ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Application of Deep reinforcement Learning in dynamic allocation of public resources and optimization of administrative efficiency

Xinyan Li[1,*]

[1] Beijing Jingbei Vocational and Technical College, Huairou 101400, Beijing, China

SUMMARY: *The dynamic allocation of public resources faces challenges such as fluctuating demand, complex departmental coordination and frequent emergencies. Traditional static rules are difficult to balance resource utilization, response efficiency and regional fairness in a timely way. In this paper, a collaborative scheduling model combining digital twin, spatio-temporal supply and demand graph and deep reinforcement learning is constructed to transform the processes of community service, medical response, traffic connection, emergency supplies and administrative service windows into sequential decision-making problems. Method, the multi-scenario simulation is completed through the digital twin environment, and the graph neural network is used to encode regional demand, resource inventory, department load and event risk. A multi-objective reward function including task completion, resource utilization, fairness, response delay and scheduling cost is designed, and the policy update is realized by combining online feedback calibration. The experimental results show that the resource utilization rate of the proposed method reaches 89.7%, the task completion rate reaches 93.6%, the average response time is reduced to 18.6 min, the comprehensive performance is restored to 0.91 after the burst disturbance, and the task completion rate of cross-region migration remains above 92.0%. The research shows that this method can provide technical support for the fine allocation of public resources and the intelligent optimization of administrative efficiency.*

KEYWORDS: *Deep reinforcement learning; Dynamic allocation of public resources; Administrative efficiency optimization; Multi-objective decision optimization*

1 Introduction

1.1 Research background of dynamic allocation of public resources and optimization of administrative efficiency

Public resource allocation involves various scenarios such as government service windows, emergency supplies, public transportation, medical assistance, community services, and urban infrastructure maintenance, and its operation effect directly affects the speed of administrative response, service fairness, and social governance efficiency. Traditional public resource allocation relies on fixed rules, manual experience and periodic statistical reports, which can meet the needs of conventional management. However, in the environment of obvious demand fluctuations, complex department coordination and frequent emergencies, it is easy to have problems such as resource idling and local shortage, too long response chain, and lagging scheduling strategy. For example, there are significant differences in population

*lxy2010324@126.com
https://doi.org/10.65102/is20261035
Published: 30 April, 2026

density, service demand, traffic conditions, and incident risk in different regions, and it is difficult for static allocation schemes to reflect supply and demand changes in a timely manner, which limits the efficiency of administrative services. With the construction of government digitalization, Internet of things perception, urban big data platform and intelligent decision-making system, a large amount of spatio-temporal data, business flow data and feedback data have been accumulated in the operation process of public resources, which provides a data basis for algorithm-driven dynamic optimization.

1.2 Research significance of deep reinforcement learning driving intelligent decision-making in public management

Deep reinforcement learning combines the feature expression ability of deep neural network with the sequential decision-making ability of reinforcement learning, which is suitable for dealing with the continuous state change, multi-objective constraints and long-term revenue optimization problems in the dynamic allocation of public resources. In public management scenarios, resource scheduling decisions are often not made once, but iterated with demand arrival, task execution, feedback update, and policy modification. By constructing the state space, action space and reward function, the deep reinforcement learning model can interact repeatedly in the simulation environment or digital twin system to learn the impact of different resource allocation strategies on task completion rate, average response time, resource utilization and service fairness. Compared with traditional heuristic algorithms and static optimization methods, deep reinforcement learning can adjust the strategy according to environmental feedback, and form a more adaptive scheduling scheme in peak demand, sudden disturbance and cross-department collaboration scenarios. This approach is helpful to push administrative management from post-statistical analysis to process prediction, real-time scheduling and intelligent optimization.

1.3 Main Contributions and research objectives of this paper

Focusing on the problem of dynamic allocation of public resources and optimization of administrative efficiency, this paper constructs an intelligent scheduling framework that integrates digital twin, spatio-temporal supply and demand graph and deep reinforcement learning. The goal of this paper is to transform the operation process of public resources into a computable, trainable and verifiable sequential decision problem, and use multi-source data to describe the relationship among regional demand, resource load, department response and service feedback. On this basis, a reward mechanism is designed to balance efficiency, fairness and response cost. The main contributions of this paper are reflected in three aspects. Firstly, a digital twin environment for public resource operation scenarios is constructed to provide a training space for deep reinforcement learning models with repeatable simulation. Secondly, the resource state coding was completed based on the spatio-temporal supply and demand graph, so that the model could identify the dynamic coupling relationship between different regions and departments. Thirdly, a deep reinforcement learning policy iteration method for cross-department collaborative scheduling is designed, and the effect of the model is verified by the indicators of common resource utilization, task completion rate, administrative response time, robustness and cross-regional generalization ability. Through the above research, this paper aims to provide a scalable computer intelligent decision-making path for the fine allocation of public resources and the improvement of administrative efficiency.

2 Literature Review

Existing research shows that deep reinforcement learning has become an important intelligent optimization method in complex resource management problems. Alwarafy et al. focused on the resource management of future heterogeneous wireless networks, summarized the application of deep reinforcement learning in spectrum, power, access and computing resource allocation, and pointed out that it can improve system adaptability through interactive learning in dynamic environments, but it still faces the problems of high state dimension, insufficient training stability and high actual deployment cost [1]. Abouleneen et al. further introduced deep reinforcement learning into UAV Internet network and emphasized that resource scheduling in the air-ground collaborative environment should consider communication links, energy consumption, task load and network topology changes at the same time, which provides a reference for cross-regional dynamic scheduling of public resources [2]. Poltronieri et al. proposed MECForge scheme to realize edge computing resource management through deep reinforcement learning, and proved that agents can adjust resource allocation strategies according to task value and system load [3]. Sellami et al. studied task scheduling and offloading in SDN-IoT network and SDN-Blockchain 5G massive iot edge network respectively, indicating that deep reinforcement learning has advantages in reducing energy consumption, improving task execution efficiency and alleviating edge node congestion [4, 5]. Danino et al. used multi-agent deep reinforcement learning to deal with the container allocation problem in cloud environment, which reflected the applicability of multi-agent collaborative decision-making for complex resource pool management [6]. Kougioumtzidis et al. proposed a resource allocation method based on deep reinforcement learning for wireless VR communication, whose core is to dynamically adjust communication resources according to the quality of user experience [7]. Murareșu et al. combined agent simulation with reinforcement learning for dynamic resource allocation in mass casualty events, indicating that such methods have begun to extend to public emergency governance scenarios [8].

In the field of transportation and urban operation, deep reinforcement learning is also widely used in dynamic decision-making problems related to administrative efficiency. Haydari and Yılmaz systematically sorted out the application of deep reinforcement learning in intelligent transportation systems, covering vehicle scheduling, path planning and traffic control [9]. Noaeen et al. reviewed urban traffic signal control and pointed out that reinforcement learning can adjust signal strategies according to real-time traffic flow, thereby reducing congestion and waiting time [10]. Waqar et al., Park and Lim introduced deep multi-agent reinforcement learning in NOMA-MEC and 5G Internet of vehicles resource allocation, respectively, to jointly optimize communication resources, vehicle clustering and edge computing capabilities [11, 12]. Mohebifard and Hajbabaie used deep reinforcement learning to carry out urban road flow metering control, and proved that this method could improve the efficiency of traffic allocation in connected urban networks [13]. Saadi et al. summarized the research on reinforcement learning and deep reinforcement learning in the collaborative control of intelligent traffic lights, and emphasized that multi-intersection collaboration and real-time feedback mechanism are the key to improving the efficiency of urban operation [14]. Ali et al., Ergun verified the adaptability of deep reinforcement learning in a highly dynamic network environment from the perspectives of fog computing task offloading and vehicle Internet of vehicles communication resource optimization [15, 16].

Researches on edge computing, network slicing and smart city service deployment have further expanded the technical foundation of public resource scheduling. Hurtado Sanchez et

al. pointed out that network slicing resource management requires a dynamic trade-off among bandwidth, delay and service quality, and deep reinforcement learning can provide support for online decision-making under complex constraints [17]. Aghapour et al., Lim et al., Hoang et al., respectively studied task offloading and online resource management in iot edge computing, mobile edge computing, and Uavs assisted edge computing, showing that deep reinforcement learning is suitable for processing continuous arrival tasks and real-time load changes [18-20]. Naderializadeh et al. proposed a multi-agent resource management framework for wireless networks, which provided algorithmic inspiration for the collaborative scheduling of multi-department public resources [21]. Kasi et al. studied the placement problem of secure mobile edge servers, indicating that both security and service coverage should be taken into account in resource deployment [22]. Bansal et al. proposed UrbanEnQoSPlace model for real-time iot application service deployment in smart cities, which reflects the value of deep reinforcement learning in urban public service quality assurance [23]. Sami et al. proposed a demand-driven fog node and service placement method, emphasizing that resource allocation should be dynamically adjusted with demand changes [24]. In summary, existing studies have proved that deep reinforcement learning has strong adaptability in communication, transportation, edge computing and smart city resource management, but the unified modeling of fairness, response cost, cross-departmental collaboration and cross-regional generalization ability in public administration scenarios is still insufficient. Therefore, it is necessary to construct a deep reinforcement learning framework for public resource dynamic allocation and administrative efficiency optimization.

3 Proposed methods

3.1 Construction method of digital twin environment for public resource operation scenario

In order to provide a simulatable, trainable and deployable operational basis for the dynamic allocation of public resources, this paper constructs a digital twin environment for administrative service scenarios. The environment takes community service center, medical response unit, traffic connection node, emergency material warehouse and administrative department as the core entities, and maps the resource supply, task arrival, regional demand and execution feedback in the real system to the virtual space. In the data layer, the system accesses iot perception data, government logs, GIS spatio-temporal information, event streams and resource inventory snapshots. In the computing layer, entity mapping, state synchronization, rule-driven update and scene deduction are completed. Thus, a closed-loop structure of "real resources-virtual mirror-policy optimization-feedback calibration" is formed.

