ISSN: 2688-9323

A Reinforcement Learning Scheduling Method for Material Handling on Assembly Lines

Jie Yuan, Yongzhuo Yang and Jiawei Zeng*

School of Management, Shanghai University, Shanghai 200444, China

Abstract

On-time and efficient material handling system ensures the continuous and stable operation of assembly manufacturing. To dynamically respond to the changes of the assembly line status and effectively balance the productivity and energy consumption of mixed-flow assembly, this paper proposes a reinforcement learning scheduling algorithm, which incorporates the design of system states, action policies, and reward functions. The simulation experimental results show that the reinforcement learning scheduling model can optimize the material handling scheduling better and effectively reduce the handling distance while ensuring the continuous and stable operation of the assembly line to achieve the maximum output.

Keywords

Workshop Material Handling System; Reinforcement Learning; Q Learning.

1. Introduction

Scheduling of material handling system is an essential part of the production control system in manufacturing enterprises, which connects different processes and workshops. The on-time and efficient material handling scheduling can effectively improve the productivity and economic efficiency of enterprises. In order to avoid assembly line downtime caused by material distribution delay and reduce handling costs and the pressure of parts inventory on the assembly line, scholars have proposed many optimization methods[1-2], such as mathematical planning methods[3-4], heuristic algorithms[5-7] and various intelligent search methods[8-10]. There may be some random events in the actual manufacturing process, such as changes in product ratios or machine and equipment failures, which may cause random changes in system parameters and states, making the problem of workshop material handling scheduling dynamic. For these dynamic scheduling problems, scholars have also proposed some scheduling methods based on hybrid intelligent search methods[11]. Due to the complexity of the problem, there is still a great need for a feasible and effective system theory and methodology. In recent years, reinforcement learning[12]has attracted much attention as a new research method for unsupervised learning in machine learning. It can discover the optimal sequence of behaviors in the current state through uncertain environmental rewards and realize online learning in a dynamic environment, which provides a new way to solve large-scale dynamic optimization problems. Therefore, this paper proposed a Q learning scheduling algorithm based on the reinforcement learning model and effectively improve the scheduling performance on workshop issues.

2. Material Handling Scheduling Problems

The assembly line is usually divided into different material handling areas according to the facility planning in the workshop material handling operation. The work between each material handling area is independent and does not interfere with each other. Therefore, in order to simplify the scheduling problem, this paper only studies the material handling scheduling

ISSN: 2688-9323

problem in one material handling area. The layout of an automotive assembly line is shown in Figure 1. A material handling area is equipped with a material supermarket and handling equipment, i.e., a multi-load trolley. This facility layout allows the multi-load trolley to choose a single departure point with multiple destinations when performing handling tasks. In other words, after the trolley departs from the material supermarket, it can directly distribute the material to the destination workstation. Therefore, the scheduling problem of the workshop material handling system studied in this paper refers to that figure out the material trolley scheduling scheme under the variable path by obtaining the state information of the assembly line. On the scheduling moment t, all scheduling information, including the departure time of the trolley, the order of the parts to be handled, and the handling distance, should be given.

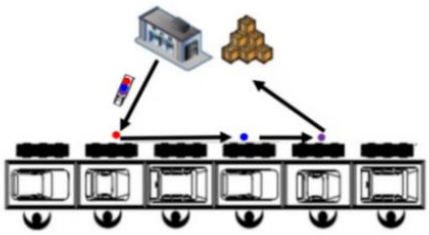

Figure 1. The layout of the automotive assembly line

Based on the requirements of the material handling system in automotive assembly lines, the basic assumptions of this paper are as follows.

(1) Handling cannot be interrupted, and no abandonment of parts distribution tasks.

(2) The multi-load trolleys do not break down during handling.

(3) The multi-load trolley travels at a constant speed.

(4) The maximum number of bins to be handled by the trolley at a time is N_c, and the same part cannot be repeated in a single handling.

(5) Each bin can only carry one type of part, and the capacity is a fixed value Q_p.

