Optimization of Laboratory Equipment Sharing Scheduling Strategy Based on Reinforcement Learning

Huaying Hu*

Wuchang Institute of Technology, Wuhan, 430000, China
Corresponding author: dreamer0o0@126.com

Abstract: *The optimal scheduling of equipment resources in shared laboratories faces complex challenges including dynamic and uncertain demands, real-time changes in inventory status, and multi-objective trade-offs. To address the limitations of traditional scheduling methods in adapting to high-dimensional state spaces and dynamically evolving environments, this paper proposes an adaptive scheduling strategy optimization method based on deep reinforcement learning. First, this method constructs a dynamic scheduling model that integrates digital representations of multi-source heterogeneous equipment, time-window-constrained random demand sequences, and real-time state-aware inventory mapping, which provides structured state inputs for intelligent decision-making. Then, it designs a state space incorporating inventory features and a reward shaping function with multi-objective constraints, and it uses the proximal policy optimization algorithm to train a scheduling network, thereby generating an adaptive scheduling strategy capable of autonomous learning. Finally, it introduces a demand-aware module based on time-series forecasting and a collaborative pricing mechanism driven by remaining inventory, enabling dynamic tuning and closed-loop iteration of the scheduling strategy. Simulation experiments show that this method outperforms baseline methods in key indicators such as resource utilization, request satisfaction rate, and system stability, and it can effectively improve the operational efficiency and supply-demand matching accuracy of the laboratory equipment sharing system.*

Keywords: *reinforcement learning; laboratory equipment; sharing scheduling; digital representation; supply-demand awareness; proximal policy optimization*

Introduction

With the continuous deepening of scientific research and teaching activities, the demand for sharing laboratory equipment resources has shown a rapid growth trend. How to efficiently allocate multi-source heterogeneous resources, including precision instruments, general equipment, and consumables, while meeting the diverse needs of users and improving the turnover rate of equipment, has become a core issue in the field of laboratory operation management. Traditional scheduling methods are mostly based on static rules or heuristic strategies, and they are difficult to cope with the randomness of demand arrivals, the complexity of time window constraints, and the real-time fluctuations of inventory status. Deep reinforcement learning, with its advantages in sequential decision-making problems, can autonomously learn optimization strategies through continuous interaction with the environment, thus providing a new solution path for dynamic scheduling problems. The research on the optimization of laboratory equipment sharing scheduling strategies based on reinforcement learning aims to build an intelligent scheduling system with state awareness, adaptive decision-making, and dynamic tuning capabilities, which has important theoretical value and practical significance for improving resource utilization efficiency and ensuring user service levels.

1. Construction of a Dynamic Scheduling Model for Shared Laboratory Resources

1.1 Digital Representation of Multi-Source Heterogeneous Equipment Resources

To uniformly schedule multi-source heterogeneous resources, including precision instruments, general equipment, and consumables in a shared laboratory, this paper constructs a standardized digital representation framework. This framework structurally encodes the inherent attributes (functional classification, specification model, technical parameters) and dynamic attributes (current status, cumulative usage time, maintenance and calibration records) of the equipment, thereby forming feature vectors that can be parsed by algorithms. By introducing ontological concepts, this framework establishes the association relationships and compatibility rule bases among equipment items, laying a semantic foundation for intelligent matching and substitution recommendations. This representation method makes implicit usage constraints and transfer rules explicit, enabling the scheduling system to make reasonable decisions based on the intrinsic characteristics of the equipment.

The digital representation is further extended to a refined description of the spatiotemporal state. The time dimension records the available time periods, reservation occupancy, and maintenance periods of the equipment, thereby forming a continuous timeline state; the spatial dimension specifies the physical storage location, the current ownership status, and the expected transfer trajectory. The dynamic label integrating spatiotemporal information enables the scheduling algorithm to perceive the instantaneous state and future availability of the equipment in real time, transforms discrete physical resources into computable digital twins, and provides structured state inputs for the reinforcement learning agent[1].

1.2 Random Demand Based on Time Windows

User sharing requests usually contain explicit time constraints composed of the borrowing start time and the expected occupancy duration, thus forming a dynamic demand sequence with time windows. The randomness of this sequence is reflected in the distribution of request arrival times, the uncertainty of equipment types and quantities, potential conflicts among time windows, and the cancellation or modification of requests.

Under the time window framework, each demand request is assigned strict time constraints, and the quality of a scheduling strategy is reflected in its ability to reasonably allocate equipment while satisfying these time constraints. For divisible or substitutable demands, the modeling process needs to consider the flexible adjustment space of the time windows, allowing minor adjustments to the expected time windows to improve resource utilization. The random demand model integrated with time windows can generate a simulated data stream that closely resembles real scenarios, provides a dynamically evolving training environment for the reinforcement learning agent, emphasizes the coupling relationship between time resources and equipment resources, and ensures that the scheduling strategy learning process fully considers the impact of time constraints on decision-making.

