2026 Volume 4, Issue 4
Educational Innovation Research
ISSN: 3029-1852 (Online)
ISSN: 3029-1844 (Print)

Research on an Automated Intraday Liquidity Scheduling Strategy for Finance Companies Based on Deep Reinforcement Learning

Bin Ge*
Zhejiang Communications Investment Group Finance Co., Ltd., Hangzhou 310001, Zhejiang, China

***Author to whom correspondence should be addressed.**

Abstract: This study rigorously formulates the complex fund-scheduling problem as a Markov decision process (MDP). It constructs a state space that integrates real-time and forecast information, an atomic action space that conforms to business logic, and a reward function that balances long-term returns against immediate risk. To address the curse of dimensionality and the credit-assignment problem in coordinated scheduling among multiple fund units, a multi-agent deep deterministic policy gradient (MADDPG) algorithm is adopted. Under a centralized-training and decentralized-execution framework, the algorithm reconciles global optimization with decentralized decision-making. In addition, a difference-reward mechanism and Kalman filtering are used to accurately measure each agent's individual contribution and reduce the impact of environmental noise on reward signals. The results show that, compared with a static rule engine and a conventional linear programming method, the proposed deep reinforcement learning strategy reduces average daily funding costs by 50.4%, lowers the payment failure rate to 0.002%, and maintains a high liquidity buffer adequacy ratio. The strategy also demonstrates clear advantages in decision timeliness, collaborative handling of complex instructions, and self-adaptation potential, thereby providing an innovative pathway for finance-company fund scheduling to progress from intelligentization to automation.

Keywords: fund scheduling; Markov decision process; reward function; Kalman filtering; finance company

Online publication: April 26, 2026

1. Introduction

As the "internal banks" of corporate groups, finance companies perform the core functions of centrally allocating funds among member entities, ensuring payment security, and improving the overall efficiency of group funds. Intraday liquidity management is the "lifeline" of a finance company's fund operations: amid rapidly changing cash inflows and outflows, cross-account positions must be scheduled accurately and in real time to minimize liquidity reserve costs and avoid payment risk [1–2]. Traditional scheduling models rely heavily on managers' experiential judgment and semi-automated systems based on fixed rules. When confronted with complex, dynamic, and highly uncertain market conditions and internal transaction networks, these models often exhibit delayed responses, suboptimal decisions, and difficulty balancing multiple objectives such as cost, risk, and efficiency.

In recent years, artificial intelligence technologies, particularly deep reinforcement learning (DRL), have achieved major advances in autonomous driving, robot control, quantitative trading, and other fields because of their strong environmental perception and autonomous learning and optimization capabilities in complex sequential decision-making problems [3–4]. This study comprehensively investigates an automated DRL-based strategy for intraday liquidity scheduling in finance companies. The scheduling problem is rigorously formulated as a Markov decision process; the state space, action space, and reward function are designed according to business rules; and a DRL agent is constructed and trained. Simulation experiments are then used to examine the feasibility of the proposed strategy in reducing funding costs, controlling payment risk, and improving scheduling efficiency. The study is intended to provide a novel solution and practical reference for the intelligent and automated transformation of fund management in finance companies.

2. Construction of a DRL-Based Intraday Liquidity Scheduling Model

2.1. Scheduling Problem Formulation

2.1.1. State Space

The funding environment faced by finance companies is highly uncertain. For example, payment instructions arrive randomly and market interest rates can change instantaneously, causing state changes to exhibit pronounced stochastic characteristics. To address this issue, the state space constructed in this study includes not only currently available and accurate account data but also key parameters that reflect the potential distribution of future states. Specifically, the state vector integrates real-time account balances, cumulative intraday cash flows, expected short-term fund movements, market interest rates, and workday categories. This design provides the DRL agent with sufficient information to assess the current situation and anticipate the consequences of its actions, enabling it to learn reliable and efficient scheduling policies under uncertainty [5].

Fund scheduling in a finance company takes place in a typical stochastic environment, in which key factors such as the arrival of payment instructions from member entities and real-time changes in market financing channels are uncertain. Accordingly, system state transitions follow a probability distribution rather than a deterministic functional transformation. This dynamic characteristic can be formalized as: $T: S \times A \rightarrow S$; that is, in state s, after scheduling action a is executed, the system enters the next state s according to the conditional probability distribution $T\left(s' \middle| s, a\right)$.