(6) The loading time l and the unloading time r of the bin are fixed.

3. Reinforcement Learning Model for Scheduling

3.1. Q Learning

This paper chooses a Q learning algorithm for the material handling scheduling problem to construct the reinforcement learning model. The Q learning algorithm uses the agent to take actions in each scheduling state and obtain rewards from environmental feedback to compute updated Q values, which are updated as follows.

$$Q(s,a) = Q(s,a) + \alpha\left(r(s,a) + \gamma \max_{a' \in A} Q(s',a') - Q(s,a)\right) \quad (1)$$

r is the reward value obtained by the system after choosing action a in state s; A is the optional action set; α is the learning rate of the agent; γ is the discount factor, the smaller γ is, the more focus on the current reward. Q is updated and stored into Q-table, after which it is continuously iteratively updated to approximate the objective function Q*. If an agent's action receives a positive reward from the environment, the tendency to perform it is enhanced; conversely, the tendency is reduced. Finally, the agent maximizes the long-term cumulative reward and learns the optimal behavioral strategy. Mapping the problem to reinforcement learning is the crucial process while using the Q learning algorithm to solve material handling scheduling problems on the workshop, which includes the following aspects: the setting of system state and action, and the setting of reward function.

ISSN: 2688-9323

3.2. State-space Design

Since the current manufacturing shop has higher material handling scheduling requirements, this paper proposes a state space and action group space considering part relaxation time to meet more robust real-time requirements. In order to simplify the scheduling decision model of the workshop material handling system while ensuring the accuracy of the decision model, this paper refers to the vital information in the material handling scheduling process, including the parts lineside inventory on each assembly workstation and the assembly task information for the future period, as well as the parts and quantities required for these tasks. The slack time ST_p for each part can be calculated with the system information, which is used as the system state characteristic of the scheduling decision model to show the urgency of the handling task. Therefore, the system state is defined as follows.

$$S = [ST_1, ST_2, ..., ST_P] \tag{2}$$

In order to keep the same production pace with the assembly line, ST_p is an integer multiple of the assembly line's cycle time, CT. In addition, if the parts lineside inventory cannot meet the assembly demand for a future period, i.e., ST_p is less than N_p*CT, then add the part to buf, which is the current task sequence to be handled. N_p is the number of forward-looking products, which is the number of products produced in a future period from known system information.

3.3. Action-space Design

The scheduling performance is directly influenced by the multi-load trolley's decision at the scheduling moment. Therefore, the action space must be set up considering all possible cases of actions, giving the agent sufficient choice space to learn the action sequence with the best long-term scheduling performance. Considering that N_c, the upper limit of the number of bins handled by a multi-load trolley at a time is usually 3, the action group is set as follows.

$$a = \{a_1, a_2, a_3, a_4, a_5\} \tag{3}$$

$\{a_1\}$: the multi-load trolley does not carry any parts; $\{a_2\}$: the multi-load trolley handles one kind of parts; $\{a_3\}$: the multi-load trolley handles two kinds of parts. Since the handling distance changes by the order when the kinds of parts to be handled reach three. So $\{a_4\}$: the multi-load trolley handles three kinds of parts, and the handling order is determined by the slack time of parts. $\{a_5\}$: the multi-load trolley handles three kinds of parts, and the handling distance determines the order. The selection of parts for $\{a_i\}$ is first generated from buf_p, the sequence of tasks to be handled. If the kinds of parts in buf_p do not meet the required specifications, they are filled from bfl_p, the substitute task sequence. Parts in bfl_p are listed in ascending order of the current online inventory, represented by $\frac{LIp}{Qp}$, the ratio of lineside inventory to bin capacity. The departure time of the multi-load trolley in each $\{a_i\}$ can be calculated from the line look-ahead information, the product bill of materials, and the distance information of the line.

