Flexible Services and Manufacturing Journal (2023) 35:1116–1138
https://doi.org/10.1007/s10696-023-09498-7

# A digital twin-driven flexible scheduling method in a human–machine collaborative workshop based on hierarchical reinforcement learning

Rong Zhang[1] · Jianhao Lv[1] · Jinsong Bao[1] · Yu Zheng[2]

Accepted: 16 April 2023 / Published online: 17 May 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract
Under the influence of the global COVID-19 pandemic, the demand for medical equipment and epidemic prevention materials has increased significantly, but the existing production lines are not flexible and efficient enough to dynamically adapt to market demand. The human–machine collaboration system combines the advantages of humans and machines, and provides feasibility for implementing different manufacturing tasks. With dynamic adjustment of robots and operators in the production line, the flexibility of the human–machine collaborative production line can be further improved. Therefore, a parallel production line is set up as a parallel community, and the digital twin community model of the intelligent workshop is constructed. The fusion and interaction between the production communities enhance the production flexibility of the manufacturing shop. Aiming at the overall production efficiency and load balancing state, a digital twin-driven intra-community process optimization algorithm based on hierarchical reinforcement learning is proposed, and as a key framework to improve the production performance of production communities, which is used to optimize the proportion of human and machine involvement in work. Finally, taking the assembly process of ventilators as an example, it is proved that the intelligent scheduling strategy proposed in this paper shows stronger adjustment ability in response to dynamic demand as well as production line changes.

**Keywords** Digital twin communities · Human–machine collaboration · Reinforcement learning · Flexible scheduling

Extended author information available on the last page of the article

## 1 Introduction

In recent years, the demand for medical equipment has grown significantly. A rapidly changing market leads to the elimination of traditional manufacturing systems. Therefore, the manufacturing system must quickly respond to the insertion of new tasks, complete rescheduling, and improve flexibility.

A flexible manufacturing system (FMS) is an automated product manufacturing management system composed of an automatic processing management device, material storage device, and a control management system. It adapts to the transformation of processing objects through reconfigurable manufacturing systems (RMS), dynamic scheduling (DS), reconfiguration of manufacturing source (ROMS), path planning (PP), and human–robot collaboration (HRC) in the manufacturing workshop where operators and machines coexist.

RMS combines a dedicated manufacturing line (DML) with FMS to improve the ability of the system to adapt to changes. This capability can be further improved by highly parametric models and reconfigurability enhancements (Wikarek et al. 2019; Zhang et al. 2019). DS refers to dynamically updating the decision-making plan according to environmental changes to ensure the completion of production (Hu et al. 2020; Luo et al. 2021). ROMS refers to generating corresponding allocation plans for manufacturing resources such as human resources, manufacturing equipment resources, and technical resources according to changing requirements (Zhang et al. 2021). PP refers to the optimization of indicators such as transportation distance, time, cost of vehicles, robots, and other machine transportation plans according to environmental changes (Zhang et al. 2021). HRC combines manual assembly and automated assembly to improve efficiency on the premise of ensuring flexibility to achieve safe, economical, and efficient assembly tasks. One of the most critical aspects is the optimal subtask assignment between humans and robots, and the Knowledge Transfer in Robotic Assembly Sequence Planning (KT-RASP) approach provides a new idea for this problem solution (Rahman and Wang 2018; Rodríguez et al. 2020).

The concept of FMS has been widely considered since it was proposed. Although extensive research has been carried out on FMS, there are some limitations:

1. Conventional assembly lines lack the flexibility to proactively adapt to different production tasks. When urgent orders appear, it is difficult to efficiently assess conflicts between current ongoing orders and newly arrived urgent orders, especially when machine sharing is required. In addition, factors such as sudden machine breakdowns and changes in manufacturing processes have become obstacles to rapid adjustment of assembly lines.
2. The existing FMS research based on the neighborhood search algorithm has some limitations and is difficult to adapt to the complex industrial environment. Combinatorial optimization of flexible manufacturing systems is a nonlinear hybrid discrete optimization problem with complex coupling relations. However, with the increase of the industrial scale, the structure of tasks becomes complex, and

the neighborhood search algorithm is limited to a fixed scheduling plan. The defect of poor anti-interference ability becomes obvious.

