International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-12

https://doi.org/10.70917/ijcisim-2026-0139
Article

Optimization Study of Intelligent Robot Production Line Scheduling Problem Based on Markov Decision Process

Shuai Yang *

Department of Intelligent Manufacturing, Jiangsu Vocational College of Electronics and Information, Huai'an, Jiangsu, 223003, China; hcityangshuai@sina.com

Abstract: With the rapid development of science and technology, industrial technology has continued to advance and improve. Addressing the issue that traditional Markov prediction methods often yield results that deviate from reality, this paper proposes an optimization solution method based on reinforcement learning. It employs a deep deterministic action strategy to compute an infinite-dimensional semi-Markov queuing model (where the queuing system control time is positive infinity), thereby achieving the objective of model optimization. Taking the parts production task O = {2, 12, 19} as an example, the effectiveness of the proposed method is validated by verifying the mold scheduling process solved using the reinforcement learning algorithm. Research shows that the average time spent on each stage of part production using the pre-optimization scheduling strategy was 27.90167, while the average time after optimization was 12.6533. Therefore, applying the reinforcement learning-based optimization method significantly improves the operational efficiency of intelligent robot production line scheduling and ensures reliability.

Keywords: Markov; reinforcement learning; production line scheduling; intelligent robot

1. Introduction

With the continuous advancement of science and technology, human productivity has been steadily improving, and robotics technology has seen exponential growth. Intelligent robotics technology, in particular, has brought about a revolution in industrial production [1-2]. Currently, intelligent robotics technology is widely applied in production planning and scheduling optimization, not only alleviating the workload of human workers on production lines, enhancing work efficiency and production quality, but also reducing corporate costs [3-4].

In today's highly competitive manufacturing environment, production scheduling, as a core component of production operations, plays a crucial role in influencing a company's production efficiency, cost control, and customer satisfaction [5-6]. Traditional production scheduling methods often rely on experience and simple rules, making it difficult to adapt to complex and dynamic production environments [7]. Intelligent robots, however, offer new opportunities and tools for optimizing production scheduling [8]. Intelligent robots can process large amounts of data and extract valuable information and patterns from it, particularly through the optimization of intelligent robot production line scheduling using Markov decision process algorithms [9-11]. The Markov decision process is a key concept in reinforcement learning, serving as a mathematical model to describe decision-making processes in uncertain environments. It has been widely applied in fields such as resource allocation, production scheduling, financial investment, and robot control [12-15]. In the optimization of intelligent robot production line scheduling, this algorithm enables the production line to transition from one state to another by making different decisions in different states, and evaluates the value of each state transition based on a reward function to identify an optimal strategy that maximizes long-term rewards in an uncertain environment, thereby optimizing the scheduling problem [16-19].

Literature [20] discusses the application of industrial robots in intelligent manufacturing, taking

intelligent industrial robots and PLCs as research objects, and proposes a joint control scheme for PLC controllers and industrial robot controllers, providing a reference for the subsequent construction of similar production line joint control systems. Literature [21] highlights the differences between intelligent robots and conventional robots, their applications, and classifications, discusses key technologies and typical application examples in the field of intelligent robots, and finally explores the future trends of intelligent robots. Literature [22] examines the application of artificial intelligence technology in manufacturing, discusses the changes brought about by the rapid development of core technologies in the new era, and explores the current state of smart manufacturing from the perspectives of application technology and industry. It also proposes suggestions for the application of artificial intelligence in smart manufacturing. Literature [23] introduces the application of intelligent robots in industrial production, which not only reduces production costs but also improves product quality. It discusses the key areas, benefits, and challenges of applying intelligent robots in Industry 4.0, laying the foundation for guiding the future development of intelligent robots in the Industry 4.0 era. The above studies primarily discuss the application of intelligent robots in the industrial field, which effectively promotes the intelligentization of industrial production and contributes to improving production efficiency and product quality.

Reference [24] emphasizes the importance of production optimization for enterprise development, especially production planning and scheduling, which are of great significance for improving the market competitiveness of enterprises. Reference [25], based on the high-order Markov decision model, explores its application in the optimal scheduling of raw materials in the production workshop, emphasizes the validity of the model, and there are certain schemes that can make the coordination degree of the optimal scheduling of raw materials, and the influence degree of the comprehensive qualification rate is at a relatively high level. Reference [26] introduces the importance of production scheduling (PS) and simulation optimization (SO) methods, and discusses existing research on the application of SO to PS, providing real-time and efficient SO-based decision support tools for PS modules in modern manufacturing. Literature [27] reviews the development of production scheduling optimization based on scheduling objective optimization, scheduling method selection, and the construction of scheduling management control systems. The study indicates that efficiency and energy consumption are important factors in scheduling objectives, and the establishment of an intelligent production scheduling management control system is of significant importance for achieving the intelligent transformation of production. Literature [28] combines genetic algorithms with particle swarm algorithms, applying this hybrid algorithm to production scheduling models. It emphasizes the efficiency and flexibility of hybrid algorithms in production scheduling, providing valuable references for optimizing similar algorithms in production scheduling. The above studies affirm the important role of production scheduling in manufacturing, highlighting its optimization as a key factor in corporate development, and introduce the application and impact of optimization algorithms such as Markov decision processes and genetic algorithms in production scheduling.

