multiple agents but also gradually approach a more stable, efficient, and fair allocation state through continuous feedback[6,7]. This integrated approach provides a new theoretical path to address the insufficient adaptability and stability of traditional methods in dynamic competitive environments.

2. Datasets and Dataset Preprocessing

2.1 Dataset

This paper selects Alibaba Cluster Trace v2018 as the research dataset. This dataset originates from an open-source cluster tracing project based on a real production cluster. It records the co-location and

operation of online services and batch processing tasks within the same cluster, realistically reflecting the operational status of cloud environments under conditions of multi-task concurrency, heterogeneous load coexistence, and limited resource competition. The dataset covers the operational trajectory of approximately 4000 machines over eight consecutive days, providing machine metadata, machine resource usage, container metadata, container resource usage, batch processing instance information, and batch processing task information in six core tables. It also includes DAG association information for batch tasks, thus supporting system modeling of the relationships between resource requests, resource consumption, task arrival, task execution, and scheduling.

This dataset is highly relevant to the research topic of cloud resource competition and allocation stability. Firstly, the co-location of online services and batch processing loads makes the competition for resources such as CPU and memory more pronounced for different task types, providing a foundation for multi-agent strategy interactions in game theory modeling. Secondly, the dataset simultaneously provides information such as resource requests, resource usage, and task-level dependencies, making it suitable for characterizing the continuous evolution of resource allocation states in dynamic environments. It also provides support for state construction, reward design, and policy updates in reinforcement learning. Research based on this dataset helps analyze the intrinsic relationships between resource contention, policy adjustments, and system stability from the perspective of real-world cloud workloads, thereby enhancing the consistency between research questions and actual cloud resource management scenarios.

2.2 Dataset Preprocessing

In the data preprocessing stage, the machine information, task information, instance information, and resource usage records in the original cluster tracking data are first aligned, and timestamps are standardized. Duplicate records, samples with significant missing data, and data with abnormal formats are removed to ensure the quality of the basic data. Then, based on key fields such as machine ID, task ID, and instance ID, multi-table joins are performed to construct a unified time-series sample reflecting task arrival, resource requests, resource usage, and changes in execution status. On this basis, continuous resource attributes such as CPU and memory are normalized, categorical status fields are numerically encoded, and status sequences are organized according to fixed time windows or continuous scheduling segments to characterize the dynamic evolution of multi-agent resource competition. Finally, combining information such as the difference between resource requests and actual usage, changes in task queuing and scheduling status, and node load levels, key features characterizing the intensity of cloud resource competition and allocation stability are extracted, providing a unified, standardized, and computable input representation for subsequent game theory modeling and reinforcement learning decision-making.

As shown in Figure 1, the overall data quality and statistical distribution were significantly improved after data preprocessing. The top left figure indicates that the distribution of the CPU Utilization Bin is more concentrated and smoother after preprocessing, with reduced abnormal fluctuations and dispersion, indicating that resource usage records, after standardization and cleaning, better reflect the actual task load characteristics. The top right figure shows that the Memory Gap Ratio Bin clusters significantly towards the lower range after preprocessing, indicating a reduction in the deviation between memory request values and actual usage values, resulting in a more accurate description of resource requirements. The bottom left figure reflects a comprehensive decrease in the Missing Ratio of each key field after preprocessing, demonstrating that missing value handling effectively improves the completeness of multi-table related data. In the bottom right figure, the fluctuation amplitude of Node CPU Std across time windows converges significantly, indicating that after anomaly cleanup, time alignment, and feature normalization, the node-level load sequence is more stable, which is beneficial for subsequently characterizing the state evolution patterns in the cloud resource competition process and providing a more reliable data foundation for game modeling and reinforcement learning strategy learning.