1.3 Real-Time State-Aware Inventory Mapping Mechanism

The information delay between physical inventory changes and the digital scheduling system may cause decisions to be based on outdated inventory states, leading to resource conflicts or allocation failures. Therefore, this study constructs a real-time state-aware inventory mapping mechanism to achieve efficient synchronization between the physical world and the digital space. This mechanism relies on IoT perception technology, deploys intelligent sensing nodes in equipment storage units, collects inbound and outbound actions, in-position statuses, and borrowing and return records in real time, and thus forms a continuous data stream. Each physical operation triggers an immediate update of the inventory state and pushes the update to the scheduling system through an event-driven approach, ensuring that the available inventory view used for decision-making remains consistent with the physical reality.

The core of the inventory mapping mechanism lies in establishing a precise state synchronization protocol, which defines the complete link from the occurrence of a physical event to the update of the digital state, covering such steps as information acquisition, collation, state calculation, and information dissemination. According to the transfer characteristics of different types of equipment, this protocol sets differentiated synchronization strategies: it adopts a real-time synchronization mode for high-turnover general equipment to ensure information timeliness, and it adopts a regular calibration mode for low-frequency-use precision instruments to balance system resource consumption. This mechanism enables the scheduling system to check the accurate inventory surplus, the in-transit quantity, and the expected return time in real time, provides high-quality state data for decision-making input of the reinforcement learning algorithm, and ensures the feasibility and robustness of the scheduling strategy in a dynamic environment[2].

2. Adaptive Scheduling Strategy Generation Based on Deep Reinforcement Learning

2.1 Design of the Scheduling State Space Integrating Inventory Features

Under the deep reinforcement learning framework, the generation of a scheduling strategy relies on the agent' s accurate perception of the current environmental state. The design of the state space needs to encode the key information involved in scheduling decisions into a vector representation that can be processed by a neural network. For the laboratory equipment sharing scenario, the state space must not only include the features of the pending scheduling request sequence (such as the equipment type, quantity, and expected time window required by each request) but also deeply integrate real-time inventory features. The inventory features can be further decomposed into a static dimension and a dynamic dimension: the static dimension covers the total inventory quantity, specification parameters, and functional attributes of each type of equipment; the dynamic dimension reflects the real-time available inventory, the current quantity being borrowed, the expected return time distribution of each equipment item, and the remaining usage duration of the equipment currently in stock. By combining these two types of features into a high-dimensional state tensor, the agent can fully perceive the resource endowment and supply-demand situation at the current moment.

To improve the efficiency and interpretability of the state representation, this paper performs structured processing and feature engineering on the raw inventory data. For the expected return time of the equipment, this paper divides it into time segments to form recognizable temporal information; for the real-time available quantity of the equipment, this paper normalizes it in combination with its historical consumption rate to generate a relative indicator reflecting the degree of resource abundance. In addition, the state space should also embed the graph structure of substitution relationships among equipment items, so that the agent can understand what alternative resources can be used for allocation when a certain type of inventory is tight. This design of the scheduling state space integrating inventory features directly introduces the dynamic evolution law of inventory into the decision-making basis, enables the reinforcement learning agent to obtain a comprehensive understanding of the global resource situation at each decision moment, and provides information-rich input for the subsequent training of the policy network.

2.2 Immediate Reward Shaping under Multi-Objective Constraints

The learning objective of the reinforcement learning agent is defined by the reward function, and the design of the reward signal directly determines the optimization direction of the scheduling strategy. In laboratory equipment sharing scheduling, the optimization objective has typical multi-dimensional characteristics, and it needs to balance resource utilization efficiency, user service level, and system operation stability. Reward shaping needs to quantify these objectives into computable immediate feedback signals. Specifically, this paper defines that positive rewards originate from the successful allocation and circulation of equipment. For example, a single allocation action of equipment can obtain a reward value proportional to its contribution to utilization, and the on-time return of equipment can also trigger positive incentives. Negative rewards correspond to various conflicts or delays caused by scheduling decisions, such as the rejection of a request due to insufficient resources, time window conflicts caused by equipment allocation, and the situation where the waiting time of a user request exceeds the tolerance threshold[3].

The trade-off relationship among multiple objectives needs to be adjusted through a weighted combination of reward functions. The utilization-oriented reward can be designed as a function of equipment occupancy duration, which emphasizes resource turnover efficiency; the service-oriented reward can be linked to the request satisfaction rate and response speed, which reflects the level of response to user demands. To prevent the agent from excessively pursuing a single indicator at the expense of overall performance, this study introduces constrained penalty terms. For example, if the scheduling strategy causes equipment to be overly concentrated in a few long-term reservations, a fairness penalty for resource occupation can be imposed; if equipment is frequently scheduled to scenarios that exceed its functional thresholds, a decay penalty based on the equipment wear model can be applied. Through refined reward shaping, the agent gradually forms a perception of multi-objective priorities during its interaction with the environment, and it learns to generate trade-off optimized scheduling behaviors under complex constraint conditions.