This paper addresses the challenge of predicting production line scheduling issues in complex intelligent manufacturing environments, where factors such as changes in product market opportunities, varying equipment update cycles, and differences in enterprise scale make such predictions difficult. The paper proposes an optimization solution based on reinforcement learning. In a semi-Markov model observing the number of customers and the number of open service counters in a queuing system, the paper utilizes a deep deterministic policy gradient (DDPG) to derive corresponding action strategies based on the environmental state of the queuing system, thereby achieving intelligent robot production line scheduling. Taking the parts production task O = {2, 12, 19} as an example, the paper analyzes the current state of production line scheduling in an intelligent manufacturing context to validate the effectiveness of the proposed method.

2. Intelligent Robot Production Line Scheduling Optimization Model

2.1. Markov Decision Process

User-oriented time performance standards are one of the primary evaluation criteria for production scheduling in networked manufacturing models. These standards have evolved alongside continuous updates and advances in software and hardware technology. The concept and application of time performance scheduling standards have surpassed the capabilities of traditional system scheduling and expanded into the realm of unlimited network applications, such as database systems, human-machine interactive interfaces, mathematical models, distributed systems, network management structures, technology, and artificial intelligence.

2.1.1. Production Problem Description Based on Markov Decision Processes

The Markov decision process is a forecasting method that uses state transition probabilities to study the likelihood of an event occurring during a forecast period. In a networked manufacturing environment, various unpredictable factors such as changes in product market opportunities, different equipment upgrade cycles, and predictions of enterprise development scale often lead to analysis and forecasting results that deviate from reality when using the Markov forecasting method to analyze specific issues, as the assumptions are often too simplistic.

Manufacturing enterprises operating under a networked manufacturing model are event-driven, so various fluctuations may occur during actual production processes. This necessitates the extension of the Markov decision process. Based on the Markov decision process, the following assumptions are proposed for production scheduling in a networked manufacturing environment:

(1) Product-related factors, such as changes in market demand, the emergence of new design schemes, and product maturity.

(2) Environmental factors and other unknown factors leading to product discontinuation, early production, etc.

Assume that the arrival of new workpieces at the start of production follows a Poisson distribution, where λ is the arrival rate. Then, the probability that x new workpieces start production in a unit of time is:

$$P(x)=e^{-\lambda}\lambda^{x}/x! \tag{1}$$

The expected value of new workpieces arriving per unit time, i.e., the arithmetic mean, is:

$$E(x)=\sum_{x=0}^{\infty}xP(x)=\lambda \tag{2}$$

That is, the average number of new workpieces arriving per unit time that begin production is equal to the arrival rate.

That is, the average number of new workpieces that begin production per unit time is equal to the production rate. Replacing the unit time with any time t, we obtain the probability of producing x new workpieces within a known time t:

$$P(x(t))=e^{\lambda t}(\lambda t)^{x}/x! \tag{3}$$

The probability that no new workpieces will be produced within time t is:

$$P(0)=e^{-\lambda t} \tag{4}$$

The probability that at least one new workpiece will be produced within time t is:

$$P(x(t)>0)=1-e^{-\lambda t} \tag{5}$$

The probability of new workpieces arriving in production in the next cycle is independent of the probability of old workpieces being produced. Thus, the Markov decision process characteristic can simplify the scheduling model for workpieces and production, whose density function is:

$$P(x(t))=\lambda e^{-\lambda t} \tag{6}$$

The expected value of t is equal to:

$$E(t)\int_0^{\infty}t\lambda e^{-\lambda t}dt=-te^{-\lambda t}\Big|_0^{\infty}+\int_0^{\infty}e^{-\lambda t}=1/\lambda \tag{7}$$

The average time interval between consecutive new workpiece production orders is $1/\lambda$. Similarly, the average production time of the production unit is $1/\mu$. It can be seen that the production system can only operate normally when $\lambda<\mu$; otherwise, the tasks waiting for production will wait indefinitely [29].