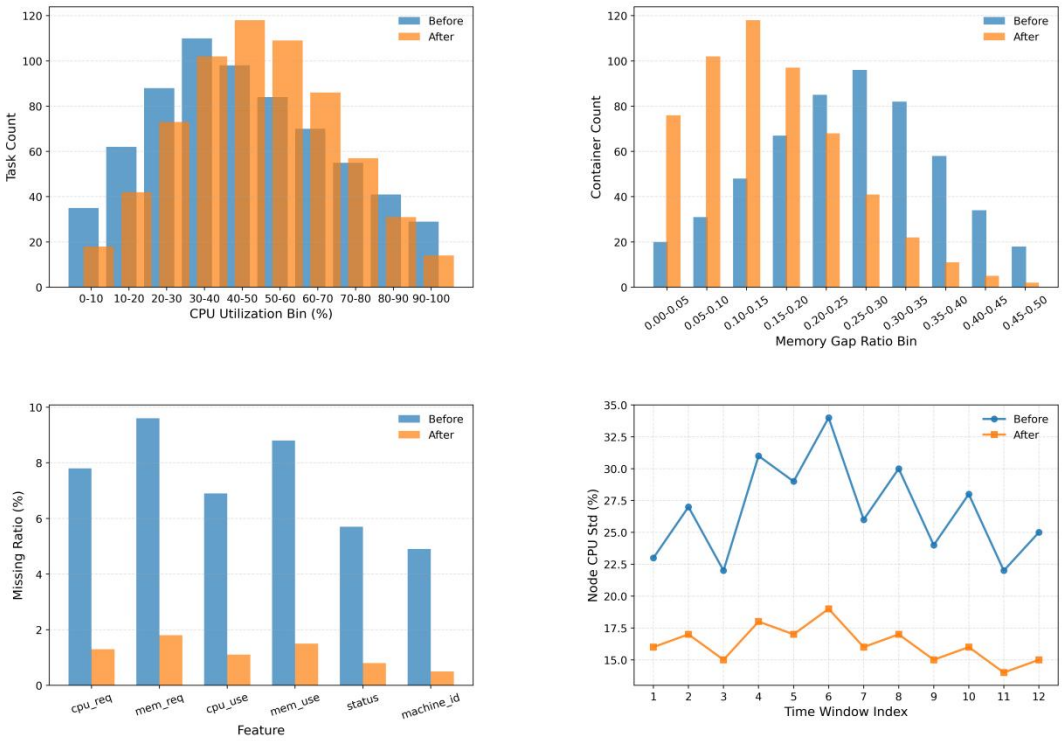


Figure 1. Statistical comparison before and after data preprocessing

3. Methodology

In a competitive cloud environment, resource allocation must jointly reflect instantaneous system conditions and the strategic reactions of heterogeneous tenants, since purely myopic scheduling often amplifies contention and induces unstable oscillations under bursty demand. To capture this coupling, the cloud platform is modeled as a multi-agent stochastic game in which each agent represents a resource requester or a scheduling entity, while the environment evolves with workload arrivals, resource occupation, and release events over time. A state vector is therefore constructed to summarize the observable operating condition, including residual CPU capacity, residual memory capacity, queue length, task urgency, and recent contention intensity, so that the decision process remains sensitive to both supply pressure and competitive interaction. The overall model architecture is also presented here, as shown in Figure 2.

The underlying objective is not merely to maximize short-term allocation gain, but to regulate the dynamic balance among efficiency, fairness, and stability, which is essential for preventing repeated reallocations and performance degradation in shared cloud infrastructure. Formally, the interactive allocation process is defined on a tuple:

$$M = \{N, S, \{A_i\}_{i=1}^{N}, P, \{r_i\}_{i=1}^{N}, \gamma\} \tag{1}$$

where $\mathcal{N}$ denotes the agent set, $\mathcal{S}$ is the global state space, $\mathcal{A}_i$ is the action space of agent i, $\mathcal{P}$ denotes the state transition law, r_i is the reward function, and γ is the discount factor controlling the contribution of long-term return. For each decision epoch t, the system state is written as:

$$s_t = [c_t^{cpu}, c_t^{mem}, q_t, d_t^{prio}, \rho_t] \tag{2}$$

in which c_t^{cpu} and c_t^{mem} denote the normalized remaining CPU and memory capacities, q_t measures queue pressure, d_t^{prio} characterizes task urgency, and ρ_t represents the current competition intensity derived from concurrent requests per allocable resource unit. This formulation is introduced to ensure that the learning agent does not optimize in isolation, but instead responds to the structural tension between local demand and shared platform capacity.