In order to further illustrate the organization of digital twin environment in the dynamic allocation of public resources, this paper integrates real public resource entities, real-time data collection, spatio-temporal feature fusion, virtual mirror modeling, policy simulation and scheduling execution into a hierarchical architecture. The overall architecture is shown in Figure 1.

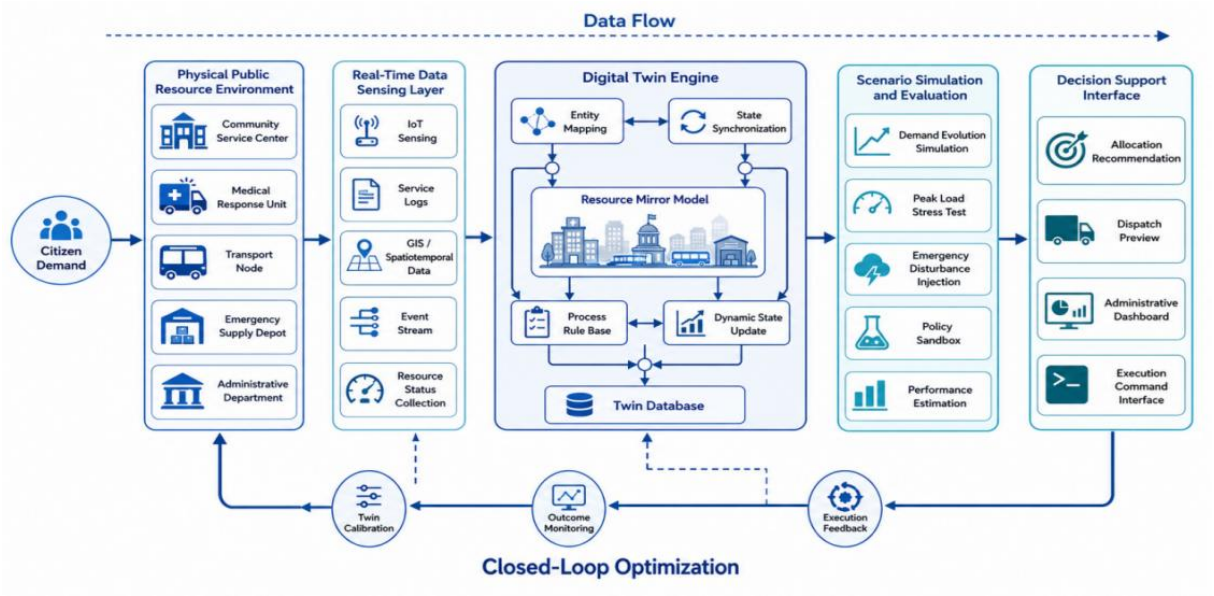

Figure 1: Architecture diagram of digital twin environment with dynamic allocation of public resources

It can be seen from Figure 1 that the digital twin environment does not stay at the static resource display level, but through the hierarchical structure of "physical resource layer, data fusion layer, twin modeling layer and deep reinforcement learning decision layer", the running state of public resources is transformed into a dynamic system that can be calculated, inferred and fed back. The right performance monitoring module continuously tracks resource utilization, task completion rate, response time, fairness, and scheduling cost, enabling the model to calibrate the virtual environment and update policy parameters based on execution results. This structure provides a stable simulation foundation for subsequent cross-departmental resource co-scheduling and administrative efficiency optimization.

In order to describe multi-source heterogeneous data uniformly, let the original observation vector at time t be o_t and the scene context vector be c_t, then the digital twin environment representation vector z_t is defined as follows:

$$z_t = \tanh(W_o o_t + W_c c_t + b_z) \tag{1}$$

where, W_o and W_c represent the observation feature mapping matrix and context mapping matrix, respectively, and b_z is the bias term. z_t is used to carry a unified expression of resource states, regional pressures, and service loads, which provides the input basis for subsequent reinforcement learning decisions.

The digital twin environment not only requires the representation to be complete, but also requires the virtual and real states to be consistent. To this end, the twin calibration error function is introduced:

$$\Delta_t = \|\mathcal{M}(\mathbf{z}_t) - \mathbf{y}_t\|_2^2 + \lambda_d \|\mathbf{z}_t - \mathbf{z}_{t-1}\|_2^2 \tag{2}$$

Here, $\mathcal{M}(\mathbf{z}_t)$ represents the mirrored output by the twin model, y_t represents the true feedback returned by the real system, and λ_d is the time smoothing coefficient. On the one hand, this formula constrains the approximation degree of the virtual environment to the real running state, on the other hand, it avoids the excessive update of the state, which is conducive to improving the stability of the simulation environment.

In the scenario deduction stage, this paper further constructs a dynamic evolution mechanism for the fluctuation of public resources. Let gt be the load vector of simulation scenario, ut be the external disturbance input, pt be the resource adjustment quantity, then the scenario evolution result at the next moment is expressed as follows.

$$g_{t+1} = \Phi(g_t, u_t, p_t) \tag{3}$$

Here, Φ(·) represents the state transition function in the digital twin environment, which can describe the peak service arrival, emergency injection, resource occupancy change and recovery process. Through this mechanism, the system can complete demand growth simulation, emergency impact test and scheduling strategy preview in the virtual space, which provides a safe, continuous and repeatable training experimental environment for subsequent deep reinforcement learning models. On the whole, the digital twin environment transforms the operation process of public resources into a computable dynamic system, which lays a model foundation for administrative efficiency optimization.

3.2 State representation and feature coding of public resources based on spatio-temporal supply and demand graph

The key to the dynamic allocation of public resources is to accurately describe "where the demand appears, where the resources are idle, and how the departments interact with each other". In order to avoid the lack of spatial relationships caused by describing resource states only by table variables, this paper constructs a spatio-temporal supply and demand graph, which abstracts service areas, resource sites, administrative departments and event occurrence points as graph nodes, and abstracts transportation accessibility, business flow relationships and cross-department collaboration relationships as graph edges. Let the spatio-temporal supply and demand graph at time t be represented as follows:

$$\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \mathbf{A}_t, \mathbf{X}_t) \tag{4}$$

where $\mathcal{V}_t$ is the set of nodes,$\mathcal{E}_t$ is the set of edges, A_t is the neighboring weight matrix, and X_t is the node feature matrix. The node characteristics are composed of population density, task arrival rate, resource inventory, department load, historical response time and incident risk level, which are used to describe the service pressure of different regions at the same time instant.Graph edge weight not only considers spatial distance, but also introduces resource complementarity and administrative collaboration efficiency.

In order to more intuitively show the components of the spatio-temporal supply and demand graph of public resources and the feature encoding process, this paper visually represents the node types, edge relationships and the generation process of graph neural network states, as shown in Figure 2.

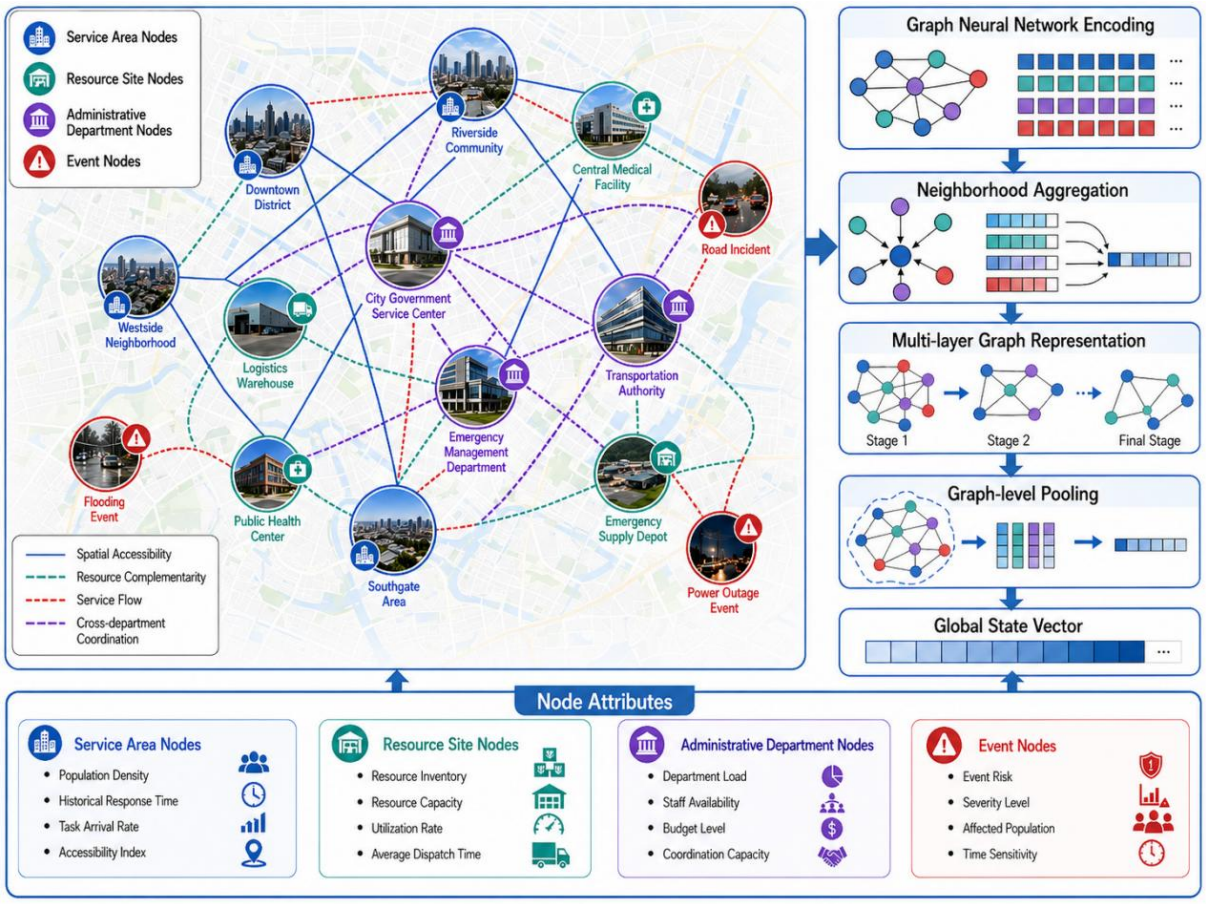

Figure 2: Schematic diagram of the structure and feature encoding of the spatio-temporal supply and demand graph of public resources

As can be seen from Figure 2, service area nodes, resource site nodes, administrative department nodes and event nodes jointly constitute the spatio-temporal supply and demand graph of public resources, and spatial accessibility, resource complementarity, business flows and cross-department collaboration relationships form multi-type associated edges. On this basis, the graph neural network achieves graph state compression through neighborhood aggregation, multi-layer representation learning and graph-level pooling, and finally generates a global state vector for reinforcement learning decisions, so as to provide structured input for subsequent cross-department collaborative scheduling.