2.3 Iterative Learning and Optimization of the Scheduling Strategy

Under the deep reinforcement learning framework, the generation of a scheduling strategy relies on the agent’ s accurate perception of the current environmental state. The choice of algorithm directly affects the learning efficiency and convergence effect of the scheduling strategy. This study adopts a policy iteration-based optimization method, which limits the magnitude of each policy update to ensure training stability while achieving efficient learning, and it is suitable for scheduling problems with continuous decision-making characteristics such as laboratory equipment sharing. During the training process, the policy network receives the current state input that integrates inventory features, and it outputs the equipment allocation scheme or the probability distribution of resource allocation for each pending scheduling request. The value network then estimates the value of the current state, which is used to evaluate the policy performance and provide an update direction. The two networks update cooperatively, enabling the agent to gradually approach the optimal scheduling strategy through its interaction with the simulation environment (which contains random demand sequences and dynamic inventory changes).

The training process requires special attention to the balance between exploration and exploitation. In the early stage, this study encourages the agent to extensively explore various allocation combinations by adding a certain degree of flexibility to action selection, so as to discover potential high-reward scheduling patterns. As training progresses, the strategy gradually converges to deterministic actions. The iterative optimization mechanism prevents the deviation of the policy caused by excessively large updates by controlling the magnitude of each single update, thereby ensuring the stability of training. In terms of training data generation, this study constructs a simulation environment containing a large number of random demand samples to simulate scheduling scenarios under different inventory levels and different demand intensities. Through repeated iterative training, the policy network learns the mapping from the complex state space to reasonable scheduling actions, and finally generates an adaptive scheduling strategy that can respond quickly in a real-time operating environment[4].

3. Dynamic Tuning Mechanism for the Scheduling Strategy Integrating Supply-Demand Awareness

3.1 Equipment Usage Prediction Based on Demand Patterns

The sharing demand for laboratory equipment usually exhibits significant temporal characteristics, including both regular fluctuations caused by teaching cycles and research rhythms and random disturbances driven by unexpected events. To improve the foresight and adaptability of the scheduling strategy, this paper constructs a demand-aware module based on time-series forecasting, which quantitatively estimates the request intensity of various equipment items in the near future. This module extracts the request frequency sequence of each equipment item within consecutive time windows based on historical reservation data, and it uses data analysis methods to extract the periodic features, change trends, and nonlinear fluctuation relationships. By combining the predicted demand intensity information with the real-time inventory status, this module jointly forms a reference basis for scheduling decisions, enabling scheduling arrangements to make more reasonable advance plans based on the prediction of future usage conditions.

The integration of the demand prediction results injects temporal insight into the scheduling strategy. When the prediction model outputs a signal that a certain type of equipment is about to enter a peak demand period, the agent can adopt a resource reservation strategy in the current decision-making process, appropriately controlling the quantity of immediate allocation to ensure supply during the peak period. Conversely, if the prediction shows that the demand in the future period is at a low level, the agent can moderately relax the allocation constraints and encourage the borrowing of equipment in the current period to improve turnover efficiency. This perception mechanism based on time-series forecasting enables the scheduling strategy to shift from passive response to active pre-configuration, achieves dynamic adaptation to demand fluctuations in an uncertain environment, thereby reduces the probability of resource conflicts caused by sudden demand peaks, and improves the overall supply-demand matching accuracy of the system.

3.2 Inventory-Guided Collaborative Allocation and Usage Adjustment

The remaining inventory level is a core indicator reflecting the degree of resource scarcity. When the real-time available inventory falls below a preset threshold, the system needs to activate an inventory-guided scheduling adjustment mechanism to cope with the resource-scarce state. The core of this mechanism lies in introducing a collaborative allocation strategy, that is, proactively providing users with alternative equipment of similar functions when resources are tight, thereby guiding demand diversion to alleviate the supply-demand contradiction. The generation of substitution recommendations depends on the association rule base established during the digital representation stage of equipment. The system comprehensively evaluates the substitutability of each candidate equipment based on the functional similarity among equipment items, usage compatibility, and historical substitution success rates. When a user request cannot be satisfied due to a shortage of the originally requested equipment, the system can incorporate the substitution recommendation list into the available options, decide whether and how to present the substitution alternatives, and update the inventory status and usage records after the user accepts the substitution, thereby achieving an improvement in capability from single-equipment allocation to multi-equipment collaborative allocation.