Beyond representing the environment, strategic dependence among agents must be explicitly encoded because cloud tenants continuously adjust their requests according to expected utility and observed system congestion. A mixed-strategy game is therefore embedded into the allocation loop, allowing each agent to select an action with a probability distribution rather than a fixed deterministic response, which improves adaptability under uncertain competition. The expected utility of agent i is defined as:

$$U_i(a_i, \pi_{-i}, s_t) = \sum_{a_{-i} \in A_{-i}} \pi_{-i}(a_{-i} \mid s_t)\, u_i(a_i, a_{-i}, s_t) \tag{3}$$

where a_i is the action of agent i, π_{-i} denotes the joint mixed strategy of the other agents, and u_i is the one-step utility determined by allocation benefit, waiting cost, and contention penalty under the state s_t. Stability at the strategic level is encouraged by requiring the policy profile to approach a state-dependent equilibrium region, which can be expressed as:

$$\pi_i^* \in \arg\max_{\pi_i} E_{a_i \sim \pi_i,\ a_{-i} \sim \pi_{-i}^*} U_i(a_i, \pi_{-i}^*, s_t) \tag{4}$$

so that no agent has an incentive to unilaterally deviate when the system reaches a balanced competitive configuration. Such a design is meaningful because resource contention in cloud systems is rarely static; instead, it emerges from repeated strategic adaptation, and equilibrium-oriented modeling provides a principled way to suppress aggressive resource grabbing and reduce allocation volatility.

To make the equilibrium mechanism operational in a dynamic setting, reinforcement learning is introduced to optimize policies through long-horizon interaction with the evolving cloud environment rather than through fixed analytical rules. The reward is shaped to preserve service efficiency while discouraging unstable behavior, since a scheduler that only maximizes immediate utilization may sacrifice fairness and trigger repeated reallocations. Specifically, the instantaneous reward is constructed as:

$$r_t = \alpha R_t^{util} + \beta R_t^{fair} - \lambda R_t^{conflict} - \mu R_t^{switch} \tag{5}$$

where R_t^{util} measures effective resource utilization, R_t^{fair} quantifies balanced access across competing agents, R_t^{conflict} penalizes excessive competition or rejected allocations, and R_t^{switch} captures policy-induced instability such as frequent reassignment or abrupt decision changes, while α, β, λ, and μ control the relative importance of these terms. Long-term optimization is then driven by the action-value target:

$$Q_i(s_t, a_t) \leftarrow (1-\eta) Q_i(s_t, a_t) + \eta \left[r_t + \gamma \max_{a_i'} Q_i(s_{t+1}, a_i') \right] \tag{6}$$

with learning rate η, so that each agent gradually updates its allocation preference according to both immediate feedback and anticipated future system states. This step is introduced because stable cloud allocation cannot be guaranteed by one-shot equilibrium analysis alone; the platform must continuously

adapt to workload drift, resource fragmentation, and changing contention patterns, and reinforcement learning supplies that adaptive capability through repeated policy refinement.

A final layer is added to explicitly regularize allocation stability, because even utility-consistent and reward-maximizing decisions may remain undesirable if they produce large temporal fluctuations in resource assignment. Instead of treating stability as an implicit byproduct, the method quantifies it as a separate optimization signal derived from the variation of consecutive allocation outcomes, thereby favoring smooth policy evolution and sustained service consistency. The stability objective is defined as:

$$L_{stab} = \frac{1}{T-1} \sum_{t=2}^{I} \|x_t - x_{t-1}\|_2^2 \tag{7}$$

where x_t denotes the allocation vector at time t, and smaller values indicate lower inter-temporal oscillation in assigned resources. On this basis, the overall training objective is written as:

$$J = \mathbb{E}\left[\sum_{t=1}^{I} \gamma^{t-1} r_t\right] - \omega L_{stab} \tag{8}$$

with coefficient ω balancing cumulative return against temporal smoothness. By coupling game-theoretic equilibrium constraints, long-horizon reinforcement learning, and explicit stability regularization within one unified framework, the proposed method is able to align rational competition with adaptive scheduling, thereby providing a more robust foundation for cloud resource allocation under persistent multi-agent contention.