Assume that S_i is a state of the system, indicating that there are $i-1$ new workpieces in the queue waiting to be produced and 1 old workpiece being produced in the production unit. From the probability density, the probability of 1 new workpiece arriving to be produced within dt time is:

$$P(dt_1)=\lambda e^{-\lambda dt}dt=\lambda dt+O(dt^2) \tag{8}$$

Similarly, the probability of producing one new workpiece within dt hours is:

$$P(dt_2) = \mu dt + O(dt^2) \tag{9}$$

The probability that no new workpieces arrive and no old workpieces are completed within dt hours is:

$$P(dt_3) = 1 - P(dt_1) - P(dt_2) = 1 - (\lambda + \mu)dt - O(dt^2) \quad (i \geq 0) \tag{10}$$

Ignore the quadratic term.

Assume that the probability of the production system being in state S_i at time $t + dt$ is $P_i(t + dt)$. Then, from the state transition diagram, we have:

$$(\lambda + \mu)dtP_i(t) = \lambda P_{i-1}(t) + \mu dtP_{i+1}(t) \tag{11}$$

Let $\lambda / \mu = \rho$, then we obtain: $P_i(t) = \rho^i P_0(t)$. The production system must be in a certain state at any given moment, therefore:

$$\sum_{i=0}^{\infty} P_i = 1 \Leftrightarrow \sum_{i=0}^{\infty} \rho^i P_0(t) = 1 \tag{12}$$

When $\rho < 1$, we have:

$$\sum_{i=0}^{\infty} \rho^i = 1/(1-\rho) \Leftrightarrow P_0(t) = 1 - \rho \tag{13}$$

That is, under stable production system conditions, the probability of no new workpieces is:

$$1 - \rho = (\mu - \lambda) / \mu \tag{14}$$

The probability of a new workpiece is λ / μ. The arithmetic mean of the production of new workpieces in the production system is:

$$n = E(i) = \sum_{i=0}^{\infty} iP_i(t) = \sum_{i=0}^{\infty} (1-\rho)^i \rho^i = \rho(1-\rho) \tag{15}$$

Let $\lambda / \mu = \rho$. When $\rho < 1$, we have:

$$\sum_{i=0}^{\infty} \rho^i = 1/(1-\rho) \Leftrightarrow P_0(t) = 1 - \rho \tag{16}$$

That is, under stable conditions in the production system, the probability of no new workpieces is $1 - \rho = (\mu - \lambda) / \mu$, and the probability of new workpieces is λ / μ.

2.1.2. Response Time Calculation for Scheduling Rules

A single-machine scheduling system with Markovian properties, where the response time R is the time from the start of scheduling a new workpiece to its completion and departure, is defined as $n = \lambda R$, where n is the number of new workpieces in the system and λ is the arrival probability of new workpieces. The response times calculated based on the Markov decision process for the three scheduling rules are as follows.

(1) FIFO Priority Scheduling Rule

In a networked manufacturing environment, production scheduling rules are based on arrival times and random factors. The first-in, first-out (FIFO) rule is more suitable from the perspective of actual production applications. Therefore, in the predictive analysis model, it is assumed that the waiting time for new workpieces to be produced is independent of the current state of the production system and the start time of production for old workpieces.

In a networked manufacturing environment, the FIFO production scheduling rule assumes that the probability of a production unit scheduling new workpiece production follows a Poisson distribution, with the production probability of new workpieces being μ and the arrival rate of new workpieces being λ. Let $\lambda / \mu = \rho$, then the average response time of the system under the FIFO mode is:

$$R = n / \lambda = 1 / (\mu(1 - \rho)) \tag{17}$$

In FIFO mode:

$$R_{FIFO} = 1/(\mu - \lambda) \tag{18}$$

(2) Priority Scheduling Rules

The Priority Scheduling Method (SRR) is based on the fundamental principle of rule-based scheduling. When making scheduling decisions, the system determines the priority weight of workpieces to be processed by comprehensively considering the different scheduling rules applied to various devices and workpieces within the system, based on the ultimate objective. It then prioritizes processing the workpiece with the highest priority level first. This scheduling method adopts a relatively balanced scheduling strategy, operating by assigning priority levels to new products according to various rules and determining the scheduling queue based on the level of priority.