The edge weight between node i and node j is defined as follows.

$$A_t(i,j) = \exp\left(-\frac{\ell_{ij}}{\tau_\ell}\right) \cdot (1 + \kappa_1 q_{ij,t} + \kappa_2 m_{ij,t}) \tag{5}$$

where, ℓ_{ij} represents the passing distance between two nodes, τ_ℓ is the spatial attenuation coefficient, $q_{ij,t}$ represents the resource complementarity intensity, $m_{ij,t}$ represents the department collaboration frequency, and κ_1 and κ_2 are the regulation coefficients. This design enables neighboring regions, complementary resources, and highly collaborative departments to obtain stronger associations in the graph.

In order to reflect the degree of regional supply and demand tension, this paper constructs the node supply and demand offset index:

$$\rho_{i,t} = \frac{d_{i,t} - a_{i,t}}{a_{i,t} + \varepsilon_a} + \mu_r v_{i,t} \tag{6}$$

Here, $d_{i,t}$ represents the real-time demand intensity of node i, $a_{i,t}$ represents the amount of available resources, ε_a is the stability term that prevents the denominator from being zero, $v_{i,t}$ represents the event risk intensity, and μ_r is the risk correction coefficient. When $\rho_{i,t}$ increases, it indicates that there is a higher demand for resource compensation in this region.

In the feature encoding stage, graph neural network is used to aggregate neighborhood information, so that a single regional state can absorb surrounding resources and cross-departmental collaboration information. Layer l node representation is updated as follows:

$$h_{i,t}^{(l+1)} = \sigma\left(B^{(l)} h_{i,t}^{(l)} + \sum_{j \in \mathcal{N}(i)} \omega_{ij,t}^{(l)} C^{(l)} h_{j,t}^{(l)}\right) \tag{7}$$

where $h_{i,t}^{(l)}$ is the layer l node embedding,$\mathcal{N}(i)$ is the set of neighbor nodes, $\omega_{ij,t}^{(l)}$ is the neighborhood aggregation weight, $B^{(l)}$ and $C^{(l)}$ are trainable parameter matrices and $\sigma(\cdot)$ is a nonlinear activation function. The encoding method can identify the impact of local congestion, regional overflow and departmental linkage on the resource scheduling state.

In order to provide unified input to subsequent deep reinforcement learning models, all node embeddings, supply-demand offset indicators and time characteristics are fused to obtain the global public resource state vector:

$$x_t = \mathrm{Concat}\left(\mathrm{Pool}\left(\{h_{i,t}^{(L)}\}_{i=1}^{N_t}\right), \rho_t, \theta_t\right) \tag{8}$$

Here, Pool(·) represents the graph-level pooling operation, $h_{i,t}^{(L)}$ is the final layer node embedding, N_t is the number of nodes, ρ_t is the supply-demand offset sequence, and θ_t is the temporal context feature. Through the above encoding, the public resource state is transformed from decentralized business data into a graph state expression that can be directly recognized by the deep reinforcement learning model, which provides structured input for subsequent cross-department collaborative scheduling decisions.

3.3 Cross-department resource collaborative scheduling decision model with deep reinforcement learning

In the dynamic allocation of public resources, departments such as medical care, transportation, government window and emergency supplies are not independent units, and the resource occupation of one department will affect the response ability of other departments. In order to improve the collaboration of cross-department scheduling, this paper constructs a multi-department collaborative decision-making model based on deep reinforcement learning, which considers each department as an agent with local observation ability, and generates a unified scheduling scheme through shared state coding, department policy network and joint action arbitration mechanism. The global public resource state vector χ_t obtained in 3.2 is used as the model input, and the action is inferred by combining the local operating state of each department itself.

Let the local observation vector of the KTH department at time t be $l_{k,t}$, and the department policy network parameter be φ_k. Then the candidate scheduling action vector generated by the department is as follows:

$$\xi_{k,t} = \mathrm{Softmax}\left(F_{\varphi_k}([\chi_t; l_{k,t}])\right), \quad k = 1,2,\ldots,K \tag{9}$$

where $\xi_{k,t}$ represents the resource scheduling proposal output by department k, $F_{\varphi k}(\cdot)$ is the department policy subnetwork, and K represents the number of departments involved in co-scheduling. The design preserves the local decision-making ability of the department, and avoids the isolated scheduling of each department only according to its own load. Figure 3 shows the structure of the deep reinforcement learning model for cross-departmental common resource co-scheduling.

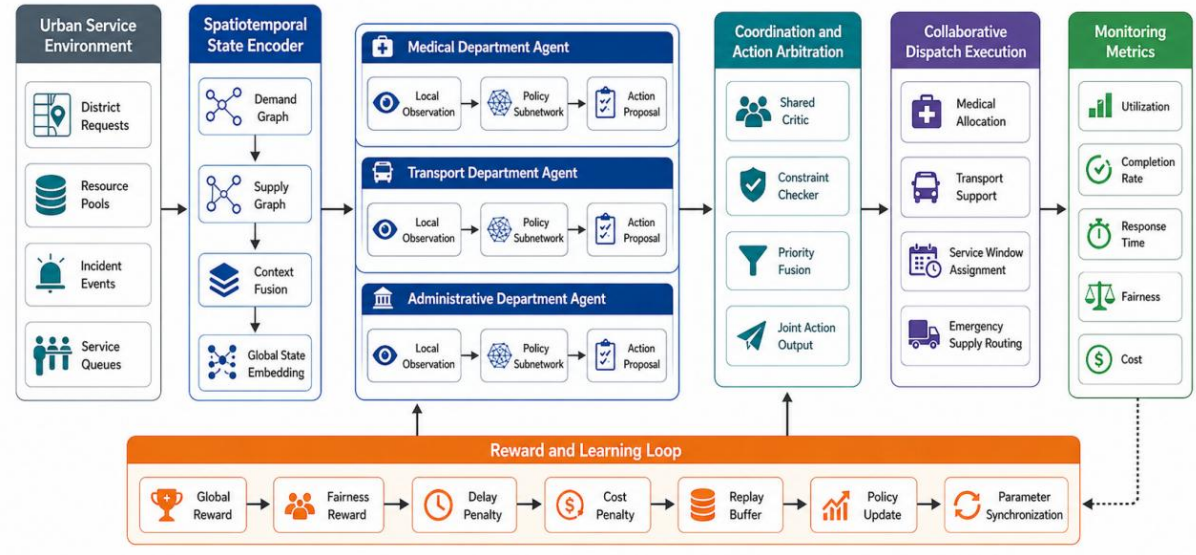

Figure 3: Structure diagram of deep reinforcement learning model for cross-departmental public resource co-scheduling

As can be seen from Figure 3, the spatio-temporal state encoder compresses the regional demand graph, resource supply graph, and context information into global state embeddings, which are fed into the medical, transportation, and administrative department agents, respectively. The agents of each department generate local action suggestions, and then the collaborative arbitration module performs constraint checking, priority fusion and joint action output. The execution results enter the monitoring indicator module and are fed back to the policy network through the reward learning loop to realize the continuous optimization of the cross-department scheduling strategy.

In the action arbitration phase, the action proposals of different departments need to be dynamically weighted according to the current demand urgency, department load and resource complementarity. Let βk,t be the action credibility weight of department k and ζt be the joint scheduling action, then we have:

$$\beta_{k,t} = \frac{\exp\left(n_a^{\top} \tanh\left(P_a[\chi_t; l_{k,t}]\right)\right)}{\sum_{g=1}^{K} \exp\left(n_a^{\top} \tanh\left(P_a[\chi_t; l_{g,t}]\right)\right)}, \quad \zeta_t = \mathrm{Proj}_{\Omega_t}\left(\sum_{k=1}^{K} \beta_{k,t}\, \xi_{k,t}\right) \tag{10}$$

Here, n_a and P_a are action arbitration parameters, Ω_t represents the resource capacity, service radius and administrative rule constraint set at the current time, and $\mathrm{Proj}_{\Omega_t}(\cdot)$ represents the constraint projection operation. With this formulation, the model can prevent the single department policy from overoccupying the common resource and guarantee that the

joint action satisfies the actual execution boundary.