To facilitate the implementation of FMS, various multi-objective optimization algorithms have been applied for their dynamic nature. However, these algorithms usually have limitations and cannot be actively adapted to different multi-objective optimization problems. In contrast, reinforcement learning (RL), as a self-learning algorithm, is able to actively adapt the strategy to find a series of suitable solutions. In this paper, RL is combined with digital twins (DT) to monitor the actual assembly process and predict the status of personnel and equipment. RL is used to realize the suitable scheduling policy by self-learning and completing the inspection in the virtual space when new tasks are inserted, machine failures occur, and disturbing signals are generated.

The remaining content of this paper is organized as follows. In Sect. 2, the related work is presented. In Sect. 3, a problem description is introduced that defines the constraints that need to be satisfied during scheduling. Section 4 introduces a flexible scheduling system in the assembly workshop driven by digital twins, and in Sect. 5, a hierarchical reinforcement learning method is proposed to address the dynamic decision-making needs in flexible manufacturing. In Sect. 6, the effectiveness of the proposed method in the rescheduling process of human–machine collaboration is demonstrated by switching from generator assembly to ventilator assembly. Finally, conclusions are presented in Sect. 7.

## 2 Related work

### 2.1 Scheduling optimization

The scheduling optimization problem of FMS is an NP-hard problem. Researchers have tried to solve such problems by using a variety of solution methods such as neighborhood search algorithms (NSA), reinforcement learning (RL), and digital twin (DT). Literature discussing scheduling optimization is listed in Table 1.

NSA utilizes the domain structure for step-by-step optimization and is used to sequence predetermined paths and jobs into a constrained combinatorial optimization problem. Dou et al. (2021) presented a multi-objective particle swarm optimization algorithm based on crowding distance and external Pareto solution archive. And the multi-objective optimization framework based on the fitness evaluation mechanism and adaptive local search strategy can improve the scheduling efficiency at a higher level (He et al. 2021). Similarly, to obtain the shortest completion time, a multi-objective mathematical programming model has been developed through the pre-distribution of inbound container clusters and outbound container clusters (Hu et al. 2019).

With the increase of industrial scale, the task structure becomes more and more complex. As a kind of local optimization algorithm, the neighbor search algorithm is easily disturbed by the local optimal solution and is difficult to

**Table 1** Scheduling optimization

| Method | Purpose | Brief introduction | Source |
|---|---|---|---|
| NSA | Efficiency | Multi-objective particle swarm optimization | Dou et al. (2021) |
| | Energy-efficient | Multi-objective optimization framework | He et al. (2021) |
| | Shortest time | Mathematical programming model | Hu et al. (2019) |
| Heuristic | Bi-object hybrid production | Meta-heuristic | Hoseinpour et al. (2020) |
| | Bi-object hybrid production | Genetic algorithm + particle swarm optimization (GAPSO) hybrid algorithm | Hoseinpour et al. (2021) |
| RL | Flexible maintenance | Flexible maintenance strategy by RL | Wang et al. (2021) |
| | Scheduling efficiency | Cyber-physical integration by deep RL | Zhou et al. (2021) |
| | Scheduling efficiency | Multi-agent system by RL | Kim et al. (2020) |

adapt to the multi-objective optimization requirements. The defect of the poor anti-interference ability of the neighborhood search algorithm becomes obvious. For the bi-object hybrid production problem, heuristic methods are used to solve the more complex problem when the number of sources and products increases (Hoseinpour et al. 2020, 2021).

Reinforcement learning algorithms do not rely on knowledge of the distribution of uncertainty, and therefore do not require predictive modeling to guide the agent on how to act in the environment to maximize reward. Wang et al. (2021) proposed a flexible preventive maintenance strategy to proactively cope with machine failures. Zhou et al. (2021) presented new cyber-physical integration in smart factories for online scheduling of low-volume/high-mix orders. Kim et al. (2020)presented a smart manufacturing system using a multi-agent system and reinforcement learning, which is characterized by machines with intelligent agents to enable a system to have autonomy in decision-making, sociability to interact with other systems, and intelligence to learn dynamically changing environments.

