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: *There are many problems with the ideal scheduling of equipment resource in shared laboratories, such as unstable and irregular demand, constant fluctuation of inventory, and conflicting goals. The problem with the previous planning method is that it cannot handle a large number of states and changes in the environment; therefore, this paper proposes a way to improve adaptive planning based on deep reinforcement learning. Based on the above method, a dynamic scheduling model can be constructed to integrate digital representations of multi-source heterogeneous equipment, time-window-constrained random demand sequences, and real-time state-aware inventory maps, and thus obtain an organised state input for intelligent decision-making. Then, the state space of inventory features and a multi-objective reward-shaping function are constructed, and finally, the proximal policy optimisation algorithm is used to train a scheduling network that can learn independently and produce an adaptive scheduling strategy. Based on the results of the time series forecast, a demand response module will be added to change the schedule dynamically and be connected with the collaborative pricing mechanism for the remaining inventory. Based on the simulation experiments, the above method has improved the utilisation of resources, the rate of request satisfaction and system stability compared with the baseline method; therefore, it can increase the operating efficiency and supply-demand matching accuracy of the laboratory equipment sharing system.*

Keywords: *Reinforcement Learning, lab equipment, sharing schedule, digital representation, supply-demand awareness, Proximal Policy Optimization*

Introduction

With the development of science and education in recent years, the demand for sharing laboratory equipment resources among people has also been rising. How to distribute all kinds of resources reasonably, such as high-precision instruments, general equipment and consumables, to meet the different needs of users and improve the utilisation efficiency of equipment has become a problem in the field of laboratory operation management. Most of the traditional scheduling methods are based on fixed rules or heuristics, so they are not very good at dealing with the randomness of demand fluctuations, the complexity of time window constraints and real-time changes in inventory. Deep reinforcement learning is suitable for problems that need a series of choices, so it can learn to optimise by interacting with the environment many times, and now it is being applied to dynamic scheduling problems. Research on the optimisation of laboratory equipment sharing scheduling strategies based on reinforcement learning aims to build an intelligent scheduling system that is aware of the current situation, can make adaptive choices, and adjust dynamically; thus, it will be possible to improve the efficiency of resource utilisation and meet the service demands of all related parties.

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

1.1 Digital Representation of Multi-Source Heterogeneous Equipment Resources

In order to distribute all kinds of resources in the shared laboratory reasonably, such as high-precision instruments, general equipment, consumables, etc., this paper puts forward a unified digital representation system. The Structure of the framework organises the inherent attributes (functional division, specification model, technical parameters) and variable attributes (current state, accumulated operation time, maintenance and calibration data) of the equipment into feature vectors that can be employed by algorithms. Ontological concepts are used to build a database of associations and compatibility rules for equipment items in this framework, and thus a semantic basis for intelligent matching and substitution recommendations has been established. Therefore, we will know the hidden constraints and transfer rules, and the scheduling system can make reasonable decisions in line with the intrinsic characteristics of the equipment.

Further expansion of the digital representation is to improve the description of the spatio-temporal state. The Time dimension will be used to store the available time slots, reservation occupancy and maintenance periods of the equipment in a continuous timeline, and the Spatial dimension will show the physical storage location, current ownership status and expected transfer path. A dynamic label with spatiotemporal information can be added to inform the scheduling algorithm of the available time and place for a piece of equipment, and thus this physical resource can be converted into a computable digital twin and organised state input for the reinforcement learning agent[1].

1.2 Random Demand Based on Time Windows

User-shared request loans generally have certain time limits on when they can be used and for how long they will be in use; thus, they are dynamic demand sequences with time windows. There is randomness in the distribution of request arrival times, uncertainty about the types and quantities of equipment, possible conflicts among time windows, and also cancellations or modifications to requests.

All the demand requests in this time window have a strict deadline, and the quality of a scheduling plan can be judged by how reasonably it distributes the equipment under this time constraint. For divisible or substitute demand, it should be possible to flexibly change the range of time windows in the model during the modelling process, and small adjustments can be made to the expected time window to optimise resource utilisation. Random demand models and time windows can be used to generate simulated data that is close to real life; thus, a dynamic training environment will be created for the reinforcement learning agent, the interdependence of time resources and equipment resources will be shown, and the learning process of scheduling strategies will fully consider the impact of time constraints on decisions.