Assume that the probability of a new workpiece arriving with priority is λ , and the probability of arriving to determine the scheduling queue order is $\lambda^{'}$. Let t_1 and t_2 be the times at which two new workpieces arrive to determine the scheduling queue order, then we have:

$$1/\lambda = t_2 - t_1 \tag{19}$$

Assuming that t_1 and t_2 are the scheduling queue sorting times determined for two new workpieces, respectively, then:

$$1/\lambda^{'} = t_2^{'} - t_1^{'} \tag{20}$$

Given the priority level linear parameter $r = 1 - b/a$, where a and b are the linear increase coefficients for new workpieces in determining priority levels and scheduling sequences. Given that $(t_2 - t_1)/(t_2^{'} - t_1^{'}) = r$, i.e., $\lambda^{'}/\lambda = r \Leftrightarrow \lambda^{'} = r\lambda$. Given that $r < 1$, the new workpiece is delayed in the process waiting area at a rate of priority level r before being added to the scheduling queue. Response time of the priority scheduling method:

$$R_{sm}(k) = R_d + R_{sLK} \tag{21}$$

Where R_d is the average waiting time of the workpiece in the waiting area, that is:

$$R_d = r \times R_s - R_s \tag{22}$$

Therefore, its response time is:

$$R_{srr}(k) = 1/(\mu - \lambda) - (1 - kq\mu)/(\mu - \lambda) \tag{23}$$

(3) Slack Scheduling Rules

The slack time (Slack) of a workpiece at the current time is defined as: the delivery date of the workpiece minus the current time, minus the time required for the remaining processes of the workpiece. The minimum slack time priority rule gives priority scheduling to workpieces with the smallest slack time. The minimum slack time scheduling (SLK) method results in a higher workpiece arrival rate than the FIFO method.

Assuming the time slice is q , the average start time is $1/\mu$, and each workpiece requires an average of k time slices, then $1/\mu = k \times q$. If a workpiece requires k time slices for processing, it will wait for k time slices in the processing area. Assume that the short job has $1/\mu_1 = k_1 \times q$ time slices, the long job has $1/\mu_2 = k_2 \times q$ time slices, the arrival probability of new workpieces is $\lambda = \lambda_1 + \lambda_2$, and the processing time is:

$$1/\mu = (\lambda_1/\mu)k_1 \times q + (\lambda_2/\mu)k_2 \times q \Leftrightarrow \lambda/\mu = \rho = \rho_1 + \rho_2 \tag{24}$$

The response time of a workpiece waiting for a time slice of k in the workpiece processing area is:

$$R_{SLK}(k) = \rho/(\lambda(1-\rho)) = k \times q/(1-\rho) \tag{25}$$

2.2. Optimization Solution Methods Based on Reinforcement Learning

Reinforcement learning algorithms are a type of machine learning algorithm that can solve large-scale or incomplete information Markov decision processes and other sequential decision problems. In this section, we will use reinforcement learning methods to optimize the semi-Markov queueing model established in this chapter. Based on this semi-Markov queueing model, we use the Deep Deterministic Policy Gradient (DDPG) to optimize and solve the model. DDPG can handle both continuous-action models and discrete-action models. Below, we will introduce the specific optimization process of the DDPG method.

I_v represents the environmental state observed by the agent at time v. In the semi-Markov model established in this paper, the specific environmental state is $I = \{(i,t) \mid i \geq 0, 0 \leq t \leq C\}$, i.e., the number of customers in the queuing system and the number of open service counters observed. a_i is the action selected by the agent at time v. After the action is executed by the environment, the environment state of the queuing system changes from I_v to I_{v+1}, and the number of customers and the number of open service counters in the system may change. The decision for the specific DDPG algorithm action a_v is a deterministic decision. Adding noise can increase the diversity of samples during the initial sampling phase [30].

For the semi-Markov queueing model, the corresponding action strategy can be searched for, and the number of service counters to be opened is determined based on the number of customers in the system. The system state for the next iteration is then provided. The specific implementation process of the DDPG algorithm is shown in Figure 1. Using the DDPG algorithm, calculations can be performed on an infinite-dimensional semi-Markov queueing model (where the queueing system control time is positive infinity).

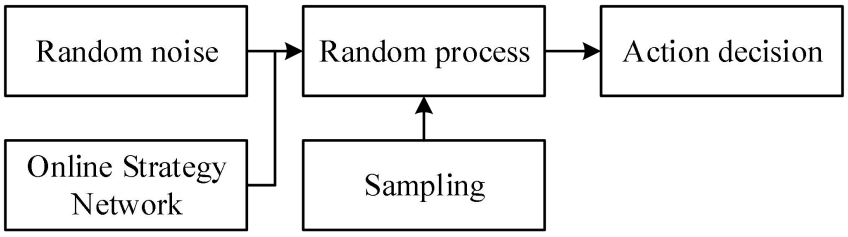


Figure 1. Action Decision-making Process of the DDPG algorithm.