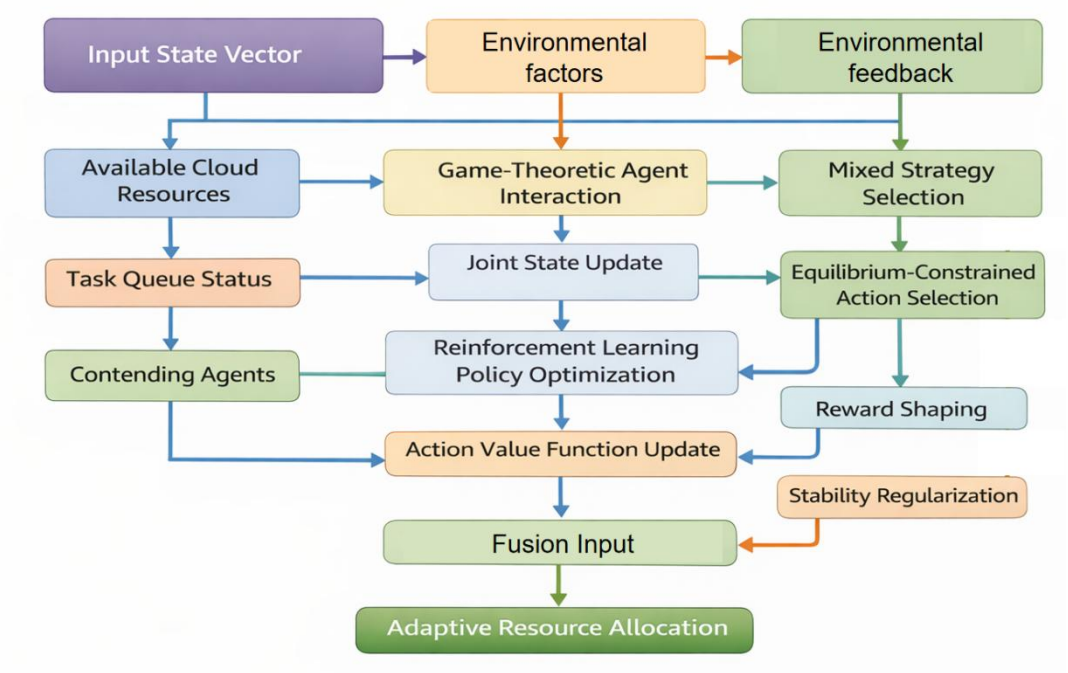


Figure 2. Overall model architecture

4. Experimental Results and Analysis

This paper first presents the results of experiments comparing the model with other models.

Table 1. Experimental results compared with other models

Method	UR	FR	SR	SI
Wei et al.[8]	83.42	80.16	84.37	79.25
Xu et al.[9]	84.08	82.41	85.12	80.33
Yan et al.[10]	85.27	83.09	86.48	81.57
Liu et al.[11]	87.14	85.63	88.05	84.22
Thein et al.[12]	86.35	84.18	87.26	83.46
Cui et al.[13]	88.29	86.47	89.31	85.74
Belgacem et al.[14]	89.16	87.38	90.24	86.91
Ours	92.74	90.83	93.41	91.26

Table 1 presents four evaluation metrics that characterize the performance of cloud resource allocation from different perspectives. UR primarily reflects resource utilization, measuring whether resources are fully utilized and idle time is minimized. FR describes allocation fairness, reflecting the degree of resource equality among different competing entities in a shared environment. SR focuses on system stability, judging whether the resource scheduling process can maintain a continuous, orderly, and low-fluctuation configuration. SI further reflects overall stability and coordination capabilities, comprehensively demonstrating the robustness and consistency of the strategy in complex competitive environments. These metrics together constitute a relatively complete evaluation system, ensuring that cloud resource allocation methods not only focus on efficiency but also consider fairness and long-term stable operation requirements.