1.3 Real-Time State-Aware Inventory Mapping Mechanism

If there is a delay in updating the information of physical inventory and the digital scheduling system, then the decisions will be based on old inventory data, and there will be resource conflicts or allocation failures. Therefore, this paper will construct an actual-time, state-aware inventory map to achieve good coordination between the physical world and the digital world. The above way is to build intelligent sensing nodes based on IoT perception technology in the equipment storage area, continuously collect real-time data on people's in-and-out activities, positions, borrowing and returning of items, etc., generate a continuous stream of data, and so on. After each physical operation, the inventory status will be updated promptly, and at the same time, an event will be sent to the scheduling system for notification; thus, the available inventory information used for decision-making will always be in line with the actual stock.

The first need for an inventory mapping scheme is to accurately link the current situation of a real-world event with changes in related digital data at all times during the process of information collection and organisation, state determination and dissemination of results. All kinds of equipment have different transfer characteristics, so many synchronization schemes have been developed; high-turnover general equipment operates in real time and needs to be quickly updated with information, so a high-frequency calibration mode is used for low-frequency applications of high-precision instruments to balance system resource consumption. In this way, the scheduling system will know how much inventory surplus there is, how much is in transit, and when it will be available again; therefore, good status data will be provided for the decision-making part of the reinforcement learning algorithm, and the scheduling plan will be practical and stable in a changing environment[2].

2. Adaptive Scheduling Strategy Generation Based on Deep Reinforcement Learning

2.1 Design of the Scheduling State Space Including Inventory Features

The agent should have good knowledge of the current state of its environment and then use deep reinforcement learning to plan how to act. The Design of the state space should contain all the relevant information of scheduling decisions in a vector form that is suitable for neural networks. The problem of shared laboratory equipment should be modelled in the state space by adding not only the attributes of the sequence of pending scheduling requests (e.g., equipment type, quantity and required time window for each request) but also the real-time inventory status. The attributes of inventory can be divided into a fixed part and a variable part; the fixed part includes the total amount of inventory, the specification parameters and function attributes of various types of equipment, and the variable part includes the current inventory quantity, how many items have been borrowed at present, the expected return time distribution for each piece of equipment, and the remaining usage period of equipment that is currently in stock. Combine the two kinds of features to form a high-dimensional state tensor, and then obtain all the resource, supply and demand conditions at the current time.

To improve the efficiency and interpretability of the state representation, structured processing and feature engineering are carried out on the raw inventory data in this paper. To make the expected return time of the equipment more obvious, this paper divides it into time periods; to obtain a relative index of resource abundance for the real-time available quantity of the equipment, this quantity is normalised with respect to the historical consumption rate. In addition, the state space should also include the graph structure of substitution relationships among equipment items so that the agent knows which alternative resources can be used for allocation when there is a shortage of a particular type of inventory. The design of the scheduling state space directly includes the inventory information, and thus the dynamic evolution law of inventory is added to the basis of decision-making; therefore, the reinforcement learning agent can obtain complete knowledge of the global resource situation at each decision time and obtain rich input for the next training of the policy network.

2.2 Immediate Reward Shaping Under Multi-Objective Constraints

The learning goal of the reinforcement learning agent is set by the reward function, so the Design of the reward signal will direct the optimisation direction of the schedule. The objective of optimising the lab equipment sharing schedule is to improve the utilisation efficiency of resources and meet all sides of the demand for service quality and system stability. Reward shaping can be used to add some numerical rewards at different times. That is to say, the positive reward is the result of successfully distributing and using the equipment. For example, one allocation operation of a piece of equipment can be given a reward based on how much it is used, and returning the equipment promptly can also be rewarded positively. Negative rewards are various kinds of conflicts or delays in the scheduling decision-making process; for example, a request may be rejected due to a lack of resources, there may be a time-window conflict caused by equipment allocation, and a person's waiting time for a service can exceed the allowed limit[3].