2.1.2. Action Space

Within the Markov decision process framework, an agent's actions directly affect the funding state, and its decisions are based on a predefined action set. Once a scheduling action is implemented, the system transitions to the next state with a certain probability, and the specific transition depends on the action taken. At each decision point, the system is in a particular state s_t, and the agent selects an action from the action set A_{s_t}, namely a. The system state then changes according to the probability distribution $\Pr\left(s_{t+1} \middle| s_t = s, a_t = a\right)$.

Within this framework, all actions can, in principle, be executed in any state; however, their effect on the current state may be neutral or even adverse, and this must be represented by the state-transition matrix. This differs from conventional artificial-intelligence planning, in which feasible actions are selected using preconditions. By substantially reducing the range of actions that must be considered in each state, preconditions lower computational complexity. When constructing the Markov decision process for liquidity scheduling, this study adopts the same principle and imposes appropriate action constraints for different funding conditions. These constraints reduce the action-search space at each decision point and improve learning efficiency.

The scheduling process considered in this study is assumed to be a time-homogeneous Markov decision process; that is, all actions have the same execution duration, and state transitions occur at fixed intervals. Each action can be regarded as the most basic operating unit of the system, completing a fundamental scheduling task within the shortest time unit. In intraday liquidity management, for example, time can be divided into fixed 15-minute intervals. During each interval, the agent can perform elementary operations such as “transferring funds to a specified account, ” “making a short-term investment, ” or “maintaining the current cash reserve.” Together, these operations constitute the basic steps of fund-management scheduling.

2.1.3. Reward Function

Unlike models that optimize only the expected total return over a finite horizon $\max E\left[\sum_{t=0}^{k-1} R_t\right]$, fund scheduling is a continuous process and is therefore better suited to an infinite-horizon optimization framework. This study introduces a discount factor γ satisfying $0<\gamma<1$, and defines the agent’s objective as maximizing the expected discounted long-run cumulative return, namely $\max E\left[\sum_{t=0}^{\infty} \gamma^t R_t\right]$. Mathematically, the discount factor γ ensures convergence of the infinite series; in business terms, it reflects the present value of future returns, consistent with the time value of money in financial management.

The solution to a Markov decision process is a policy π, which is a mapping from the state space S to the action space A, expressed as $T: S->A$. When making a scheduling decision, the agent observes the current funding state s, selects and executes an action according to the policy $\pi(s)$, after which the environment transitions to a new state s_{t+1} and provides an immediate reward R_t. This process repeats continuously. If the policy is stationary—that is, if the selected action depends only on the current state and not on the time point t—the consistency of the scheduling rule can be ensured.

2.2. DRL Agent

2.2.1. Deep Deterministic Policy Gradient Algorithm

When multiple fund units or accounts within a finance company must be coordinated, a single-agent model may be unable to optimize overall liquidity effectively. This study therefore adopts a multi-agent deep deterministic policy gradient method to address the “curse of dimensionality” arising in multi-agent collaborative decision-making.

This approach is consistent with the practical requirements of liquidity management. Suppose there are N scheduling agents, whose policy-parameter set is denoted by $\theta=\{\theta_1,\cdots\theta_n\}$. During training, a centralized critic network receives the observations of all agents $o=\{o_1,\cdots o_n\}$ and their joint action $a=\{a_1,\cdots a_n\}$, and thereby estimates a more accurate joint action-value function $Q_i^\pi(o,a_1,\cdots a_n)$ for evaluating the long-term benefit of the overall scheduling operation. This enables each agent to learn indirectly how to cooperate with the other agents during training, thereby optimizing the company’s overall liquidity. During execution, each agent independently makes scheduling decisions using only local information, such as account balances and asset maturity dates, which ensures decision timeliness and robustness. Under this architecture, the reinforcement-learning agents can improve global coordination while remaining suitable for decentralized operating conditions.

2.2.2. Target Policy

In the MADDPG multi-agent system, the company’s overall liquidity indicator—that is, the global reward—is determined by the actions of the individual fund-unit agents and by unobservable market factors. To identify each agent’s specific contribution to the decision outcome and address the credit-assignment problem in multi-agent cooperation, this study adopts a difference-reward method. For agent i, its reward $R^i\left(a^i \middle| s,a^{-i}\right)$ is evaluated as the difference in global reward between the actual action a^i it takes and a predefined baseline action c^i, as shown below:

$$R^i\left(a^i \middle| s,a^{-i}\right)=R(s,a)-R\left(S,\left\langle a^{-i},c^i\right\rangle\right) \tag{1}$$

Let a^{-i} denote the action set of the other agents, and let $R(s,a)$ denote the overall return expected when agent i adopts the baseline policy, such as maintaining its current position. This encourages each agent to learn actions that make a positive marginal contribution to overall liquidity.