In order to improve the stability of deep reinforcement learning training, this paper introduces a shared value evaluation network to estimate the long-term benefits of joint actions. Set the shared value network as Q_ψ, the target network parameters as $\bar{\psi}$, the immediate collaborative reward of the ι training sample as v_ι, and the number of mini-batch samples as Mb, then the loss function of the value network is:

$$\mathcal{L}_c(\psi) = \frac{1}{M_b}\sum_{\iota=1}^{M_b}\left[\boldsymbol{v}_\iota + \gamma_c \mathcal{Q}_{\bar{\psi}}(\mathcal{X}_{\iota+1}, \boldsymbol{\zeta}_{\iota+1}) - \mathcal{Q}_\psi(\mathcal{X}_\iota, \boldsymbol{\zeta}_\iota)\right]^2 \tag{11}$$

where, γ_c is the long-term return discount factor and $\mathcal{L}_c(\psi)$ is used to constrain the estimation error of the shared value network. Through the shared value evaluation, the model can simultaneously consider the task completion rate, resource utilization rate, response time and department collaboration cost, so that the scheduling strategy can form a long-term optimization ability for continuous administrative service process. Overall, the model transforms multi-department resource scheduling into a trainable deep reinforcement learning problem, which provides a decision-making basis for subsequent fairness constraints and response cost control.

3.4 Deep Reinforcement Learning Optimization Algorithm for Fairness Constraint and Response Cost Control

Public resource scheduling should not only pursue the task completion rate, but also take into account regional fairness, administrative response time and execution cost.In order to more clearly show the execution process of deep reinforcement learning policy optimization under multi-objective reward constraints, this paper visualizes the process of state input, action generation, environment interaction, reward calculation and policy update, as shown in Figure 4.

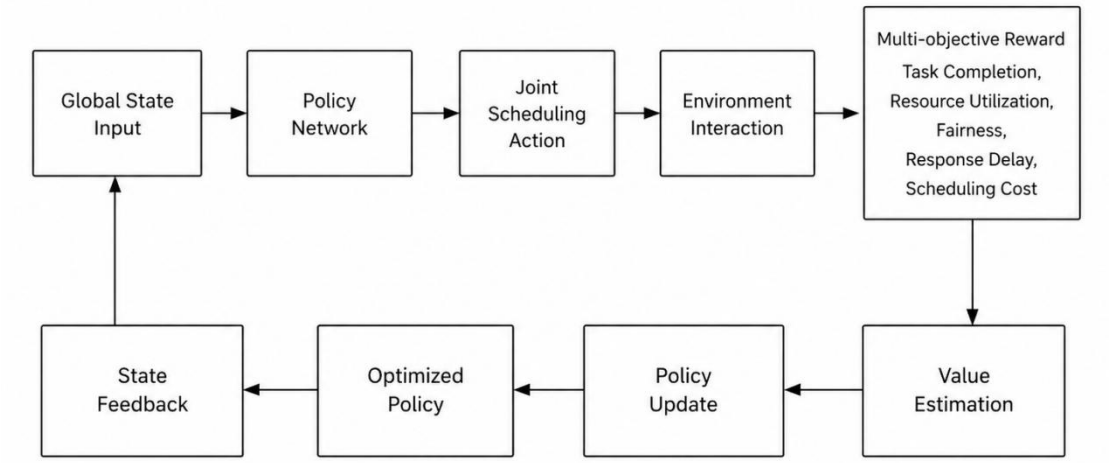

Figure 4: Flowchart of multi-objective reward-driven deep reinforcement learning policy optimization

Figure 4 shows that the constructed optimization process takes the global state vector as input, and generates actions such as resource allocation, cross-region scheduling and department collaboration through the policy network. On the basis of the scheduling environment feedback, the multi-objective reward signal is formed by integrating task completion benefit, resource utilization benefit, fairness benefit, response delay penalty and

scheduling cost penalty. Then, the value evaluation and policy update mechanism are combined to realize the iterative optimization of the scheduling policy, so that the model can better control the regional service differences and administrative operation costs while taking into account the processing efficiency.

In this paper, based on the joint scheduling action ζ_t in 3.3, the fairness constraint and response cost control mechanism are introduced to extend the optimization objective of deep reinforcement learning from a single efficiency gain to a multi-objective constrained gain. Let the integrated cooperative reward at time t be υt, which is calculated as follows:

$$v_t = \vartheta_1 \mathrm{Comp}_t + \vartheta_2 \mathrm{Util}_t + \vartheta_3 \mathrm{Fair}_t - \vartheta_4 \mathrm{Delay}_t - \vartheta_5 \mathrm{Cost}_t \tag{12}$$

Here, Comp_t represents the task completion benefit, Util_t represents the resource utilization benefit, Fair_t represents the fairness benefit, Delay_t represents the response delay penalty, Cost_t represents the scheduling cost penalty, and ϑ_1 to ϑ_5 are nonnegative weights and satisfy the normalization constraint. The reward structure enables the model to improve the administrative efficiency while restraining the excessive concentration of resources to a single region or a single department.

In order to explain the composition of the reward function and its corresponding optimization direction, this paper summarizes the contents of various rewards and punishments, as shown in Table 1.

Table 1: Deep reinforcement learning reward function composition and optimization objective description table

Reward Component	Calculation Meaning	Optimization Direction	Corresponding Optimization Objective	Administrative Service Function
Task Completion Reward	Proportion of completed public service tasks among all arriving tasks	Higher is better	Improve task processing capacity	Reduce backlogged matters and unresponded requests
Resource Utilization Reward	Effective matching degree between allocated resources and available resources	Higher is better	Improve public resource utilization efficiency	Avoid resource idleness and repeated dispatching
Fairness Reward	Balance degree of service satisfaction levels across different regions	Higher is better	Control regional service differences	Prevent resources from being concentrated in highly active regions for a long time
Response Delay Penalty	Degree to which actual response time exceeds the target service time	Lower is better	Reduce administrative response time	Improve the speed of public service access
Scheduling Cost Penalty	Execution consumption caused by cross-region transfer, manpower allocation, and departmental coordination	Lower is better	Reduce resource allocation cost	Improve the economic efficiency of administrative operation

It can be seen from Table 1 that the reward function unifis the efficiency objective, fairness objective and cost objective in the common resource scheduling into the same training framework, which can provide a more stable optimization direction for the deep reinforcement learning strategy.

To further quantify regional fairness, this paper constructs a fairness score based on the service satisfaction rate of different service areas:

$$\mathrm{Fair}_t = 1 - \frac{\sqrt{\frac{1}{J}\sum_{j=1}^{J}\left(v_{j,t} - \bar{v}_t\right)^2}}{\bar{v}_t + \varepsilon_f} \tag{13}$$

where J represents the number of service patches, $v_{j,t}$ represents the service satisfaction rate of patch j at time t, $\bar{v}_t$ represents the average service satisfaction rate of all patches, and ε_f is a stable term. The formula inhibits the scheduling bias by measuring the service difference of the area, so that the model can still maintain the basic service balance in the resource-constrained scenario.

The administrative response costs mainly come from service overtime, departmental collaborative waiting and cross-district resource allocation. In this paper, the part beyond the service time limit is used as the delay penalty:

$$\mathrm{Delay}_t = \frac{1}{J}\sum_{j=1}^{J}\frac{\max\left(0, \mathrm{RT}_{j,t} - \mathrm{SLA}_j\right)}{\mathrm{SLA}_j + \varepsilon_{rt}} \tag{14}$$

where $RT_{j,t}$ represents the actual response time of patch j, SLA_j represents the target service time limit set by the patch, and ε_{rt} is the time-normalized stable term. The penalty term can force the model to give priority to the requests with higher timeout risk, and reduce the waiting loss in the administrative service chain.

In the policy update stage, the joint action constraint is incorporated into the deep reinforcement learning optimization process to avoid the model output actions that exceed the resource capacity, service radius or administrative rule boundary. The KTH department policy network parameters are updated as follows.

$$\varphi_k^{new} = \varphi_k + \alpha_p \nabla_{\varphi_k}\left(\mathbb{E}_t[v_t] - \varsigma\Omega\mathbb{E}_t[\mathrm{Viol}(\zeta_t, \Omega_t)]\right) \tag{15}$$

Here, α_p denotes the policy update step and $\mathrm{Viol}(\zeta_t,\Omega_t)$ denotes the degree of violation of the constraint set Ωt by the joint scheduling action, $\varsigma\Omega$ denotes the constraint penalty strength. Through this update mechanism, the model can continuously improve the comprehensive reward while reducing the proportion of unexecutable actions, so that the scheduling results of deep reinforcement learning are more in line with the actual operating boundary of public administration scenarios.

3.5 A Deep Reinforcement Learning Policy Iterative Deployment Framework with Online Feedback Calibration

The common resource scheduling model is still affected by the change of demand structure, the disturbance of emergencies and the deviation of department execution after offline training. If a fixed strategy is used for a long time, the model is prone to degradation. To this end, this paper constructs a deep reinforcement learning policy iterative deployment framework based on online feedback calibration, and divides the model deployment process

into five parts: "execution monitoring, deviation identification, experience update, policy retraining, and security release". In order to further illustrate the closed-loop relationship between online feedback calibration, experience update and policy release, this paper presents a deep reinforcement learning policy iterative deployment framework, as shown in Figure 5.