The proposed algorithm demonstrates strong advantages across all metrics, indicating that this method can simultaneously consider utilization efficiency, fairness constraints, and system stability in a resource-competitive environment. The reason for this is that the game theory mechanism enhances the ability to characterize the competitive relationships among multiple parties, enabling resource allocation decisions to better adapt to the strategic interactions between different tasks; while the reinforcement learning mechanism further improves the adaptive adjustment ability of strategies in dynamic environments, making the scheduling process more continuous and coordinated. The integration of the two not only improves the quality of resource allocation but also effectively alleviates the instability caused by competitive conflicts and state fluctuations, thus exhibiting superior overall performance.

To further verify the contribution of each component module to the overall method, Table 2 presents the ablation settings that strictly correspond to the model design. These settings are implemented from three levels: game-theoretic interaction modeling, equilibrium constraint decision-making, and stability regularization, directly reflecting the impact of different mechanisms on the cloud resource competition and allocation process. As shown in Table 2, the complete model maintains superior performance in resource utilization, fair allocation, system stability, and comprehensive coordination, indicating strong synergy among the sub-modules, which collectively support the method's stable decision-making capability in a dynamic competitive environment.

Table 2. Ablation test results

Method	UR	FR	SR	SI
w/o Game-Theoretic Agent Interaction	89.46	87.18	90.12	88.35
w/o Equilibrium-Constrained Action Selection	90.03	88.24	90.87	89.11
w/o Stability Regularization	89.72	87.95	89.68	88.74
Ours	92.74	90.83	93.41	91.26

The ablation experiments show that the proposed complete method performs well overall, indicating that each component plays an irreplaceable role in the cloud resource competition and allocation process. Game-theoretic interaction modeling enhances the ability to characterize the resource competition relationship among multiple parties, making resource allocation decisions more consistent with the strategy coupling characteristics of the actual cloud environment; equilibrium constraint decision-making further improves the coordination and allocation rationality in the scheduling process, enabling the system to maintain a more stable resource configuration state under competitive conditions; stability regularization effectively suppresses fluctuations and imbalances in the allocation process, enhancing the model's continuous decision-making ability in dynamic environments. With the combined effect of all modules, the model achieves superior performance in resource utilization, fairness assurance, system stability, and overall coordination, demonstrating that the proposed method has strong structural integrity and mechanism synergy.

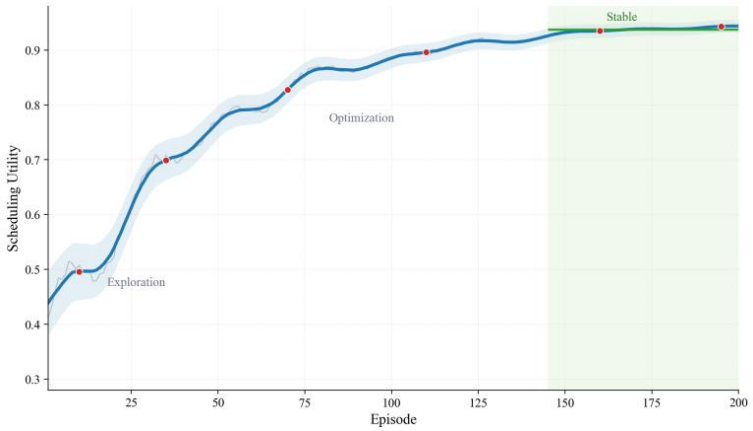


Figure 3. Visualization experiment of resource scheduling convergence behavior during reinforcement learning strategy optimization.

As shown in Figure 3, the proposed algorithm achieves relatively ideal resource scheduling convergence during reinforcement learning strategy optimization, demonstrating strong decision-making stability and scheduling adaptability in dynamic cloud environments. The consistently high scheduling efficiency in the figure indicates good synergy between the constructed game modeling mechanism, equilibrium constraint decision-making mechanism, and reinforcement learning optimization mechanism, enabling them to jointly support the effective learning and continuous optimization of resource allocation strategies. This performance demonstrates that the proposed method not only perceives the cloud resource competition state but also maintains high-quality resource scheduling output under complex interactive conditions.