Reinforcement learning effectively solves the problem of dependence on early scheduling plans. When the physical state changes continuously, the iterative efficiency of reinforcement learning is higher than the update efficiency of the state due to the delay of the data transmission, which will lead to the invalidity of the action and interfere with the learning iterative process. Therefore, there is a need for a method for reducing the loss and delay of physical assembly data as it flows to a virtual environment. The digital twin realizes the mapping of the physical assembly process in the virtual assembly space and reduces the loss and delay of the physical assembly data when it flows to the virtual environment (Liu et al. 2023).

### 2.2 Human–machine coexistence in FMS

In a flexible manufacturing system with human participation, humans and machines have common goals, but also have their own goals. The common goal is to complete the overall assembly task efficiently, while the respective goal is to minimize the cost of humans and machines. The research on the task allocation method is the key to solving this problem, and some related studies are shown in Table 2.

Ham and Park (2021) proposed two exact central approaches by formulating the task allocating problem as a flexible job shop with a sequence-dependent setup to capture heterogeneous agents and travel time. Çil et al. (2020) proposed a bee algorithm that utilizes a newly employed bee phase to accelerate the evolution of the swarm and a new scout phase to escape from being trapped into local optima and produce a high-quality and diverse population. Disassembly sequence planning for human–robot collaboration is solved by a modified discrete bees algorithm based on Pareto (MDBA-Pareto) (Xu et al. 2020). By considering the task execution constraints, the variability in task execution by the human, and the job quality of the human, a two-layered architecture is proposed to dynamically schedule assigned tasks in a collaborative cell (Pupa et al. 2021).

With the development of human–robot collaboration technology, the structure of assembly tasks has become more complex (such as human–machine parallel execution of shared tasks), and the task space has become larger (such as multiple workers and multiple robots). Under the condition of fully considering the necessary human and robot information and assembly constraints, the human–robot collaborative assembly task assignment will have a high-dimensional and complex state space, which is difficult to solve by traditional combinatorial optimization methods.

Reinforcement learning techniques have shown increasingly significant advantages in dealing with high-dimensional complex spatial problems and large-scale decision-making problems, providing a new direction for overcoming the limitations of traditional task assignment methods, as shown in Table 3.

Lin et al. (2022) proposed a hidden semi-Markov model for a human-centered HRC assembly system by RL that enables the robot to accurately predict the human assembly rate. Ghadirzadeh et al. (2020) presented a reinforcement learning-based framework for human-centered collaborative systems. El-Shamouty et al. (2020) proposed a framework that uses deep RL as an enabling technology to enhance the intelligence and safety of robots in HRC scenarios. Yu et al. (2020) formatted the human–robot collaborative assembly working process into a chessboard, and the

**Table 2** Traditional assembly task allocation method

| Task | Purpose | Brief introduction | Source |
|---|---|---|---|
| Task allocation | Precisely allocated | Two exact central optimization | Ham and Park (2021) |
| | Minimize times | Bee algorithm | Çil et al. (2020) |
| Sequence planning | Minimize cost | MDBA-Pareto algorithm | Xu et al. (2020) |
| | Effectiveness | Optimization architecture | Pupa et al. (2021) |

**Table 3** Reinforcement learning for HRC

| Task | Purpose | Brief introduction | Source |
|---|---|---|---|
| HRC assembly | Reduce time | HRC assembly system by RL | Lin et al. (2022) |
| Working | Minimize the time | Reinforcement learning-based framework | Ghadirzadeh et al. (2020) |
| | Safety | Deep RL-based framework | El-Shamouty et al. (2020) |
| | Efficiency | Monte Carlo treesearch (MCTS) algorithm | Yu et al. (2020) |

selection of moves in the chessboard is used to analogize the decision-making by both humans and robots in the HRC assembly working process.