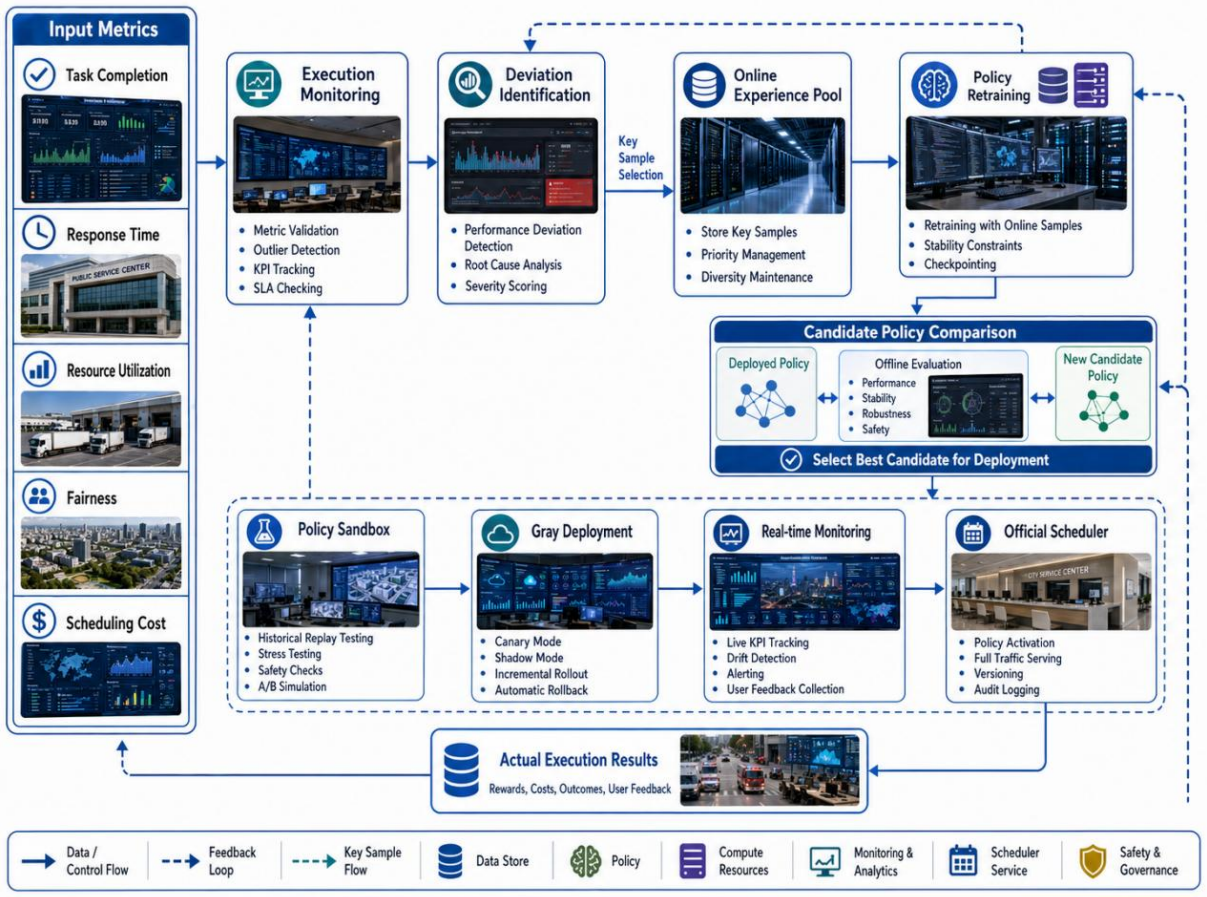

Figure 5: Framework diagram of iterative deployment of deep reinforcement learning policies based on online feedback calibration

It can be seen from Figure 5 that the system takes execution indicators such as task completion, response time, resource utilization, fairness and scheduling cost as input, and completes the screening of key samples through execution monitoring, deviation identification and online experience pool update, and forms a closed loop between policy retraining, candidate policy comparison, policy sandbox verification, grayscale deployment and official release. The framework not only ensures that the policy can be continuously optimized according to the environmental feedback, but also reduces the operational risk caused by direct online update through the phased verification and real-time monitoring mechanism.

After each round of scheduling, the system collects the execution feedback, including task completion, actual response time, resource occupancy rate, regional fairness and execution cost, and compares it with the expected results of the model to determine whether the current policy needs to be updated.

Let the execution feedback vector at time t be r_t^{fb}, the model prediction feedback vector be $\hat{r}_t^{fb}$, and the online calibration deviation score be defined as follows:

$$\mathcal{E}_t^{\mathrm{fb}} = \left\|\mathbf{r}_t^{\mathrm{fb}} - \hat{\mathbf{r}}_t^{\mathrm{fb}}\right\|_1 + \lambda_e\left|\widehat{\mathrm{Fair}}_t - \widehat{\mathrm{Fair}}_t\right| \tag{16}$$

Here, $\mathcal{E}_t^{\mathrm{fb}}$ represents the feedback bias degree, λ_e is the fairness bias correction coefficient, and $\widehat{\mathrm{Fair}}_t$ represents the fairness level predicted by the model. This formula can simultaneously identify the efficiency bias and fairness bias, and avoid the one-sided update of the model only according to the task completion rate.

When the feedback deviation exceeds the set threshold, the system writes the current state, joint action, comprehensive reward and next state into the online experience pool. Set the experience pool of the line as Bt, and the experience update method as follows:

$$\mathcal{B}_{t+1} = \mathrm{TopK}_{\kappa_b}\left(\mathcal{B}_t \cup \left\{\mathcal{X}_t, \zeta_t, v_t, \mathcal{X}_{t+1}, \mathcal{E}_t^{\mathrm{fb}}\right\}\right) \tag{17}$$

Among them, $\mathrm{TopK}_{\kappa b}(\cdot)$ indicates that the first κ_b experiences are filtered according to feedback bias, event importance and sample freshness to avoid the experience pool being occupied by a large number of low-value conventional samples. Through this mechanism, the model can preferentially learn key scenarios such as sudden demand, cross-regional resource shortage and departmental synergy imbalance.

In the policy release phase, in order to avoid the frequent fluctuation of model parameters affecting the stable operation of the administrative system, this paper adopts the smooth online update. Let the old policy parameters of department k be $\varphi_{\mathrm{k}}^{\mathrm{old}}$, the candidate parameters obtained by training based on the online experience pool be $\varphi_{\mathrm{k}}^{\mathrm{cand}}$, and the final deployment parameters be:

$$\varphi_{\mathrm{k}}^{\mathrm{dep}} = (1-\omega_{\mathrm{p}})\varphi_{\mathrm{k}}^{\mathrm{old}} + \omega_{\mathrm{p}}\varphi_{\mathrm{k}}^{\mathrm{cand}}, \quad 0 < \omega_{\mathrm{p}} < 1 \tag{18}$$

Here, ω_{p} is the deployment update coefficient and $\varphi_{\mathrm{k}}^{\mathrm{dep}}$ represents the policy parameters that are actually published to the scheduling system. This method can reduce the mutation risk while maintaining the continuous evolution of the strategy.

In the overall deployment process, the system does not directly replace the original administrative scheduling rules, but first completes small-scale simulation and verification through the policy sandbox, and then enters the gray deployment and real-time monitoring stage. When the resource utilization, task completion rate, response time and fairness reached the publishing threshold, the policy was pushed to the official scheduler. Thus, the deep reinforcement learning model is transformed from a one-time training model to a continuously calibrated administrative intelligent decision-making component, which can maintain strong adaptability when the supply and demand structure of public resources changes.

4 Results

4.1 Experimental environment Configuration and public resource scheduling dataset construction

In order to verify the effectiveness of the proposed deep reinforcement learning model in the dynamic allocation of public resources, this paper constructs a simulation experiment platform for public resource scheduling based on the digital twin environment. The experimental objects cover five types of public resources, such as community services,

medical response, traffic connection, emergency supplies and administrative Windows. Multi-scenario samples are generated according to regional population density, task arrival rate, resource supply capacity and emergency intensity. The data set contains 12 service areas, 48 resource nodes and 86 types of administrative service tasks. The task arrival process is simulated according to three types of patterns: peak on weekdays, fluctuation on weekends and sudden disturbance. For model training, 70% of the samples are used as the training set, 15% as the validation set, and 15% as the test set, and peak demand and node failure scenarios are added in the test phase to check the robustness of the scheduling strategy. The experimental comparison algorithms include Rule-based, GA, DQN, PPO and the deep reinforcement learning collaborative scheduling model proposed in this paper, and the evaluation indicators include resource utilization, task completion rate, average response time, fairness index and scheduling cost. The experimental environment and dataset configuration are shown in Table 2.

Table 2: Experimental environment configuration and common resource scheduling dataset description

Configuration Item	Specific Content	Value/ Description	Technical Function
Hardware Environment	Intel Xeon Silver 4214R, NVIDIA RTX 3090	12 cores and 24 threads, 24 GB GPU memory	Supports simulation computation and deep reinforcement learning training
Software Environment	Windows Server 2019, Python 3.10, PyTorch 2.1	64-bit system	Builds the policy network, value network, and simulation environment
Service Area	Urban public service areas	12 areas	Represents the spatial distribution of public resources
Resource Nodes	Medical, transportation, administrative service window, and emergency supply nodes	48 nodes	Forms a cross-departmental schedulable resource pool
Task Samples	Cleaned public service scheduling records	36,000 records	Serves as the basis for model training and testing
Task Types	Consultation handling, medical response, transportation connection, material allocation, etc.	86 task types	Covers multiple types of administrative service demands
Data Split	Training set, validation set, and test set	70%/15%/15%	Ensures independence between model training and performance evaluation
Experimental Scenarios	Regular demand, peak demand, emergency events, and cross-region task surge	4 scenarios	Tests model efficiency, robustness, and adaptability

It can be seen from Table 2 that the experimental configuration covers the computing environment, service area, resource node, task sample and disturbance scene, which can support the simulation training and performance verification of the dynamic allocation

process of public resources. The dataset contains both regular tasks and sudden disturbance tasks, which is helpful to test the scheduling stability of the proposed model under different public governance situations.

4.2 Comparative analysis of common resource utilization and task completion rate under different algorithms

In order to verify the advantages of the proposed model in the efficiency of common resource scheduling, Rule-based, GA, DQN, PPO and the method in this paper are selected for comparative experiments, and resource utilization and task completion rate are used as the core evaluation indicators. Among them, resource utilization is used to measure the effective invocation degree of various public resources, and task completion rate is used to reflect the system's ability to process service requests in a given period of time. The experimental results of different algorithms on two metrics are shown in Figure 6.