The effects of unobservable factors—such as market volatility and sudden payment requests—on the global reward signal are modeled as environmental noise. Accordingly, for agent i at time t and state j, the observed global reward g_t can be decomposed into two additive components, b_t^i and b_t:

$$g_t = r_t^i\left(j\right) + b_t^i \tag{2}$$

The noise term b_t follows a Gaussian distribution. To dynamically and more accurately estimate an agent's actual local reward from the noisy global reward, this study applies a Kalman filter to the reward signal. The system model can be expressed in linear form as follows:

$$\begin{cases} x_t = Ax_{t-1} + Bu + \varepsilon_t, \cdots, \varepsilon_t - N\left(0, \Sigma_1\right) \\ g_t = Cx_t + v_t, \cdots, v_t - N\left(0, \Sigma_2\right) \end{cases} \tag{3}$$

In the equation, the state vector x_t contains estimates of local rewards in the various states together with noise components. In the state-transition term Ax_{t-1}, the matrix A is taken as the identity matrix, with the updated state represented by x_t. In the observation term Cx_t, the entry corresponding to agent i in state j is 1. The observation-error covariance $\Sigma\delta = 0$ is set to zero, allowing the filter to be updated in real time during training. Finally, the filtered "clean" local reward $r_t^i\left(j\right)$, rather than the original global reward g_t, is used to update the agent's action-value function, thereby improving the stability and accuracy of policy learning.

3. Experimental Results and Analysis

To verify the effectiveness of the proposed automated intraday liquidity scheduling strategy based on a multi-agent deep deterministic policy gradient algorithm and Kalman-filter-based reward correction, this study constructs a highly realistic simulated fund-scheduling environment for a finance company and designs a series of comparative experiments. The proposed strategy and the baseline methods are comprehensively evaluated and analyzed along three core dimensions: funding cost, payment-risk control, and scheduling efficiency.

3.1. Funding Cost

The experiment simulates the daily operation of a financial institution with five principal fund pools (accounts) over one trading day, from 9:00 to 17:00. The decision interval is 15 minutes, yielding 32 decision points. Fund inflows—including deposits from member entities and proceeds from maturing investments—and outflows—payment instructions issued by member entities—are generated as stochastic processes based on historical data, with large-value payment instructions exhibiting sudden-arrival characteristics. The market overnight financing rate fluctuates randomly around a benchmark value. The following methods are compared:

(1) Static rule engine: Following common industry practice, fixed minimum and maximum liquidity-reserve limits are set for each account. If an account balance falls below the lower limit, funds are transferred from a surplus account or external financing is initiated. If the balance exceeds the upper limit, the excess funds are used for short-term investment or debt repayment.

(2) Linear programming model: This model minimizes the total daily cost, including opportunity cost and financing cost. At each decision point, a linear programming model is formulated using a deterministic forecast of fund movements over the subsequent period, such as the next hour, to identify the optimal scheduling plan. This model represents a conventional optimization approach based on deterministic forecasts.

(3) Proposed DRL strategy: The MADDPG framework described in Section 1.2 is used to train the multi-agent system. Each fund pool is controlled by an independent agent, and Kalman-filter-corrected difference rewards are used during training.

Table 1 presents the average daily funding costs of the methods after 100 simulated trading days. The proposed DRL strategy demonstrates excellent control of funding costs, with an average daily total cost of only RMB 58,000—50.4% lower than that of the static rule engine and 46.8% lower than that of the linear programming model. This result directly demonstrates the considerable potential of the DRL strategy for forward-looking optimal scheduling in dynamic and uncertain environments. The strategy performs well in reducing both the opportunity cost of liquidity reserves and the cost of emergency financing. The opportunity cost falls substantially to RMB 43,000, indicating that the agents learn to maintain excess cash at a relatively low level while ensuring payment security, thereby improving fund utilization. Emergency financing cost decreases markedly to RMB 15,000, indicating that the agents make more accurate forecasts and scheduling decisions regarding inter-account fund movements and effectively avoid costly external financing caused by localized liquidity shortages.