Reinforcement learning overcomes the limitations of traditional combinatorial optimization methods, while the accuracy and efficiency of data acquisition seriously restrict the performance of reinforcement learning. Digital twin technology makes it possible to solve this problem due to its high capability for virtual–real mapping. Therefore, our reinforcement learning scheduling framework is based on digital twins.

## 3 Problem description

The flexible workshop scheduling system contains $n$ jobs and $m$ machines. Because the availability of the machine is limited and subject to real-time operational status, it is necessary to dynamically adjust the production plan to adapt to the actual production conditions, and then generate a new scheduling plan. In the actual assembly workshop, it is also necessary to consider the state of the operator. Assuming that the number of operators performing operations is $h$, their working state will be a factor affecting the scheduling strategy. At the same time, other disturbances (such as the change of processing time and the insertion of new jobs) will also affect the initial scheduling. Due to the flexible insertion of new jobs, the rescheduling process will be frequent and unstable. The notations related to assembly line operation in the assembly workshop are shown in Table 4.

The dynamic scheduling problem is constrained by relevant assumptions:

1. Machine and operator are available and idle at the beginning of the assembly.
2. A single machine can only handle one job at a time.
3. An operation of a job needs to wait until the previous operation is completed before starting.
4. When operating on the machine, the job will not be terminated due to machine failure.
5. The processing time of each operation is pre-designed, but it can be changed during actual assembly.

It is assumed that the time of machine maintenance is known.

**Table 4** Utilized notations

| Type | Notation | Explanation |
|---|---|---|
| Indexes | $J_i$ | The $i$th job $(1 < i < n)$ |
| | $M_i$ | The $j$th machine $(1 < i < m)$ |
| | $H_i$ | The $j$th operator $(1 < i < h)$ |
| | $O_{ik}$ | The $k$th operation of $J_i\left(1 \le k \le l_o\right)$ |
| Sets | $\boldsymbol{M}$ | Machines in the workshop |
| | $\boldsymbol{J}$ | Jobs in the workshop |
| | $\boldsymbol{H}$ | Operators in the workshop |
| Parameters | $l_o$ | Number of operations of $J_i$ |
| | $E$ | The limit value of energy consumption |
| | $t$ | The time to start assigning jobs |
| | $p$ | The number of parameters related to the human state |
| | $q$ | The number of parameters related to the machine state |
| | $\sigma$ | The pre-set tolerance range |
| Decision variables | $O_i$ | Completion time of $J_i$ in original scheduling |
| | $C_i$ | Completion time of $J_i$ in rescheduling |
| | $D_{all}$ | The upper limit of total delay |
| | $S_{O_{ik}}{}'$ | Start time of $O_{ik}$ in original scheduling |
| | $S_{O_{ik}}$ | Start time of $O_{ik}$ in rescheduling |
| | $P_{O_{ik}}$ | Processing time of $O_{ik}$ in rescheduling |
| | $W_{M_i}$ | The waiting time of $M_i$ |
| | $W_{H_i}$ | The waiting time of $H_i$ |
| | $E_{J_j}$ | The energy consumption caused by $M_j$ in $J_j$ |
| | $\alpha_{M_i}$ | The utility rate of $M_i$ |
| | $\alpha_{H_i}$ | The utility rate of $H_i$ |
| | $\alpha_{ave}$ | The minimum average utility rate |
| | $U_i$ | 1 if $J_j$ is assigned, 0 otherwise |
| | $U_{ijt}$ | 1 if $J_j$ is assigned to $M_i$ or $H_i$ at time $t$, 0 otherwise |