The other important dimension of the inventory-guided mechanism is the introduction of a usage adjustment method, which guides user behavior through incentive signals to achieve supply-demand balance. The usage adjustment is not based on real fees but is embedded in the scheduling system in the form of virtual points or usage priorities. When the remaining inventory is tight, the system can dynamically increase the borrowing point consumption of that equipment or reduce its allocation priority, thereby encouraging users to actively choose alternative equipment or adjust the borrowing time; when the inventory is abundant, the system lowers the usage threshold to attract demand inflow. This adjustment strategy cooperates with the overall objectives of the scheduling system, and the system treats the adjustment magnitude as a tunable decision element. By observing user responses to the adjustment changes (such as accepting substitutions, canceling requests, or modifying time), the system accumulates experience and gradually forms reasonable usage adjustment rules. This inventory-guided collaborative allocation and usage adjustment mechanism enables the system to flexibly guide demand-side behavior when resources are scarce, promotes utilization improvement when resources are abundant, and achieves dynamic coordination and optimization on both the supply and demand sides.

3.3 Closed-Loop Feedback-Based Policy Iteration and Efficiency Evaluation

The scheduling strategy needs to continuously evolve in a real-time operating environment to adapt to changes in demand patterns and resource structures. To this end, this paper constructs a closed-loop feedback-based policy iteration mechanism, which transforms the massive amounts of scheduling logs, inventory change records, and user behavior feedback generated during system operation into a data foundation for model optimization. This mechanism first establishes an offline evaluation dataset, periodically collecting diverse information such as historical scheduling records, request satisfaction statuses, waiting time distributions, and substitution acceptance rates. Using importance sampling or offline policy evaluation methods, this mechanism performs counterfactual inference on the performance of the current reinforcement learning policy on historical data, and it identifies the decision-making shortcomings of the policy under specific inventory levels and specific demand patterns. Based on the evaluation results, this mechanism fine-tunes the parameters of the policy network or retrains the model using newly collected data, thereby achieving incremental updates and continuous evolution of the policy.

Efficiency evaluation requires the construction of a multi-dimensional indicator system to quantify the optimization effectiveness of the scheduling strategy. The core evaluation dimensions include resource utilization efficiency (average occupancy rate of equipment, turnover frequency, and distribution of idle rate), service level (request satisfaction rate, average waiting time, proportion of overdue reservations, and substitution recommendation acceptance rate), and system stability (number of scheduling conflicts, frequency of inventory imbalance, and response delay during peak hours). At the same time, ablation experiments need to be conducted on the supply-demand awareness module, and the marginal contribution of each module to overall performance is quantified by comparing the performance of policy versions with and without time-series prediction and with and without pricing adjustment on a unified test set. The establishment of the closed-loop feedback mechanism ensures that the scheduling strategy can be dynamically tuned as the environment changes, achieves a positive cycle from data collection and policy evaluation to model updating, and provides continuous technical support for the long-term efficient operation of the laboratory equipment sharing system.

Conclusion

This study focuses on the laboratory equipment sharing scheduling problem, constructs a dynamic scheduling model that integrates digital representations of multi-source heterogeneous equipment, time-window-constrained random demand sequences, and real-time inventory mapping, proposes an adaptive scheduling strategy generation method based on deep reinforcement learning, and designs a dynamic tuning mechanism that integrates supply-demand awareness. By integrating inventory features into the state space, shaping the reward function to consider multi-objective constraints, and using proximal policy optimization to ensure training stability, this method enables the scheduling strategy to possess the ability to make autonomous decisions in complex dynamic environments. The introduction of the time-series forecasting module enhances the foresight of the strategy, the collaborative pricing mechanism driven by remaining inventory achieves flexible adjustment on both the supply and demand sides, and the closed-loop feedback iterative framework ensures the continuous evolution of the strategy. Future research can be further extended in the following directions: exploring a multi-agent collaborative scheduling framework to cope with distributed laboratory resource management scenarios, studying a personalized recommendation mechanism that integrates user preference learning to improve the acceptance rate of substitution recommendations, and introducing meta-learning ideas to enhance the rapid adaptability of the scheduling strategy in new scenarios.

References

[1] Wang Yanlin, et al. "Design of Information Management System for Open Laboratories in Universities." Digital Technology and Applications, vol. 41, no. 10, 2023, pp. 173-176+192.
[2] Wu Runqiang, et al. "Management of Electrical and Electronic Laboratory Equipment in Universities Based on Long-term Safety Mechanism." Laboratory Science, vol. 25, no. 6, 2022, pp. 178-180.
[3] He Yajing, and Chen Meng. "Construction of University Laboratory Safety Culture System Based on Informatization." Shandong Chemical Industry, vol. 50, no. 5, 2021, pp. 235-236.
[4] Xiong Jie, et al. "Design and Implementation of Open Laboratory Management Platform in Universities." Wireless Internet Technology, vol. 17, no. 6, 2020, pp. 50-52.