In the figure, $d(I_v)$ is the deterministic policy. When random noise is introduced, the execution policy α_v is obtained, where N is the number of random samples and r_v is the single-step reward value. Next, we use the DDPG algorithm for the semi-Markov queueing model to optimize the problems that exist in the actual production process of steel enterprises.

2.3. Construction of a Scheduling Optimization Model

In industrial manufacturing processes, workpieces are distributed according to Poisson distribution on intelligent robot production lines. If each cycle is one period, all movements of intelligent robots are at the cycle point. A model can be constructed as follows:

$$X = \{X(n), \phi, D, P^v, f^v\} \tag{26}$$

In this context, X denotes the production line model; $X(n)$ denotes the state of the production line at a given moment; ϕ denotes the state space; D denotes the action set; P^v denotes the transition probability matrix; f^v denotes the performance matrix; and v denotes the stable strategy.

At the n th decision point, the state of the production line is:

$$X(n) = \{X_p(n), X_R(n)\} \tag{27}$$

Among these, $X_p(n)$ denotes the distribution of workpieces at the n th decision point; $X_R(n)$ denotes the availability of the intelligent robot gripper at the n th decision point. Joint encoding of

$X_p(n)$ and $X_R(n)$ is performed, and the scheduling strategy is incorporated into it. The production line's motion state is represented by $v[x(n)]$. When $v[x(n)]=2$, the robot is performing placement tasks; when $v[x(n)]=1$, the robot is performing picking operations; when $v[x(n)]=0$, the robot is not performing any actions. Thus, the intelligent robot's action set is defined as: $D=\{0,1,2\}$. When the robot is in a certain action state, the next action state is influenced by the transition probability matrix. Therefore, the transition probability matrix must be solved first before the scheduling model can be solved:

$$P^v = P_{x(n)x(n+1)}\{v[x(n)]\} \tag{28}$$

Among these, $P_x(n)x(n+1)\{v[x(n)]\}$ represents the action of the robot in state $x(n)$, and $v[x(n)]$ is the probability of transitioning to $x(n+1)$. In the production line, workpieces are primarily transported under the influence of a Poisson flow. Therefore, the current action and workpiece state determine the state of the next workpiece. When the robot performs various operations, it often incurs corresponding rewards and costs. Incorporating these into the scheduling model can enhance its accuracy. After the calculations are completed, the scheduling model can be used to implement optimized scheduling solutions.

3. Numerical Experiments and Case Studies

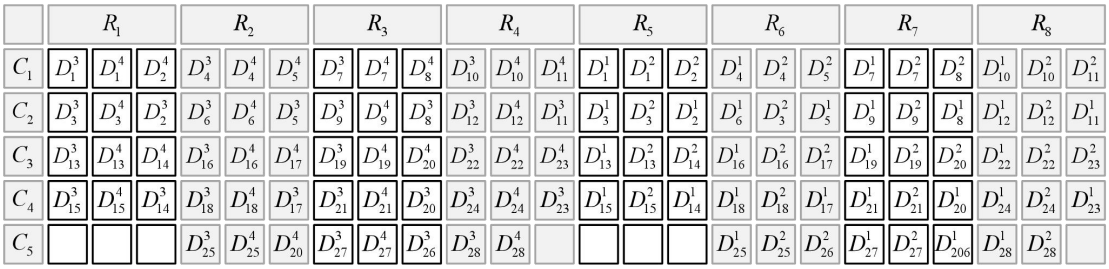
3.1. Numerical Experimental Analysis

Based on actual production experience, the loading and unloading times for vehicles are shown in Table 1.

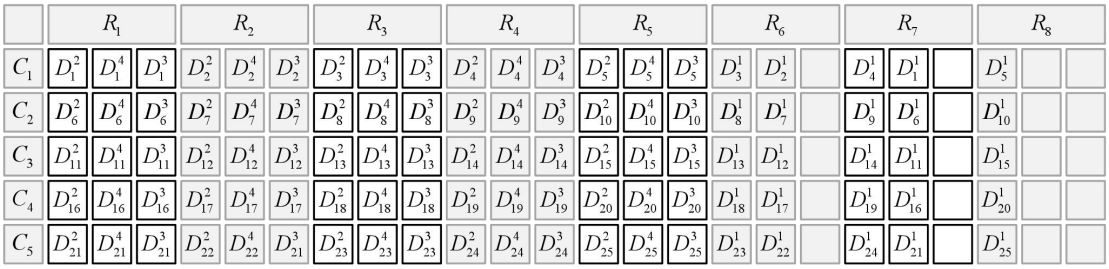
Table 1. Loading and unloading duration.