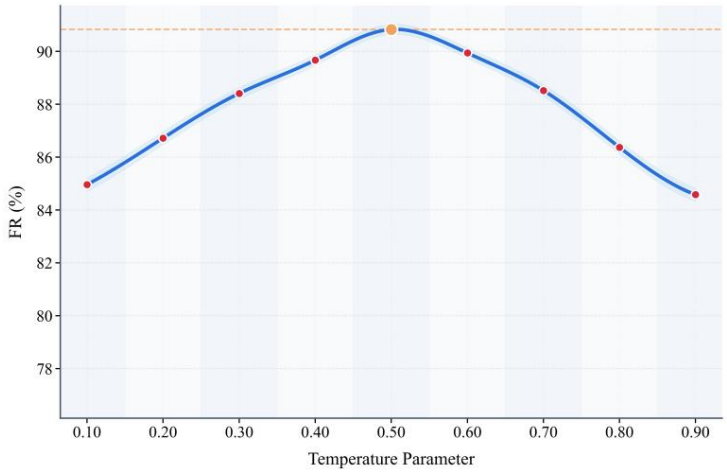


Figure 4. The effect of temperature parameter sensitivity of the hybrid strategy on FR

The proposed algorithm maintains good fairness in resource allocation under mixed strategy temperature parameter settings, demonstrating its strong strategy adjustment and allocation coordination capabilities in resource competition environments, as shown in Figure 4. The introduction of the temperature parameter allows the model to achieve a more reasonable balance between strategy exploration and stable decision-making, ensuring that the resource allocation process has the necessary flexibility without compromising overall fairness constraints. By combining game interaction modeling, equilibrium constraint action selection, and reinforcement learning strategy optimization mechanisms, the proposed method can more effectively adapt to complex state changes in cloud resource competition allocation and exhibits good overall performance in ensuring fairness.

5. Conclusion

This paper addresses the stability problem in resource allocation competition within a cloud computing environment, proposing a unified modeling method that integrates game theory mechanisms and reinforcement learning strategy optimization. To overcome the limitations of traditional resource scheduling methods in handling multi-agent competition, dynamic load fluctuations, and long-term stability control, this paper constructs a resource allocation framework for complex cloud environments from three levels: competitive relationship characterization, equilibrium constraint decision-making, and stability regulation. This method not only reflects the strategic coupling relationships among resource requesters from a multi-agent interaction perspective but also leverages reinforcement learning to continuously optimize resource scheduling behavior, thereby improving the adaptive, coordination, and stability maintenance capabilities of the allocation process. The results demonstrate that the proposed method exhibits good overall performance in resource utilization, fairness assurance, and system stability, indicating that this research provides a new approach with strong interpretability and practical value for the cloud resource allocation competition problem.

From a theoretical perspective, this work promotes the further development of cloud resource management research from static rule-driven to dynamic intelligent decision-driven approaches. The introduction of game theory transforms resource allocation from a simple process of a single scheduler allocating resources to a continuous process of competition and collaboration among multiple actors under limited resource constraints. The integration of reinforcement learning further enhances the model's responsiveness to the evolutionary characteristics of complex environments, enabling resource allocation strategies to be continuously revised and improved through long-term interactions. This fusion improves the resource scheduling model's relevance to real-world cloud environments and provides a new analytical framework for studying the evolution of competitive behavior, equilibrium formation mechanisms, and stability control paths in cloud platforms. Therefore, this paper not only focuses on the

resource allocation results themselves but also emphasizes the dynamic coordination logic and system operational quality during the resource allocation process, which is significant for building a more robust intelligent cloud scheduling theoretical system.

From an application perspective, the method proposed in this paper has strong reference value for various practical scenarios. In multi-tenant cloud platforms, this method helps alleviate resource contention and allocation imbalances under high-concurrency access conditions, thereby improving platform service quality and user experience. In application scenarios such as edge cloud collaboration, container orchestration, task offloading, and heterogeneous computing resource scheduling, this method also provides support for real-time decision-making in complex resource environments, thus enhancing the continuity and reliability of system operation. For resource governance tasks in industrial internet, smart cities, intelligent manufacturing, online service platforms, and large-scale data centers, a stable and efficient resource allocation mechanism is directly related to platform throughput, service continuity, and operating cost control. The research conclusions presented in this paper have a certain practical driving effect on promoting the intelligent upgrading of cloud computing platforms, improving the precision of resource scheduling, and supporting the stable operation of critical businesses.