The dynamic scheduling process needs to meet the following constraints:

$$Min\left(Max_{1 \le x \le n}\left(C_i - O_i\right) + |S_{Oik} - S_{Oik}{}'|\right) \tag{1}$$

$$Min\left(\sum_{i=1}^{m} W_{M_i}\right) \tag{2}$$

$$Min\left(\sum_{i=1}^{h} W_{H_i}\right) \tag{3}$$

$$\sum_{i=1}^{n} Max\{\left(C_i - O_i\right), 0\} \le D_{all} \tag{4}$$

$$S_{O_{i(k+1)}} = S_{O_{ik}} + P_{O_{ik}} + W_{M_i} \quad \forall i \in [1, m], \forall k \in [1, l_o] \tag{5}$$

$$\frac{\sum_{i=1}^{m} \alpha_{M_i}}{m} \ge \alpha_{ave} \tag{6}$$

$$\frac{\sum_{i=1}^{h} \alpha_{H_i}}{h} \ge \alpha_{ave} \tag{7}$$

$$\sum_{i=1}^{n} E_{J_i} \le E \tag{8}$$

$$\sum_{i=1}^{m} U_i \ge 1 \quad \forall j \in [1, n] \tag{9}$$

$$\sum_{i=1}^{h} U_{ijt} \le 1 \quad \forall j \in [1, n] \tag{10}$$

$$\sum_{i=1}^{m} U_{ijt} \le 1 \quad \forall j \in [1, n] \tag{11}$$

Equations (1)–(3) are the objective functions, and Eqs. (4)–(11) are the basic conditions to be met during scheduling. Specifically, Eq. (1) represents minimizing the maximum completion time after triggering rescheduling, while Eqs. (2)–(3) represent minimizing the total waiting time of the operator and machine during assembly. Equation (4) means that the total delay time of assembly operation is within a reasonable range, and Eq. (5) shows that after rescheduling, the start time of the next task is the sum of the start time of the previous task, processing time, and machine waiting time. Equations (6)–(7) require that the utilization rate of the operator and machine in the assembly workshop is higher than the average utilization rate. Equation (8), from the perspective of energy consumption, limits the total energy consumption of the machine to be maintained within a reasonable range. Equations (9)–(11) define the relationship between operators and machines during assembly.

# 4 Flexible scheduling system of the assembly workshop driven by a digital twin

## 4.1 Digital workshop modeling

The assembly operation in assembly workshops is completed by humans and automatic machines. According to the product characteristics, it can be divided into multiple assembly processes. The data in the actual assembly process is multi-source, including assembly sequence data, resource scheduling data, and assembly accuracy data. However, there is limited connection between the physical assembly process and the virtual assembly process, which leads to insufficient utilization of digital information and the inability to make production adjustments in time under the conditions of machine failure and environmental disturbance. Through the effective use of the digital twin interaction mechanism and data fusion, the state of the complex machine can be monitored and predicted driven by the digital twin.

As shown in Fig. 1, the construction process of the digital twin model of the assembly workshop is carried out according to the characteristics of the assembly workshop. The DT model of the assembly workshop can be expressed as Eq. (12) and can be considered as a composition of five parts, including the real assembly plant, the virtual plant model, the functional modules, the twin data, and the linking relationships between different modules.