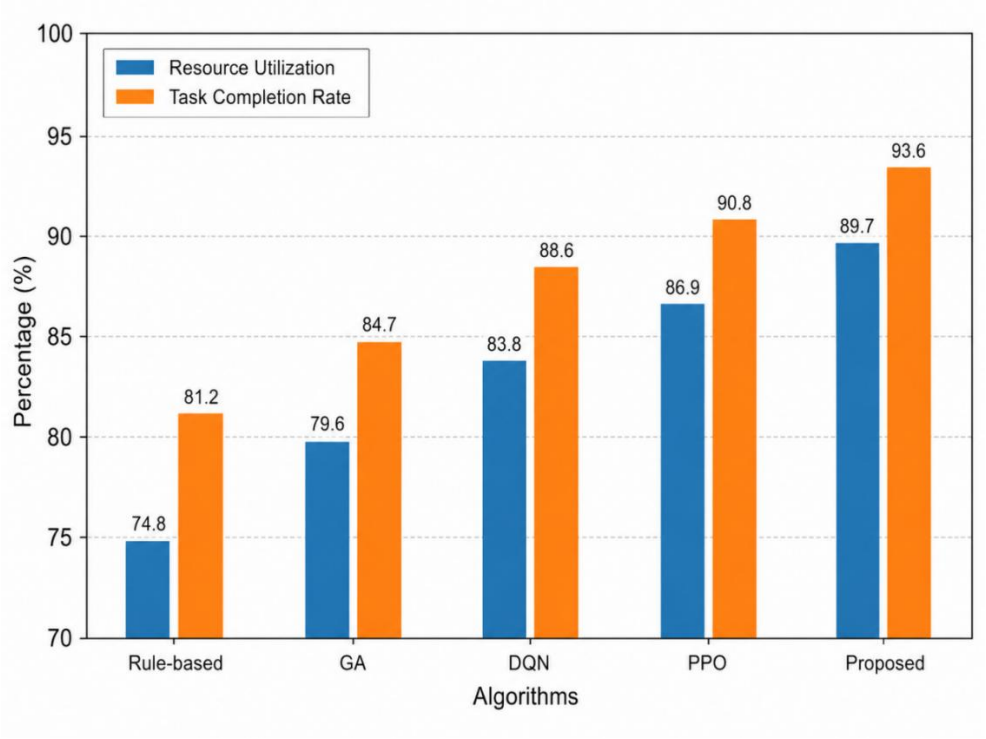

Figure 6: Bar charts comparing public resource utilization and task completion rate under different algorithms

Figure 6 shows that the traditional Rule-based method is difficult to adapt to demand fluctuations due to its dependence on fixed scheduling rules, and its resource utilization and task completion rate are only 74.8% and 81.2%, respectively. GA improves the partial scheduling quality by global search, and the two indexes are increased to 79.6% and 84.7% respectively. DQN and PPO can use environmental feedback to continuously optimize the strategy, and the performance is further improved. The resource utilization rate of PPO reaches 86.9%, and the task completion rate reaches 90.8%. In contrast, under the support of cross-department state fusion and collaborative scheduling mechanism, the method in this paper achieves the optimal results, the resource utilization rate reaches 89.7%, and the task completion rate reaches 93.6%, which are 14.9 and 12.4 percentage points higher than Rule-based, and 2.8 and 2.8 percentage points higher than PPO, respectively. The results show that the proposed deep reinforcement learning model can more fully tap the potential of

public resources, and maintain a higher service completion ability in a dynamic task environment.

4.3 Analysis on the optimization effect of deep reinforcement learning model on administrative response time

The administrative response time can directly reflect the improvement degree of the public resource scheduling policy to the actual service efficiency. In order to analyze why the deep reinforcement learning model can shorten the response time delay, this paper further counts the response strength of the policy network to key state features, including variables such as task arrival rate, resource inventory, department load, event risk, regional population density and cross-region scheduling distance. The response relationship of different key state features in the deep reinforcement learning strategy is shown in Figure 7.

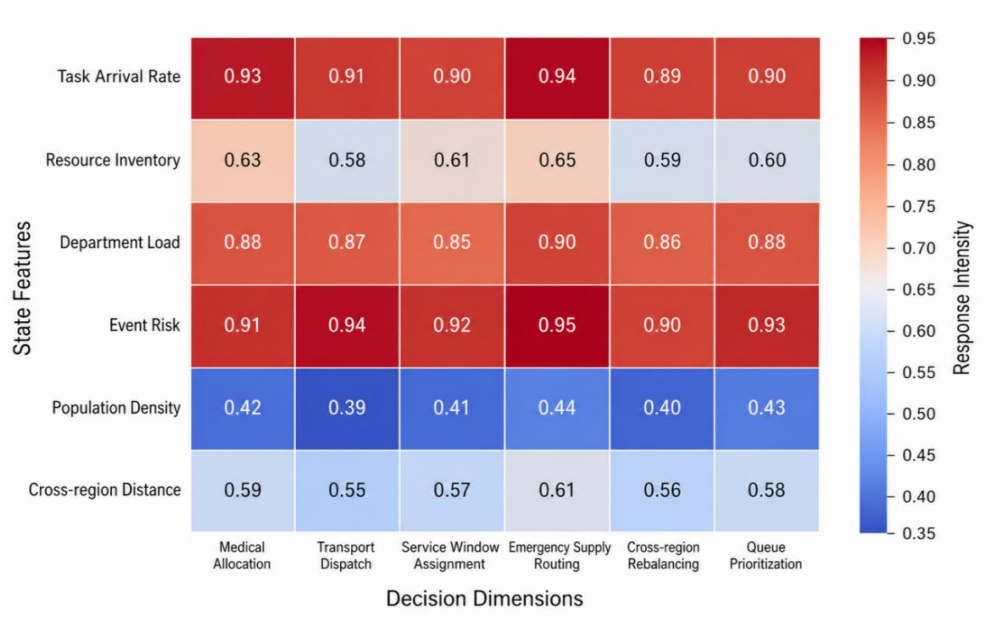

Figure 7: Heatmap of response of deep reinforcement learning policy to key state features

It can be seen from Figure 7 that the response intensity corresponding to task arrival rate, event risk and department load is the highest, indicating that the model pays more attention to high pressure and high timeliness scenarios when generating scheduling actions. Resource inventory and cross-region scheduling distance are the second, which indicates that the strategy takes into account resource consumption control while ensuring reachability. Population density mainly acts as an environmental correction. Affected by this, the proposed model can preferentially identify high-urgency areas and allocate available resources in advance, thereby shortening the administrative processing chain. Experimental results show that the average response time of the proposed method is reduced to 18.6 min, which is 15.6 min, 11.1 min, 6.2 min and 2.8 min shorter than Rule-based, GA, DQN and PPO, respectively. This shows that the deep reinforcement learning strategy is not simply to increase resource allocation, but to achieve more effective response time optimization through the differential perception of key state features.

4.4 Verification of model robustness under peak demand and emergency scenarios

In order to further test the stability of the proposed model in complex public governance environment, this paper sets up two types of disturbance scenarios, peak demand and emergency, and divides the scheduling process into regular operation stage, peak impact stage, emergency injection stage and recovery stage. In the experiment, the task arrival rate in the peak phase is 35% higher than that in the regular phase, and the additional resource node failure and cross-zone emergency task growth in the emergency phase are superimposed to investigate the scheduling retention ability of each algorithm under continuous disturbance. The performance variation of different algorithms in the perturbation scenario is shown in Figure 8.

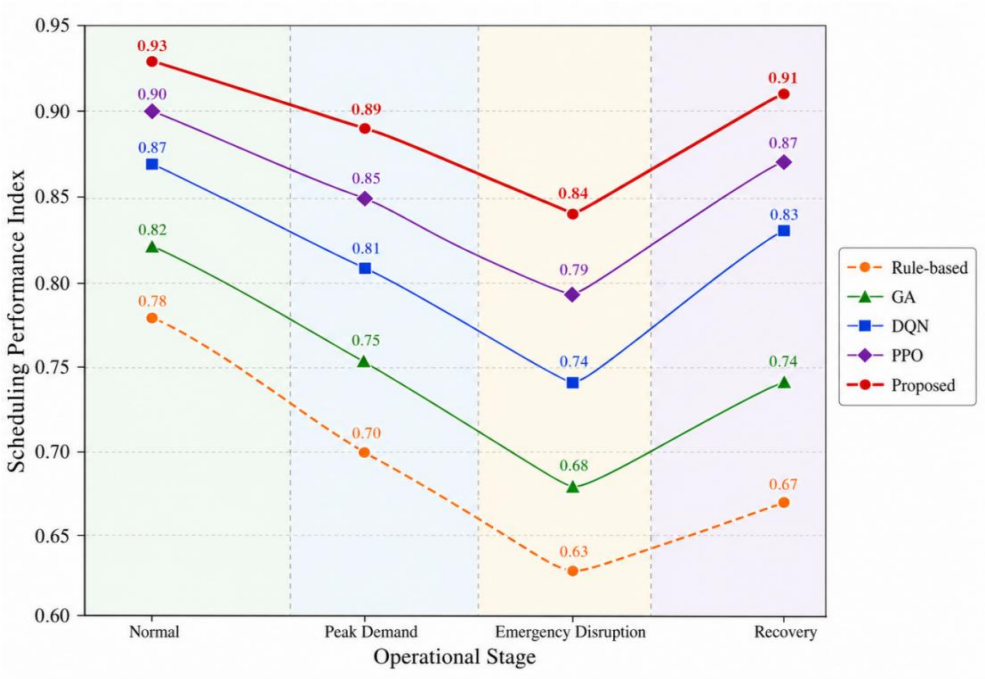

Figure 8: Graph of scheduling performance variation under peak demand and burst disturbance

It can be seen from Figure 8 that all algorithms show varying degrees of performance degradation after the arrival of peak demand and emergencies, but the proposed model has the smallest overall fluctuation and the fastest recovery speed. In the conventional phase, the comprehensive scheduling performance index of the proposed method is maintained at about 0.93. After entering the peak demand stage, it dropped to 0.89, a decline of only 4.3%. It was as low as 0.84 during the burst injection and then quickly recovered to 0.91 during the recovery phase. In contrast, PPO drops to 0.79 in the burst phase, DQN to 0.74, GA and Rule-based to 0.68 and 0.63, respectively. At the same time, the average recovery time of the proposed model is controlled within 3 scheduling cycles, which is 1 cycle shorter than PPO and 2 cycles shorter than DQN. This shows that the proposed deep reinforcement learning model can still maintain strong scheduling resilience and environmental adaptability under high-pressure and abnormal conditions.