All the goals have to be balanced, so different weights will be given to the reward functions. The utilisation-oriented reward can be set as a function of how long the equipment has been used, and the service-oriented reward can be based on the level of user satisfaction with their requests and the speed of service response. To avoid having too much emphasis on a single indicator and ignoring others, constrained penalty terms are added in this paper. For instance, if the result of the scheduling plan is that many pieces of equipment need to be reserved for a long time, a fairness penalty for resource occupation can be added; if the equipment is used frequently under conditions that exceed its design parameters, a decay penalty based on the equipment wear model can be employed. Through reward shaping, it gradually learns which reasons to consider when interacting with the world, and thus can learn a schedule that is good for all of them.

2.3 Iterative Learning and Optimisation of the Scheduling Strategy

Deep reinforcement learning is employed to develop a schedule, and in order to do so, one needs to know the current state of the environment. The Choice of the Algorithm will directly affect the learning efficiency and convergence of the scheduling method. To limit the range of each policy update and maintain the stability of training and good learning results, Policy Iteration Optimisation is employed in this paper; thus, it is suitable for scheduling problems that require continuous decision-making, such as those of shared laboratory equipment. The current state of the policy network is given the current inventory features during training, and then it outputs either the equipment allocation plan or the probability distribution of resource allocation for each pending scheduling request. The value network then estimates the value of the current state, and this is used to assess the performance of the policy and determine the direction of update. The two networks will be combined in the simulation to determine the best way to handle the random demand and inventory fluctuations in the simulation.

Training will be carried out with the dual purposes of exploration and exploitation. At the beginning of this study, some freedom has been added to the selection of actions to help the agent try out many different distribution combinations and find a good schedule with a high reward. With the progress of training, the strategy will be more certain. By changing the upper bound of the policy update in each round, we can reduce the size of the update and prevent large fluctuations in the policy; thus, training will be more stable. A simulation environment is employed to generate a large number of random demand samples for the creation of training data, and it can also simulate scheduling under various inventory levels and different degrees of demand. After many rounds of iterative training, the policy network has been able to map the large and complex state space to reasonable scheduling actions, and finally, it can output an adaptive scheduling strategy that performs well in real time[4].

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

3.1 Equipment Use Prediction Based on Demand Patterns

Generally speaking, the demand for shared laboratory facilities is periodic; it will change according to changes in the schedule of teaching and research, and will rarely occur irregularly due to unforeseen circumstances. To improve the accuracy of the prediction and the timeliness of the schedule, a demand-aware module based on time-series forecasting has been added in this paper to forecast the demand for all kinds of equipment in the near future. Extract the sequence of request frequency for each equipment item in consecutive time windows from the historical reservation data, and then use all sorts of data analysis methods to obtain periodic features, change trends and nonlinear fluctuation relationships. According to the predicted demand intensity information obtained in the above module, combine it with the current inventory status to form a reference basis for scheduling decisions, and then make more reasonable advance plans for scheduling in light of the forecast of future utilisation conditions.

Add the demand forecast results to the schedule and give it a sense of time. When the prediction model finds that a particular kind of equipment will be in high demand at a certain time, the agent can choose a resource reservation plan in its current decisions to increase the amount reserved in advance and ensure that there is sufficient supply during this peak period. At the same time, if the forecast shows that the demand in the future period will be relatively low, the agent can moderately loosen the allocation constraints and encourage the borrowing of equipment in the current period to improve turnover efficiency. Based on the results of the time series forecast above, the perception mechanism will be changed from passive response to active pre-configuration; thus, it can be dynamically adjusted according to changes in uncertain demand to reduce resource conflicts caused by sudden increases in demand and improve the overall supply-demand matching accuracy of the system.

3.2 Inventory-Guided Collaborative Allocation and Usage Adjustment

The remaining stock level is a kind of shortage in the company's resources. When the amount of stock available at any time falls to a low level, inventory-driven rescheduling will be carried out to deal with shortages of resources. At the bottom of this mechanism is cooperative allocation; that is to say, when resources are scarce, we will proactively offer users other equipment that can achieve the same effect to shift demand and thus narrow the supply-demand gap. The generation of substitution recommendations is based on the association rule set that was constructed in the digital representation stage of equipment. Based on the function similarity of the equipment items, compatibility and the success rate of historical substitution will be used to evaluate the substitutability of all candidate equipment. If the required item cannot be obtained because the original equipment is unavailable, a list of alternative recommendations will be added to the list of options, and it will be decided whether and how to display the substitute choices; after the user selects a substitute, the inventory status and usage record will be updated to expand the function from single-equipment allocation to multi-equipment collaborative allocation.