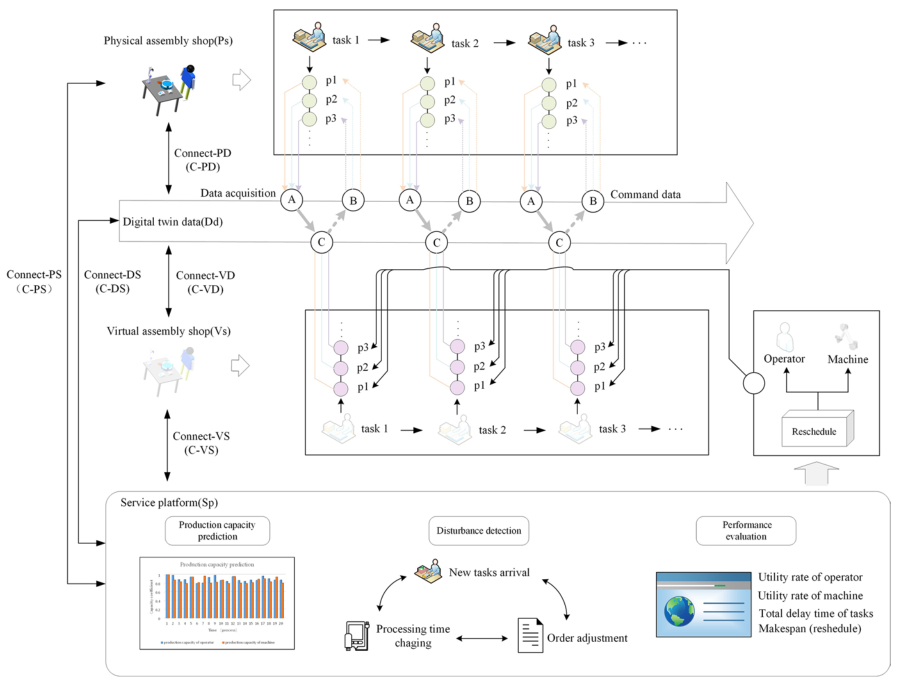


**Fig. 1** Digital twin modeling of the assembly workshop

$$A\_S_{DT} = [P_s + V_s + D_d + C + S_p] \tag{12}$$

$P_s$ represents the real assembly workshop where the corresponding sensors have been deployed;

$V_s$ represents the virtual assembly workshop defined by multiple models, including a geometric model, physical model, behavior model, and rule model;

$S_p$ represents the relevant functions provided by the DT, presents data through visual components, and outputs control instructions at the same time;

$D_d$ represents twin data, including $P_s$, $V_s$, $S_p$ data, and fusion data; and

$C$ represents the link between different modules. Here, it refers to the connection between the four modules $P_s$, $V_s$, $S_p$, and $D_d$.

In the digital twin model of the assembly workshop, the position and orientation of objects are captured by visual sensors. The power and temperature of machine operation are obtained by power and temperature sensors, respectively. The geometric model of the virtual workshop is presented using a three-dimensional (3D) model, the physical model is simulated by ANSYS software, and the behavior model is reflected in the assembly process after the implementation of scheduling rules. The rule model is the scheduling rules extracted from the data and experience.

As shown in Fig. 1, the data layer of the digital twin provides three different data interfaces. The port Ⓐ is used to collect sensing data from the physical workshop, including internal and external sensing data. The port Ⓑ is used to input command data to the controller of physical machines. The port Ⓒ is the interactive interface of the virtual workshop, which can accept data from the port Ⓐ for the status update of the virtual space, and at the same time, it can obtain feedback data from the virtual workshop and feed back to the physical workshop through the port Ⓑ to guide the actual assembly.

In addition to the virtual model, the digital twin workshop also includes a service platform. The platform is used to predict the production capacity of operators and machines, detect environmental disturbances, and evaluate performance. The platform also realizes the interaction with users through input and output interfaces, providing humans with the function of manually adjusting the virtual workshop.

$P_s$, $V_s$, $S_p$, and $D_d$ are connected mainly through C and enabled via data transfer protocols, database interfaces, and software interfaces. Therefore, the service platform further detects disturbance signals of the current environment (new job insertion, processing time change, order adjustment) based on the prediction of operator and machine productivity, and optimizes the scheduling policy through intelligent algorithms. Among them, the new scheduling policy will enter $V_s$ via $C - V_S$ to quickly verify the effect of the rescheduling policy.

### 4.2 State prediction

The real state (i.e., productivity) of the operators and the processing machine are the key elements that affect the assembly results. The digital twin assembly system checks whether the $P_s$ and $V_s$ are synchronized when acquiring data from the physical and virtual assembly spaces. Equation (13) illustrates that the data related to operators and machines is part of the assembly plant data and that the data contained in each operator or machine is a multidimensional vector, as in Eqs. (14) and (15). As shown in Eq. (16), $P_p$ and $P_q$ are used as a representation of the parameter, $p$ is the number of parameters related to the human state, while $q$ is the number of parameters related to the machine state. Equation (17) is used to determine whether the virtual and real assembly spaces in the digital twin are consistent, and $\sigma$ is a preset tolerance range.