4.5 Influence analysis of multi-objective reward mechanism on co-improvement of fairness and efficiency

In order to verify the effectiveness of the multi-objective reward function constructed in Section 3.4, this paper further analyzes the collaborative optimization effect of the model between fairness and administrative efficiency under different combinations of reward weights. In the experiment, the regional service fairness index was taken as the horizontal axis, the task completion rate was used to represent the administrative efficiency, and the scheduling cost was used as the auxiliary constraint quantity to obtain the Pareto front formed by different policy points. Figure 9 shows the co-optimization results of fairness and administrative efficiency under different reward weights.

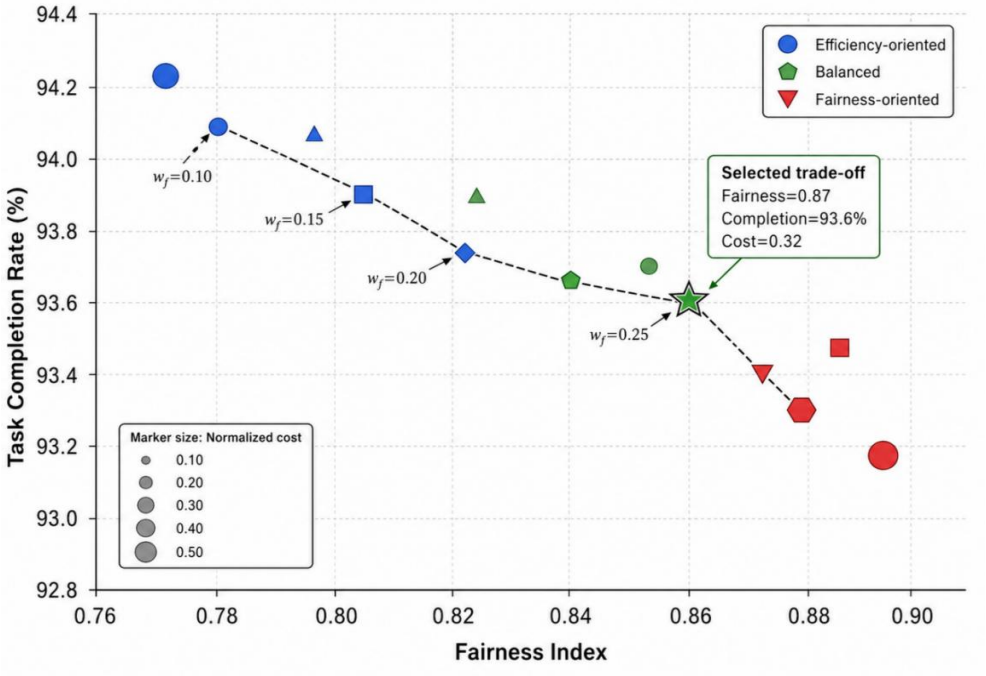

Figure 9: Diagram of the Pareto front for co-optimization of fairness and administrative efficiency

As can be seen from Figure 9, if the reward function only emphasizes efficiency gains, although the task completion rate can be improved to 94.1%, the fairness index is only 0.78, indicating that resources are easier to concentrate on areas with high demand and high activity. When the fairness weight gradually increases, the Pareto front as a whole moves to the upper right, indicating that the model can improve the regional service balance while maintaining a high task completion rate. When the fairness reward weight is increased from 0.10 to 0.25, the fairness index is increased from 0.78 to 0.89, and the task completion rate only slightly decreases from 94.1% to 93.3%, and the decrease is controlled within 0.8 percentage points. Further combined with the analysis of scheduling cost, the strategy point in the middle of the frontier performs best, and the compromise scheme selected in this paper corresponds to the fairness index of 0.87, the task completion rate of 93.6%, the average response time of 18.6 min, and the normalized scheduling cost of 0.32. This result shows that the multi-objective reward mechanism does not significantly weaken the administrative efficiency, but realizes the coordinated improvement of fairness, efficiency and cost control by inhibiting the excessive concentration of resources and ineffective cross-regional allocation.

4.6 Validation of the Generalization Ability of Deep Reinforcement Learning Models in Cross-region Transfer scenarios

In order to test the applicability of the proposed model in different public service areas, this paper sets up five transfer scenarios: original training area, adjacent area, long distance area, high density area and resource shortage area, and directly conducts policy transfer test without changing the core network structure. The distribution of model generalization performance under different cross-region transfer scenarios is shown in Figure 10.

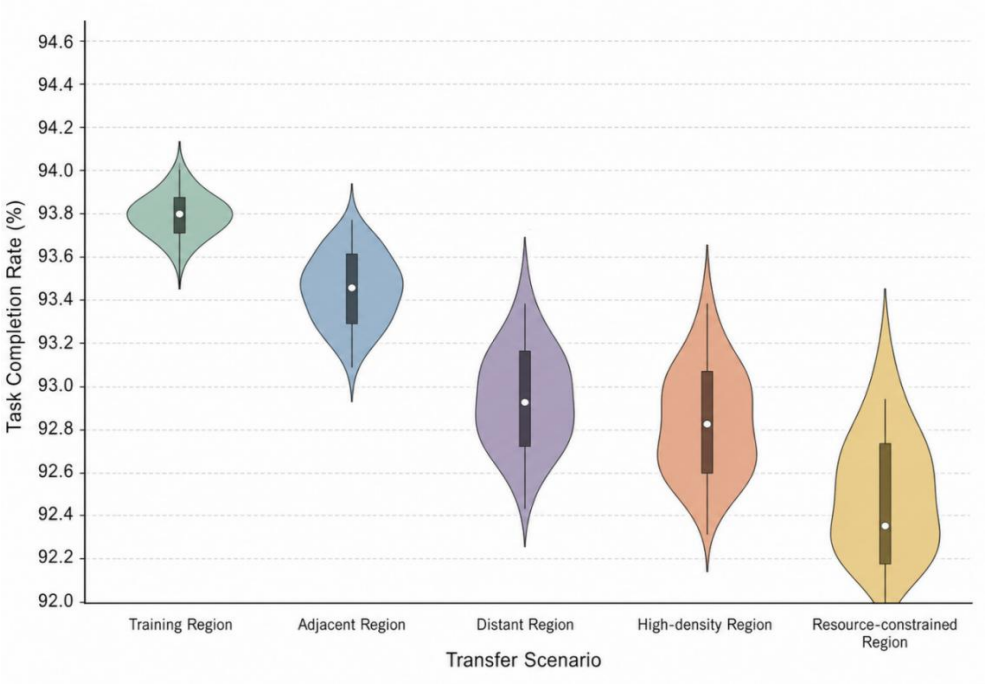

Figure 10: Violin plot of model generalization performance distribution in the cross-region transfer scenario

Figure 10 shows that the task completion rate of the proposed model in the original training area has the most concentrated distribution, with a median of 93.8%. After migrating to adjacent regions, the median is 93.4%, and the performance degradation is small. Due to the increasing difference of service structure in long-distance regions, the median decreased to 92.9%. High-density areas and resource-constrained areas remained at 92.7 percent and 92.3 percent, respectively. From the distribution width, the violin shape between the original training region and the neighboring regions is narrower, indicating that the policy output is more stable, while the distribution in the resource-constrained region is slightly wider, indicating that complex constraints will increase the scheduling fluctuation. Overall, the task completion rate of the proposed model in the five types of scenarios remains above 92.0%, and the maximum median reduction is only 1.5 percentage points, indicating that the proposed deep reinforcement learning method has good cross-regional transfer ability and deployment generalization potential.

5 Discussion

Focusing on the problem of dynamic allocation and administrative efficiency optimization of public resources, this paper constructs a collaborative scheduling framework combining

digital twin, spatio-temporal supply and demand graph and deep reinforcement learning. From the overall results, the proposed method can maintain good scheduling stability in complex public service environments. The key lies in transforming the running process of public resources from a static allocation problem to a continuous interaction problem, so that resource status, task requirements, department load and execution feedback can be dynamically updated in a unified framework. The digital twin environment provides a safe and controllable simulation space for model training, which can simulate multiple operating scenarios such as regular demand, peak pressure, node failure and emergencies, and reduce the risk caused by directly testing the strategy in the real administrative system. The spatio-temporal supply and demand graph further enhances the state representation ability, so that the model can identify the traffic distance, resource complementarity, service spillover and department collaboration between regions, and avoid one-sided scheduling judgment caused by only relying on a single point feature.