Loading layer	Uninstall layer	Loading time/min	Unloading length/min	Total time/min
Third layer	Third layer	9.02	9.16	18.17
Third layer	The first or the second	8.54	4.15	12.63
The first or the second	The first or the second	3.78	3.78	7.63
The first or the second	Production line	3.78	3.78	7.63

An example of a four-station mold storage system is shown in Figure 2.



(a) Group and store instances



(b) Aggregate and store instances

Figure 2. The example of the four station mould is stored.

Taking the parts production task O={2,12,19} as an example, the mold scheduling process solved by the reinforcement learning algorithm is shown in Table 2. This example demonstrates the effectiveness of the reinforcement learning algorithm in solving the problem of vehicle coordination scheduling.

Table 2. station mould gathers the scheduling instance.

Parts	Driving	Molds	Starting point	End point	Start time/min	End time/min
2	Front	D_2^3	R2C1L3	Number 3	0	12.59
	After	D_2^1	R6C1L2	Number 1	0	12.26
	Front	D_2^4	R2C1L2	Number 4	12.56	20.14
	After	D_{12}^3	R2C1L1	Number 2	16.32	23.78
12	Front	D_2^2	R2C3L3	Number 3	23.84	36.28
	After	D_{12}^1	R6C3L2	Number 1	23.84	36.36
	Front	D_{12}^4	R2C3L2	Number 4	36.27	43.89
	After	D_{12}^2	R2C3L1	Number 2	40.12	48.28
-2	Front	D_2^4	Number 4	R2C1L2	47.56	58.89
	After	D_2^2	Number 2	R2C1L1	47.50	55.39
	Front	D_2^3	Number 3	R2C1L3	58.86	71.28
	After	D_2^1	Number 1	R6C1L2	58.86	66.38

19	Front	D_{16}^{1}	R7C4L2	R8C4L2	66.28	74.59
	After	D_{19}^{3}	R4C4L3	Number 3	71.34	84.96
	Front	D_{19}^{1}	R7C4L1	Number 1	73.89	84.56
	After	D_{19}^{4}	R4C4L2	Number 4	83.87	92.28
	Front	D_{19}^{2}	Number 4	Number 2	87.56	98.76
-12	After	D_{12}^{4}	Number 2	R2C3L3	95.51	107.39
	Front	D_{12}^{2}	Number 3	R2C3L1	95.51	103.56
	After	D_{12}^{3}	Number 1	R2C3L3	107.36	120.89
	Front	D_{12}^{1}	Number 4	R6C3L2	107.36	114.89
-19	After	D_{19}^{4}	Number 2	R4C4L2	119.26	134.28
	Front	D_{19}^{2}	Number 3	R4C4L1	119.13	128.26
	After	D_{19}^{3}	Number 1	R4C4L3	132.25	143.26
	Front	D_{19}^{1}	Number 3	R7C4L1	132.25	138.56

To further validate the superiority of reinforcement learning algorithms, this paper compares reinforcement learning algorithms with rule-based heuristic optimization algorithms. Considering the heuristic idea that having two vehicles operate simultaneously can effectively utilize time, the following rules can be used for optimization: minimize vehicle position conflicts. Specifically, the leading vehicle should prioritize selecting molds or buffer positions located at the front of the mold library, while the trailing vehicle should make the opposite selection. Input data for stamping production tasks of different scales are shown in Table 3.

Table 3. Different scale test case input.

Serial number	Scale	Production task
1	2	3,7
2	4	2,5,9,13
3	6	1,10,15,23,25
4	8	4,6,11,14,18,20,23,25
5	10	2,5,9,13,14,17,19,21,23,25

After repeating the solution 100 times, the reinforcement learning algorithm solution results are shown in Table 4.

Table 4. Intensive learning.

Serial number	Min/min	Max/min	Avg/min	SD/min	Time/s
1	85.23	97.52	86.35	2.42	1.14

2	186.27	218.00	202.78	6.13	1.49
3	234.78	302.56	28996	6.09	1.67
4	394.56	422.07	411.06	4.45	1.85
5	482.57	522.89	500.27	8.41	2.09

The results of the rule-based heuristic algorithm are shown in Table 5.

Table 5. Based on the rule optimization algorithm to solve the result.

Serial number	Min/min	Max/min	Avg/min	SD
1	85.23	85.23	85.23	0.12
2	194.56	235.78	208.56	10.48
3	284.26	336.38	304.08	11.30
4	422.56	472.63	445.32	12.14
5	500.27	612.14	556.39	20.98

Among them, Min and Max represent the shortest and longest scheduling times, respectively. Avg represents the average scheduling time. SD represents the standard deviation. Time represents the average solution time. The comparison results show that the reinforcement learning algorithm proposed in this paper can obtain a shorter scheduling time under the same input conditions. The rule-based optimization algorithm only makes relatively optimal choices at each step, but the reinforcement learning algorithm can accumulate previous scheduling experience and comprehensively consider the overall situation.