Future research can be further developed in several directions. Firstly, it can be applied to cloud environments with larger scales and more types of resources, further enhancing the model's ability to express the coupling relationships of heterogeneous resources, making it applicable to more complex production scenarios. Secondly, task dependencies, service level constraints, energy consumption control, and carbon emission constraints can be incorporated into a unified decision-making framework, thereby promoting the expansion of resource competition allocation research from a purely performance-oriented approach to a green, low-carbon, and sustainable approach. Thirdly, federated learning, online learning, and distributed collaborative optimization mechanisms can be combined to enhance the model's deployment capabilities under cross-regional, multi-cluster, and privacy-constrained conditions. Fourthly, the integration of the methodology with actual cloud platform scheduling systems can be further strengthened, promoting the transformation of relevant theoretical results into engineering applications. With the continuous expansion of cloud infrastructure and the increasing complexity of business scenarios, intelligent scheduling methods focused on the stability of competitive allocation will play an increasingly important role in future digital infrastructure construction.

References

[1] Z. Xu, J. Tang, C. Yin, et al., "ReCARL: Resource allocation in cloud RANs with deep reinforcement learning," IEEE Transactions on Mobile Computing, vol. 21, no. 7, pp. 2533-2545, 2020.

[2] J. Cen and Y. Li, "Resource allocation strategy using deep reinforcement learning in cloud-edge collaborative computing environment," Mobile Information Systems, vol. 2022, no. 1, Art. no. 9597429, 2022.

[3] Y. Jiang, Y. Mao, G. Wu, et al., "A collaborative optimization strategy for computing offloading and resource allocation based on multi-agent deep reinforcement learning," Computers and Electrical Engineering, vol. 103, Art. no. 108278, 2022.

[4] W. Guo, W. Tian, Y. Ye, et al., "Cloud resource scheduling with deep reinforcement learning and imitation learning," IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3576-3586, 2021.

[5] H. Zhang, J. Wang, H. Zhang, et al., "Security computing resource allocation based on deep reinforcement learning in serverless multi-cloud edge computing," Future Generation Computer Systems, vol. 151, pp. 152-161, 2024.

[6] F. J. Maldonado-Carrascosa, S. García-Galán, M. Valverde-Ibáñez, et al., "Game theory-based virtual machine migration for energy sustainability in cloud data centers," Applied Energy, vol. 372, Art. no. 123798, 2024.

[7] G. Zhou, W. Tian, R. Buyya, et al., "Deep Reinforcement Learning-based Methods for Resource Scheduling in Cloud Computing: A Review and Future Directions," ACM Computing Surveys, vol. 54, no. 11s, Art. no. 245, 2022.

[8] G. Wei, A. V. Vasilakos, Y. Zheng, et al., "A game-theoretic method of fair resource allocation for cloud computing services," The Journal of Supercomputing, vol. 54, no. 2, pp. 252-269, 2010.

[9] X. Xu and H. Yu, "A game theory approach to fair and efficient resource allocation in cloud computing," Mathematical Problems in Engineering, vol. 2014, no. 1, Art. no. 915878, 2014.

[10] W. Yan, J. Wang, and J. Sun, "A game-theoretic based resource allocation strategy for cloud computing services," Scientific Programming, vol. 2016, no. 1, Art. no. 1629893, 2016.

[11] N. Liu, Z. Li, J. Xu, et al., "A hierarchical framework of cloud resource allocation and power management using deep reinforcement learning," in Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems, 2017, pp. 372-382.

[12] T. Thein, M. M. Myo, S. Parvin, et al., "Reinforcement learning based methodology for energy-efficient resource allocation in cloud data centers," Journal of King Saud University-Computer and Information Sciences, vol. 32, no. 10, pp. 1127-1139, 2020.

[13] D. Cui, Z. Peng, J. Xiong, et al., "A reinforcement learning-based mixed job scheduler scheme for grid or IaaS cloud," IEEE Transactions on Cloud Computing, vol. 8, no. 4, pp. 1030-1039, 2017.

[14] A. Belgacem, S. Mahmoudi, and M. Kihl, "Intelligent multi-agent reinforcement learning model for resources allocation in cloud computing," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 6, pp. 2391-2404, 2022.