The cross-department co-scheduling model is an important support for the proposed method. Medical response, transportation connections, administrative windows, and emergency supplies do not operate independently of each other, and resource changes in one department may affect the service efficiency of other departments. In this paper, the department policy sub-network is used to generate local scheduling suggestions, and then the collaborative arbitration module completes the joint action output, so that the resource calls of different departments can be coordinated under uniform constraints. This mechanism not only retains the professionalism of department business, but also reduces the problems such as resource preemption, repeated distribution and response chain breakage. Compared with single department optimization, cross-department collaboration is more consistent with the actual operation logic of public resource scheduling, and can also improve the resource organization ability in complex scenarios.

Fairness constraint and response cost control make the model closer to the public administration scenario. Public resource scheduling should not only focus on processing speed, but also take into account regional balance, service accessibility and administrative operation costs. If the model pursues efficiency excessively, resources may be concentrated in areas with active demand for a long time, resulting in insufficient basic service protection in some areas. In this paper, fairness, delay and cost factors are added to the reward function, so that the policy training is constrained by both the efficiency objective and the public service equilibrium objective. The online feedback calibration mechanism further improves the continuous adaptation ability of the model. When the demand structure, resource status or execution effect change, the system can gradually modify the strategy through feedback deviation identification and experience pool update, so as to reduce the degradation risk of the model in long-term operation.

It should be noted that the method in this paper still has some limitations. The effectiveness of digital twin environment depends on data quality, entity mapping accuracy and rule modeling integrity. If the basic data is missing, lagging or inconsistent, the model training results may deviate from the real operating state. Cross-departmental collaboration also involves authority boundaries, process specifications, manual review and emergency disposal responsibility division. The scheduling suggestions generated by the algorithm need to be connected with the administrative system, and cannot be independent from the actual management process. Subsequent research can further combine privacy computing, interpretable reinforcement learning, and human-machine collaborative audit mechanism to improve the credibility, transparency, and deployability of the model in real public governance scenarios. In general, the proposed method provides a technical path with engineering potential for dynamic allocation of public resources.

6 Conclusion

Focusing on the problem of dynamic allocation and administrative efficiency optimization of public resources, this paper proposes a collaborative scheduling method that integrates digital twin, spatio-temporal supply and demand graph, deep reinforcement learning and online feedback calibration. In this method, scenes such as community services, medical response, traffic connection, emergency supplies and administrative windows are mapped to the digital twin environment, and regional demand, resource accessibility, department load and event risk are expressed through the spatio-temporal supply and demand graph, which provides structured state input for reinforcement learning strategy. The multi-objective reward function further integrates task completion rate, resource utilization, fairness, response delay and scheduling cost into the unified optimization, so that the model maintains service balance while improving efficiency. The experimental results show that on the data set of 12 service areas, 48 resource nodes and 86 types of administrative service tasks, the resource utilization rate of the proposed method reaches 89.7% and the task completion rate reaches 93.6%, which are 14.9 and 12.4 percentage points higher than those of the Rule-based method respectively. The average response time was reduced to 18.6 min, which was 2.8 min shorter than that of PPO. In the peak demand and emergency test, the minimum comprehensive performance of the model remained at 0.84, and rose to 0.91 in the recovery phase. Under the compromise strategy, the fairness index reaches 0.87, and the normalized scheduling cost is 0.32. In the cross-region transfer experiment, the task completion rate of five types of scenes remains above 92.0%, indicating that the model has good generalization ability. In general, the proposed method can improve the real-time performance, balance and environmental adaptability of public resource scheduling, and provide reference for intelligent decision-making deployment in government services, emergency response and urban public governance.

About the Author

Xinyan Li was born in Beijing, China in 1991. She obtained a master's degree from Beijing University of Chemical Technology. I am currently working as an administrative management teacher at Beijing Jingbei Vocational and Technical School. Her research interests include administrative management, educational management, and business management. E-mail lxy2010324@126.com

References

[1] Alwarafy A, Abdallah M, Ciftler B S, et al. The frontiers of deep reinforcement learning for resource management in future wireless HetNets: Techniques, challenges, and research directions[J]. IEEE Open Journal of the Communications Society, 2022, 3: 322-365. DOI: 10.1109/OJCOMS.2022.3153226.

[2] Aboueleneen N, Alwarafy A, Abdallah M. Deep reinforcement learning for Internet of Drones networks: Issues and research directions[J]. IEEE Open Journal of the Communications Society, 2023, 4: 671-683. DOI: 10.1109/OJCOMS.2023.3251855.

[3] Poltronieri F, Stefanelli C, Suri N, et al. Value is king: The MECForge deep reinforcement learning solution for resource management in 5G and beyond[J]. Journal

of Network and Systems Management, 2022, 30(4): 63. DOI: 10.1007/s10922-022-09672-6.

[4] Sellami B, Hakiri A, Ben Yahia S, et al. Energy-aware task scheduling and offloading using deep reinforcement learning in SDN-enabled IoT network[J]. Computer Networks, 2022, 210: 108957. DOI: 10.1016/j.comnet.2022.108957.

[5] Sellami B, Hakiri A, Ben Yahia S. Deep reinforcement learning for energy-aware task offloading in join SDN-Blockchain 5G massive IoT edge network[J]. Future Generation Computer Systems, 2022, 137: 363-379. DOI: 10.1016/j.future.2022.07.024.

[6] Danino T, Ben-Shimol Y, Greenberg S. Container allocation in cloud environment using multi-agent deep reinforcement learning[J]. Electronics, 2023, 12(12): 2614. DOI: 10.3390/electronics12122614.

[7] Kougioumtzidis G, Poulkov V K, Lazaridis P, et al. Deep reinforcement learning-based resource allocation for QoE enhancement in wireless VR communications[J]. IEEE Access, 2025, 13: 25045-25058. DOI: 10.1109/ACCESS.2025.3538546.

[8] Murarețu I, Vultureanu-Albiși A, Ilie S, et al. ABMS-driven reinforcement learning for dynamic resource allocation in mass casualty incidents[J]. Future Internet, 2025, 17(11): 502. DOI: 10.3390/fi17110502.

[9] Haydari A, Yılmaz Y. Deep reinforcement learning for intelligent transportation systems: A survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(1): 11-32. DOI: 10.1109/TITS.2020.3008612.

[10] Noaeen M, Naik A, Goodman L, et al. Reinforcement learning in urban network traffic signal control: A systematic literature review[J]. Expert Systems with Applications, 2022, 199: 116830. DOI: 10.1016/j.eswa.2022.116830.

[11] Waqar N, Hassan S A, Pervaiz H, et al. Deep multi-agent reinforcement learning for resource allocation in NOMA-enabled MEC[J]. Computer Communications, 2022, 196: 1-8. DOI: 10.1016/j.comcom.2022.09.018.

[12] Park H, Lim Y. Deep reinforcement learning based resource allocation with radio remote head grouping and vehicle clustering in 5G vehicular networks[J]. Electronics, 2021, 10(23): 3015. DOI: 10.3390/electronics10233015.

[13] Mohebifard R, Hajbabaie A. Deep reinforcement learning technique for traffic metering in connected urban street networks[J]. Transportation Research Record: Journal of the Transportation Research Board, 2024, 2678: 1872-1888. DOI: 10.1177/03611981241253581.

[14] Saadi A, Abghour N, Chiba Z, et al. A survey of reinforcement and deep reinforcement learning for coordination in intelligent traffic light control[J]. Journal of Big Data, 2025, 12: 84. DOI: 10.1186/s40537-025-01104-x.

[15] Ali E M, Abawajy J, Lemma F, et al. Analysis of deep reinforcement learning algorithms for task offloading and resource allocation in fog computing environments[J]. Sensors, 2025, 25(17): 5286. DOI: 10.3390/s25175286.

[16] Ergün S. Resource allocation optimization for effective vehicle network communications using multi-agent deep reinforcement learning[J]. Journal of Dynamics and Games, 2025, 12(2): 134-156. DOI: 10.3934/jdg.2024017.

[17] Hurtado Sánchez J A, Casilimas K, Caicedo Rendon O M. Deep reinforcement learning for resource management on network slicing: A survey[J]. Sensors, 2022, 22(8): 3031. DOI: 10.3390/s22083031.

[18] Aghapour Z, Sharifian S, Taheri H. Task offloading and resource allocation algorithm based on deep reinforcement learning for distributed AI execution tasks in IoT edge computing environments[J]. Computer Networks, 2023, 223: 109577. DOI: 10.1016/j.comnet.2023.109577.

[19] Lim D, Lee W, Kim W T, et al. DRL-OS: A deep reinforcement learning-based offloading scheduler in mobile edge computing[J]. Sensors, 2022, 22(23): 9212. DOI: 10.3390/s22239212.

[20] Hoang L T, Nguyen C T, Pham A T. Deep reinforcement learning-based online resource management for UAV-assisted edge computing with dual connectivity[J]. IEEE/ACM Transactions on Networking, 2023, 31(6): 2761-2776. DOI: 10.1109/TNET.2023.3263538.

[21] Naderializadeh N, Sydir J J, Simsek M, et al. Resource management in wireless networks via multi-agent deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2021, 20(6): 3507-3523. DOI: 10.1109/TWC.2021.3051163.

[22] Kasi M K, Abu Ghazalah S, Akram R N, et al. Secure mobile edge server placement using multi-agent reinforcement learning[J]. Electronics, 2021, 10(17): 2098. DOI: 10.3390/electronics10172098.

[23] Bansal M, Chana I, Clarke S. UrbanEnQoSPlace: A deep reinforcement learning model for service placement of real-time smart city IoT applications[J]. IEEE Transactions on Services Computing, 2023, 16(4): 3043-3060. DOI: 10.1109/TSC.2022.3218044.

[24] Sami H, Mourad A, Otrok H, et al. Demand-driven deep reinforcement learning for scalable fog and service placement[J]. IEEE Transactions on Services Computing, 2022, 15(5): 2671-2684. DOI: 10.1109/TSC.2021.3075988.