3.4. Reward Function Design

The system feedback obtained by the agent after selecting an action is reflected by the reward function, which guides the selection of the action sequence in the overall learning process by the positive or negative reward value. In the material handling problem, the most important thing is to ensure the continuous and stable operation of the assembly line to get the maximum capacity of production. Secondly, the handling cost in the material transportation process is also an important consideration, mainly reflected by the handling distance. At the same time,

ISSN: 2688-9323

the size of the line side inventory of the assembly line also needs to be considered. In this paper, we choose the sparse reward function for the design, which contains the reward and punishment terms of three dimensions:

$$R = A * TS + B * Dis + C * \left[\left(sumLI_P^{'} \right) - sum\left(LI_P \right) \right] \tag{4}$$

TS is out-of-stock time, Dis is the handling distance, and LI is the lineside inventory. A, B, C are the weights of these three reward items, and the values of A, B, C are listed in descending order according to the priority of the optimization objectives.

4. Simulation Experimental Analysis

4.1. Simulation Assumptions

Since the movement of in-process products between workstations on the conveyor belt is synchronized, the workstation that finishes assembly first must wait for the upstream workstation to finish the current assembly task before assembling the following product. In order to keep the whole assembly line running smoothly, the assembly time CT for each workstation is assumed to be 72s. The parameter configurations of the simulation experiments are shown in Table 1 and 2.

Table 1. BOM, Q_p,and Dis

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}
M_1	1	1	1	1	0	0	1	1	0	0	1	0	0	1
M_2	1	0	0	0	1	1	1	0	1	0	0	1	0	1
M_3	1	0	0	0	1	0	1	1	0	1	0	0	1	0
Q_p	95	16	18	18	60	22	20	70	20	50	90	90	36	100
Dis/m	135	132	130	110	108	98	96	100	102	111	114	117	132	136

Table 2. Configuration parameters

parameters	value	instructions
A	104	Weight of TS
B	102	Weight of Dis
C	1	Weight of changes in online inventory
P_t	3	Number of product models
P_m	(0.2,0.5,0.3)	Product Ratio
N_p	18	Number of forward-looking products
S_t	6	Number of workstations
P	14	Number of parts types
N_c	3	The maximum load capacity of small vehicles
CT (s)	72	Assembly tempo
l (s)	37	The loading time of parts
r (s)	43	Unloading time of parts
v (m/s)	3	Trolley speed

ISSN: 2688-9323

4.2. Simulation Experiments and Analysis of Results

The simulation model of automobile mixed-flow assembly line is built through Arena simulation software, and the VBA module in it is used for secondary development to complete the interaction process between the scheduling method and the simulation model to realize the dynamic scheduling of the automobile assembly line. Furthermore, compared with the common dynamic scheduling method based on genetic algorithm, the performance of the dynamic scheduling method based on reinforcement learning proposed in this paper is verified. The simulation duration is 100h.

The following metrics are chosen to evaluate the performance of the scheduling method considering the production characteristics of the automotive assembly line: average lineside inventory ALI, throughput TH, total handling distance TD, the total number of out-of-stocks NS, and comprehensive cost CI. In actual manufacturing, the assembly line runs continuously and steadily to utilize production efficiency. Therefore, the weight of the cost of the out-of-stock, handling distance, and inventory for workshop material handling scheduling is reduced in order. Comprehensive cost CI is expressed as follows. a, b, c are the weights of NS, DS, and the sum of the average lineside inventory, and the values are set as follows: $a=10^6$, $b=10^2$, $c=1$.

$$CI = a * NS + b * DS + c * \sum_{p \in P} ALI_p \tag{5}$$

The simulation results are shown in Table 3, Table 4 and Figure 2. In the experiment of 100 hours of simulation, both the reinforcement learning algorithm RL and the genetic algorithm GA do not run out of stock and achieve the maximum yield. The handling distance of RL is shorter than GA, while the average online inventory is slightly higher than GA. Nevertheless, the comprehensive scheduling cost of the RL method is lower. Since the simulation parameters are set the same, the material demand is smooth throughout the scheduling process. The lineside inventory is slightly higher because the agent may choose the case where multiple parts are transported at once rather than the single part several times to pursue shorter handling distances.