3.2. Case Study Analysis

Optimizing the scheduling of intelligent robot production lines for industrial manufacturing. To validate the optimization results, the optimized scheduling strategy was applied to an intelligent robot production line at a certain industrial manufacturing plant, and the effects before and after optimization were compared. During the research process, parameters were set based on an actual robotic automation production line, with the principle of balancing the production line's arrival rate and the average production rate of the robots to avoid excessive workpiece arrival rates causing significant blockages and losses, or excessive robot movement speeds leading to overheating and wear. The parameters for the intelligent robotic production line are shown in Table 6.

Table 6. Smart machine life production line parameter setting table.

Project	Parameter
Work grab time/s	1
Workpiece time/s	2
Workpiece size/cm	3.6
Transmission speed/(ms-1)	0.05
Robot running speed/(ms-1)	1.4

After completing the preparatory work, both production lines (the pre-optimization production line and the post-optimization production line) began operating simultaneously. Following the completion of the operation, the time spent on each stage of the workpiece production process was recorded, with the results presented in Table 7. Analysis of the data recorded in the table shows that the time spent on each stage of workpiece production using the pre-optimization scheduling strategy was significantly longer than that spent using the post-optimization scheduling strategy. This indicates that applying the post-optimization scheduling strategy can significantly improve the production efficiency of intelligent robots and the overall efficiency of the production line, thereby bringing higher economic benefits to industrial manufacturing plants.

Table 7. Scheduling and optimization of the process of production.

Link classification	Time consuming before optimization	Optimized after time
Wrkpiece	26.35	11.04
Workpiece handling and marking	12.48	6.39
Workpiece to specify position (position)	38.79	18.67
Placement of workpiece	5.32	2.15
Workpiece to specify position (position)	55.78	24.38
Grab the workpiece to the processing center	28.69	13.29
Mean	27.90167	12.6533

4. Conclusion

This paper employs reinforcement learning optimization methods to optimize the scheduling problem of intelligent robot production lines based on traditional Markov decision processes. The parts production tasks O = {2, 12, 19} are selected as the research example, and the mold scheduling process is solved using reinforcement learning algorithms. By comparing the time consumption of each production stage for each workpiece between the two production lines (the pre-optimization production line and the post-optimization production line), the study shows that the time consumption of each production stage for workpieces under the pre-optimization scheduling strategy is significantly higher than that under the post-optimization scheduling strategy. This demonstrates that the application of the proposed method achieves dynamic optimization of production tasks and efficient integration of resources, with the optimization method significantly enhancing production efficiency and flexibility.

Funding

This research was supported by the 6th Jiangsu Province '333 Talents' 2022 Training Support Funding Plan (Project Number: s.u Caihang, [2022]122).

References

1. Cheng, H., Jia, R., Li, D., & Li, H. (2019). The rise of robots in China. Journal of Economic Perspectives, 33(2), 71-88.
2. Wang, T. M., Tao, Y., & Liu, H. (2018). Current researches and future development trend of intelligent robot: A review. International Journal of Automation and Computing, 15(5), 525-546.
3. Kim, J. H., Yang, W., Jo, J., Sincak, P., & Myung, H. (2015). Robot intelligence technology and applications 3. In Springer International Publishing.
4. Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2023). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. Artificial intelligence and international HRM, 172-201.
5. Fuchigami, H. Y., & Rangel, S. (2018). A survey of case studies in production scheduling: Analysis and perspectives. Journal of Computational Science, 25, 425-436.
6. Harjunkoski, I., Maravelias, C. T., Bongers, P., Castro, P. M., Engell, S., Grossmann, I. E., ... & Wassick, J. (2014). Scope for industrial applications of production scheduling models and solution methods. Computers & Chemical Engineering, 62, 161-193.
7. Jiang, Z., Yuan, S., Ma, J., & Wang, Q. (2022). The evolution of production scheduling from Industry 3.0 through Industry 4.0. International Journal of Production Research, 60(11), 3534-3554.
8. Gao, K., Huang, Y., Sadollah, A., & Wang, L. (2020). A review of energy-efficient scheduling in intelligent production systems. Complex & Intelligent Systems, 6, 237-249.
9. Baldea, M., & Harjunkoski, I. (2014). Integrated production scheduling and process control: A systematic review. Computers & Chemical Engineering, 71, 377-390.
10. Lohmer, J., & Lasch, R. (2021). Production planning and scheduling in multi-factory production networks: a systematic literature review. International Journal of Production Research, 59(7), 2028-2054.