Another kind of inventory-guided mechanism is to provide a way to change use that motivates people to alter their behaviour and thus regulates the supply and demand of goods. It will not be significantly more expensive; rather, it will be added to your virtual points or given higher priority in the schedule system. If the stock is too low, the system will automatically increase the borrowing point consumption for that equipment or reduce its allocation priority, and the user will have to choose another one or extend the borrowing period; if the stock is too high, the usage threshold will be lowered to increase demand. Therefore, it will be in line with the general purpose of the scheduling system and the range of modification can be set by users. Observe how the users are affected by the change in plan, and according to the above data, reasonably adjust the service schedule for both users and companies. Inventory-guided collaborative allocation and utilisation adjustment can flexibly change the behaviour of demand in times of resource shortage to make better use of abundant resources and achieve dynamic coordination and optimisation of both supply and demand.

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

According to changes in the operating environment and demand at different times, the schedule of operation will need to be adjusted promptly. Therefore, in order to make full use of the above data, a closed-loop feedback-based policy iteration mechanism will be constructed, and a large amount of scheduling logs, inventory change records and user behaviour feedback obtained from the operation of the system will be used as the data foundation for model optimisation. The first is to build an offline evaluation set and regularly collect all kinds of data, such as old schedule records, request satisfaction status, waiting time distribution, substitution acceptance rate, etc. Importance sampling and offline policy evaluation methods are used to perform counterfactual inference on the performance of the current reinforcement learning policy in historical data to identify deficiencies in the decision-making of the policy at certain inventory levels and demand patterns, etc. Based on the evaluation results above, the parameters of the policy network will be updated or the model will be retrained with the new data to achieve incremental updates and continuous improvement of the policy.

To know how good a schedule is, some criteria for evaluation should be set, and then scores should be given. The three main parts of the general assessment are: Resource Utilisation Efficiency (average equipment utilisation rate, turnover frequency and distribution of idle rate), Service Level (such as the proportion of satisfied customers and average waiting time for service), and System Stability (number of scheduling conflicts, frequency of inventory imbalance and response delay during peak hours). At the same time, ablation studies will also be carried out on the supply-demand awareness module, and the marginal contribution of each module to the overall performance can be determined by comparing the results of policy versions with and without time-series prediction, and with and without pricing adjustment on a unified test set. A closed-loop feedback mechanism will be added to change the schedule according to changes in the environment, and in the long run, a positive cycle of data collection, policy evaluation, model optimisation and continuous improvement for the technical support and good operating conditions of the laboratory equipment sharing system will be established.

Conclusion

First, there is a problem in the scheduling of laboratory equipment sharing; second, based on the digital representation of multi-source heterogeneous equipment, a dynamic scheduling model will be built that considers time-window-constrained random demand sequences and real-time inventory mapping, and finally, an adaptive scheduling strategy generation method based on deep reinforcement learning and a dynamic tuning mechanism for supply and demand will be introduced. Add an inventory function to the state space, design a multi-objective reward function, use proximal policy optimisation to maintain the stability of training, and thus the scheduling strategy will be able to make autonomous decisions in complex and dynamic environments. Add a time-series forecasting module to increase the lead time of the strategy, dynamically adjust the collaborative pricing mechanism according to the remaining inventory at both ends of the supply and demand, and set up a closed-loop feedback iteration framework for continuous improvement of the strategy. The next three research directions will be to build a multi-agent collaboration and scheduling framework for the management of distributed laboratory resources; to design a personalised recommendation method that learns from the behaviour of users to increase the acceptance rate of substitution suggestions; and to add meta-learning to improve the rapid adaptation ability of scheduling strategies in light of changes in circumstances.

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.