In summary, the RL method has better balanced the cost indexes under considering out-of-stock cost, handling cost, and inventory cost. The RL method gave full play to the carrying advantages of the multi-load material trolley, thus achieving better scheduling performance on the whole.

Table 3. Average online inventory ALI of GA and RL

scheduling method	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}
GA	51	9	9	10	32	12	63	37	12	26	47	47	20	54
RL	50	9	9	10	33	12	67	37	12	26	47	47	20	52

Table 4. Scheduling performance of GA and RL

Scheduling method	TH	Sum(ALI)	TD	NS	TS	CI
GA	4995	429	97893	0	0	9.79×10^6
RL	4995	431	95236	0	0	9.52×10^6

ISSN: 2688-9323

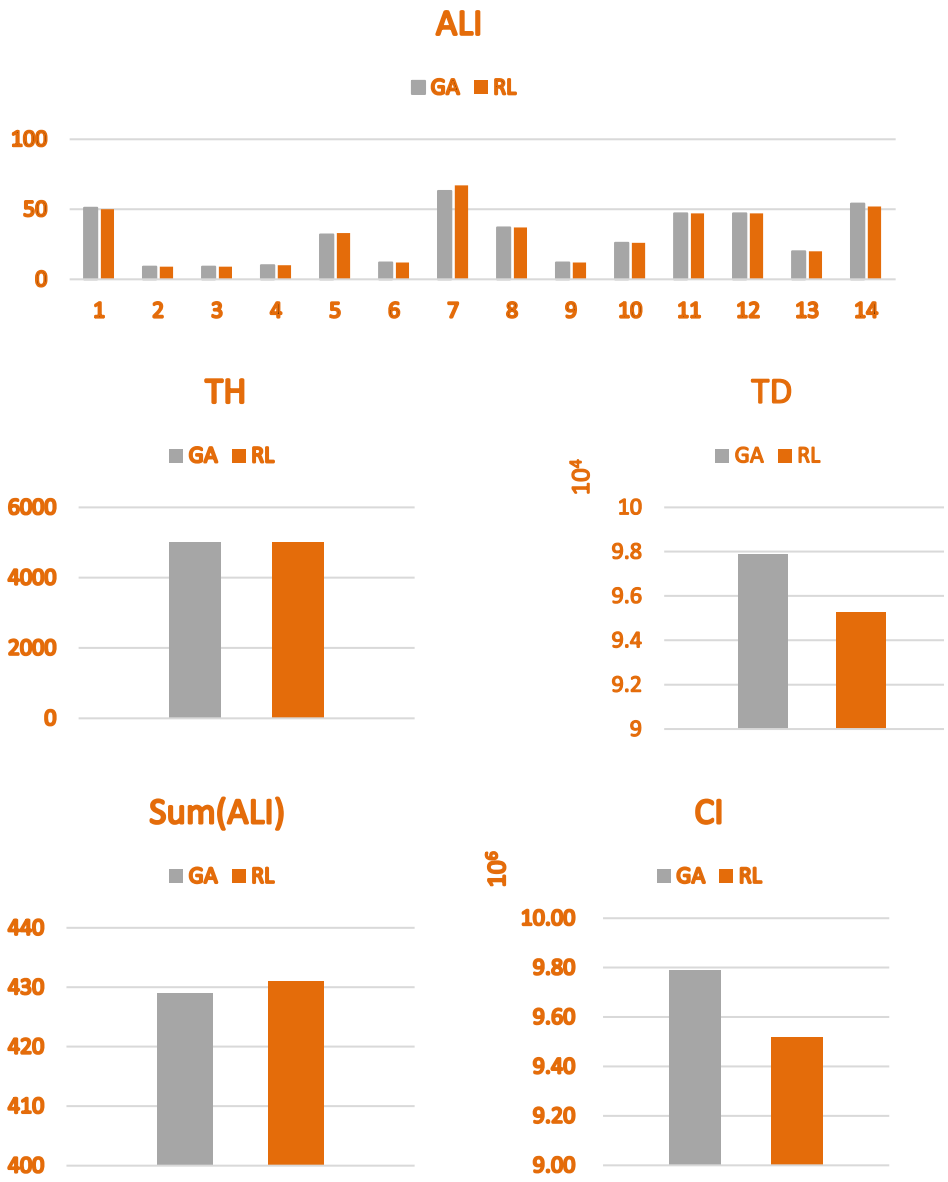

Figure 2. Scheduling performance of RL and GA

5. Conclusion

This paper applies the reinforcement learning algorithm to the scheduling problem of material handling in assembly lines to achieve real-time dynamic scheduling of the workshop material handling system. Finally, it is verified through simulation that the reinforcement learning scheduling method proposed in this paper can effectively avoid the out-of-stock stoppage problem caused by untimely distribution to ensure the smooth operation of the assembly line. Moreover, it also optimizes the output, handling distance and greatly reduces material distribution costs while maintaining high output and low online inventory.

References

[1] Fu Jianlin,Zhang Hengzhi,Zhang Jian,Jiang Liangkui. A review of automatic guided vehicle scheduling optimization research[J]. Journal of Systems Simulation, 2020, 32(09): 23-25.

[2] Cao LJ, Liu Y. New Advances in Automatic Guided Vehicle Scheduling for Manufacturing Shops [J]. Computer Engineering and Applications, 2021: 1-10.

ISSN: 2688-9323

[3] Miyamoto T, Inoue K. Local and random searches for dispatch and conflict free routing problem of capacitated AGV systems[J]. Computers & Industrial Engineering, 2016, 91: 1-9.