11. Folch, J. P., Tsay, C., Lee, R., Shafei, B., Ormaniec, W., Krause, A., ... & Mutny, M. (2024). Transition constrained Bayesian optimization via Markov decision processes. Advances in Neural Information Processing Systems, 37, 88194-88235.
12. Garcia, F., & Rachelson, E. (2013). Markov decision processes. Markov Decision Processes in Artificial Intelligence, 1-38.
13. Boucherie, R. J., & Van Dijk, N. M. (2017). Markov decision processes in practice. Springer.
14. LaMar, M. M. (2018). Markov decision process measurement model. Psychometrika, 83(1), 67-88.
15. Steimle, L. N., Kaufman, D. L., & Denton, B. T. (2021). Multi-model Markov decision processes. IISE Transactions, 53(10), 1124-1139.
16. Khorolskyi, A., Hrinov, V., Mamaikin, O., & Demchenko, Y. (2019). Models and methods to make decisions while mining production scheduling. Mining of Mineral Deposits, 13(4), 53-62.
17. Yang, H., Li, W., & Wang, B. (2021). Joint optimization of preventive maintenance and production scheduling for multi-state production systems based on reinforcement learning. Reliability Engineering & System Safety, 214, 107713.
18. Tirkolaee, E. B., Aydin, N. S., & Mahdavi, I. (2022). A hybrid biobjective markov chain based optimization model for sustainable aggregate production planning. IEEE Transactions on Engineering Management, 71, 4273-4283.
19. Sodachi, M., Pirayesh, A., & Valilai, O. F. (2024). Using Markov Decision Process Model for Sustainability Assessment in Industry 4.0. IEEE Access.
20. Zhou, L., Wang, F., Wang, N., & Yuan, T. (2021, August). Application of industrial robots in automated production lines under the background of intelligent manufacturing. In Journal of Physics: Conference Series (Vol. 1992, No. 4, p. 042050). IOP Publishing.
21. Lai, R., Lin, W., & Wu, Y. (2018). Review of research on the key technologies, application fields and development trends of intelligent robots. In Intelligent Robotics and Applications: 11th International Conference, ICIRA 2018, Newcastle, NSW, Australia, August 9–11, 2018, Proceedings, Part II 11 (pp. 449-458). Springer International Publishing.
22. Li, B. H., Hou, B. C., Yu, W. T., Lu, X. B., & Yang, C. W. (2017). Applications of artificial intelligence in intelligent manufacturing: a review. Frontiers of Information Technology & Electronic Engineering, 18(1), 86-96.
23. Soori, M., Dastres, R., Arezoo, B., & Jough, F. K. G. (2024). Intelligent robotic systems in Industry 4.0: A review. Journal of Advanced Manufacturing Science and Technology, 2024007-0.
24. Ojstersek, R., Brezocnik, M., & Buchmeister, B. (2020). Multi-objective optimization of production scheduling with evolutionary computation: A review. International Journal of Industrial Engineering Computations, 11(3), 359-376.
25. FAN, Y., & WU, W. (2020). APPLICATION OF HIGH-ORDER MARKOV DECISION MODEL IN THE OPTIMAL SCHEDULING OF RAW MATERIALS IN PRODUCTION WORKSHOP. Academic Journal of Manufacturing Engineering, 18(1).
26. Ghasemi, A., Farajzadeh, F., Heavey, C., Fowler, J., & Papadopoulos, C. T. (2024). Simulation optimization applied to production scheduling in the era of industry 4.0: A review and future roadmap. Journal of Industrial Information Integration, 100599.
27. Ma, H., Huang, X., Hu, Z., Chen, Y., Qian, D., Deng, J., & Hua, L. (2023). Multi-objective production scheduling optimization and management control system of complex aerospace components: a review. The International Journal of Advanced Manufacturing Technology, 127(11-12), 4973-4993.
28. Shang, J., Tian, Y., Liu, Y., & Liu, R. (2018). Production scheduling optimization method based on hybrid particle swarm optimization algorithm. Journal of Intelligent & Fuzzy Systems, 34(2), 955-964.
29. Michel Minoux. (2018). Robust and stochastic multistage optimisation under Markovian uncertainty with applications to production/inventory problems. International Journal of Production Research,56(1-2),565-583.
30. Yongxin Lu,Yiping Yuan,Jiarula Yasenjiang,Adilanmu Sitahong,Yongsheng Chao & Yunxuan Wang. (2025). An Optimized Method for Solving the Green Permutation Flow Shop Scheduling Problem Using a Combination of Deep Reinforcement Learning and Improved Genetic Algorithm. Mathematics,13(4),545-545.