Table 1. Comparison of Average Daily Funding Costs across Scheduling Strategies

Method	Average daily opportunity cost of liquidity reserves (RMB 10,000)	Average daily emergency financing cost (RMB 10,000)	Average daily total cost (RMB 10,000)
Static rule engine	8.5	3.2	11.7
Linear programming model	6.1	4.8	10.9
Proposed DRL strategy	4.3	1.5	5.8

3.2. Payment-Risk Control

Table 2 compares the payment-risk control performance of the scheduling strategies. The key to controlling payment risk is preventing payment failures—that is, cases in which payment instructions from member entities cannot be executed promptly because of insufficient account balances. Managing this risk requires an adequate liquidity buffer throughout the scheduling process. The proposed DRL strategy achieves the lowest payment failure rate (0.002%) and the highest liquidity buffer adequacy ratio (95%), outperforming the comparison methods overall. These results indicate that, after the reward function assigns a severe negative penalty to payment failures, the agents learn not only to avoid failures but also to maintain an adequately sized and appropriately distributed liquidity buffer across accounts in a forward-looking manner, thereby guarding against future uncertainty.

Table 2. Comparison of Payment-Risk Control across Scheduling Strategies

Method	Payment failure rate	Liquidity buffer adequacy ratio	Meets risk-tolerance requirement
Static rule engine	0.01%	92%	Yes
Linear programming model	0.02%	78%	No
Proposed DRL strategy	0.00%	95%	Yes

3.3. Scheduling Efficiency

Scheduling efficiency concerns decision speed, adaptability, and the ability to process complex instructions, all of which determine the system’s practical usability. **Table 3** summarizes the qualitative comparison and key quantitative results for the methods in the efficiency dimension. During execution, the DRL strategy requires only a forward pass through a neural network, with an average decision time of approximately 15 ms—far shorter than the 15-minute decision interval and fully satisfying real-time requirements. Although training is time-consuming and performed offline, once deployed, the trained model makes decisions far more efficiently than the linear programming model, which must solve an optimization problem online and whose computation time increases sharply as the numbers of accounts and constraints grow.

Table 3. Comparison of Efficiency Characteristics across Scheduling Strategies

Method	Average time per decision	Complex-instruction handling	Adaptability	Overall assessment
Static rule engine	< 1 ms	Weak	None	Extremely fast, but its logic is simple and rigid; it cannot handle complex scheduling or adapt to change.
Linear programming model	120–500 ms	Strong (in theory)	Weak	The optimization problem must be resolved at every decision point; computation is relatively slow and increases sharply with scale, and the model cannot adapt autonomously.
Proposed DRL strategy	~15 ms	Strong	Strong	Fast decision-making, with the ability to handle high-dimensional, complex decisions and to learn and adapt online.

4. Conclusion

This study evaluates the proposed automated intraday liquidity scheduling strategy for finance companies, based on multi-agent deep reinforcement learning, from the three perspectives of funding cost, risk control, and scheduling efficiency, using both quantitative and qualitative analyses. The experimental results show that the strategy reduces average daily funding costs by 50.4%, lowers the payment failure rate to 0.002%, and maintains a liquidity buffer adequacy ratio of 95%, thereby achieving an effective balance between cost and risk. The strategy also demonstrates clear advantages in real-time decision-making, collaborative handling of complex scheduling instructions, and future self-learning and adaptation. These findings provide strong evidence that deep reinforcement learning—particularly multi-agent coordination and reward-signal processing techniques—is both feasible and advantageous for solving the complex stochastic decision problem of intraday liquidity scheduling in finance companies, and they offer practical support for the intelligent transformation of fund management.

Disclosure statement

The author declares no conflict of interest.

References

[1] Wang YZ, Song X B, Sun Y, et al., 2025, A Survey of Capital Efficiency and Financial Risk among Chinese Listed Companies: 2024. Accounting Research, 2025(12): 180-188.

[2] Jiang AL, 2025, Practice of Building a World-Class Financial Management System in Company A, a Financial Leasing Company. Finance & Accounting, 2025(15): 27-29.

[3] Zhao M, Xie L, Lin WJ, et al., 2024, A Deep Reinforcement Learning Portfolio Model Based on a Dynamic Selection Predictor. Computer Science, 51(4): 344-352.

[4] Long J, Xie L, Xu HJ, 2024, An Ensemble Deep Reinforcement Learning Portfolio Model. Journal of Computer Applications, 44(1): 300-310.

[5] Bu Z, Zhang SF, Li XY, et al., 2023, Adaptive Stock Index Forecasting Based on Deep Reinforcement Learning. Journal of Management Sciences in China, 26(4): 148-174.