[4] Huo, K. G., Zhang, Y. Q., Hu, C. H.. Research on multi-load AGV scheduling problem in automated container terminals[J]. Journal of Dalian University of Technology, 2016,56(3): 244-251.

[5] Ho Y C, Liu H C, Yih Y. A multiple-attribute method for concurrently solving the pickup-dispatching problem and the load-selection problem of multiple- load AGVs[J]. Journal of Manufacturing Systems, 2012, 31(3):288-300.

[6] Xiao HN, Lou Peihuang, Man Zengguang, et al. A real-time multi-attribute task scheduling approach for automated guided vehicle systems[J]. Computer Integrated Manufacturing Systems, 2012, 18(10): 2224-2230.

[7] Namita S, Sarngadharan P V, Pal P K. AGV Scheduling for Automated Material Distribution: a Case Study[J]. Journal of Intelligent Manufacturing (S0956-5515), 2011, 22(2): 219-228.

[8] Yang Y S, Zhong M S, Dessouky Y, et al. An Integrated Scheduling Method for AGV Routing in Automated Container Terminals [J]. Computers & Industrial Engineering (S0360-8352), 2018, 126: 482-493.

[9] Zhang F Q, Li J J. An Improved Particle Swarm Optimization Algorithm for Integrated Scheduling Model in AGV-Served Manufacturing Systems[J]. Journal of Advanced Manufacturing Systems (S0219-6867), 2018, 17(3): 375-390.

[10] Jin J, Zhang X H. Multi AGV scheduling problem in automated container terminal[J]. Journal of Marine Science and Technology-Taiwan (S1023-2796), 2016, 24(1): 32-38.

[11] Zhang J, Ding G F, Zou Y S, et al. Review of Job Shop Scheduling Research and its New Perspectives Under Industry 4.0[J]. Journal of Intelligent Manufacturing (S0956-5515), 2017: 1-22.

[12] Xue T F, Peng Z, Yu H B. A Reinforcement Learning Method for Multi-AGV Scheduling in Manufacturing[C].2018 IEEE International Conference on Industrial Technology (ICIT). Los Alamitos, CA: IEEE Computer Society, 2018: 1557-1561.