$$[\boldsymbol{H},\boldsymbol{M}] \in PD, \quad [\boldsymbol{H}',\boldsymbol{M}'] \in VD \tag{13}$$

$$\boldsymbol{H} = \{H_1, H_2, \cdots, H_h\} \quad and \quad \boldsymbol{M} = \{M_1, M_2, \cdots, M_m\} \tag{14}$$

$$\boldsymbol{H}' = \{H_1{}', H_2{}', \cdots, H_h{}'\} \quad and \quad \boldsymbol{M}' = \{M_1{}', M_2{}', \cdots, M_m{}'\} \tag{15}$$

$$H_i(H_i{}') = \{P_1, P_2, \cdots, P_p\} \quad and \quad M_i(M_i{}') = \{P_1, P_2, \cdots, P_q\} \tag{16}$$

$$f = \begin{cases} 0, & ||PD - VD|| \le \sigma \\ 1, & ||PD - VD|| > \sigma \end{cases} \tag{17}$$

When $f = 0$, the operating data of devices will be extracted and recorded, which include operating power ($W$), operating temperature ($t$), etc. When the two states do not match, the fault location function will be activated to determine the source of the error. If the actual machine fails, the failure characteristics (including power and temperature) will be extracted and put into the failure characteristics database. As shown in Fig. 2, the relationship between working time, operator skills, and work efficiency is mined by comparing the state information of operators in the virtual and real space. The formed feature database or relational inference table will be used to train the neural network model and fuse the data from physical and virtual space to achieve the purpose of real-time state detection in the actual operation process.

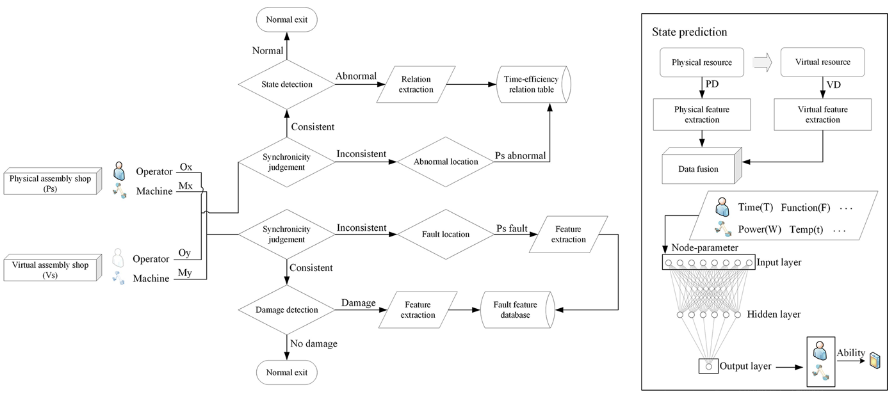


**Fig. 2** DT-based state prediction

### 4.3 Disturbance detection

Environmental disturbance information for shop floor operations includes adjustments to job orders, changes in processing times, and insertion of new job tasks. The digital twin is unique in detecting environmental disturbances. Data from physical and virtual spaces are graphically displayed once they are fused. Since the operating power of machines, etc., differs at different stages of the assembly plant, it is possible to compare the characteristic values of machine operation in the virtual and real spaces (e.g., operating power) and extract the characteristic segments $\Delta x$ with differences between them, to determine the type of environmental disturbance.

As shown in Fig. 3, each of the three different types of perturbations corresponds to a different characteristic segment $\Delta x$. When a new task is inserted, the operating power of the machine in the physical assembly plant will vary with the new task; when the task's processing time changes, the processing time of a particular job will be extended; when the job order changes, there will be a sudden change in the power of the machine in the physical assembly unit. Based on the above three different environmental disturbances, the digital twin system will detect the type of disturbance that occurs based on the characteristic curve of the imaginary and real space. Once the type of disturbance is identified, a new scheduling plan will be generated by the data management module, which will also act on the virtual assembly plant to complete the update of operator and machine status parameters.

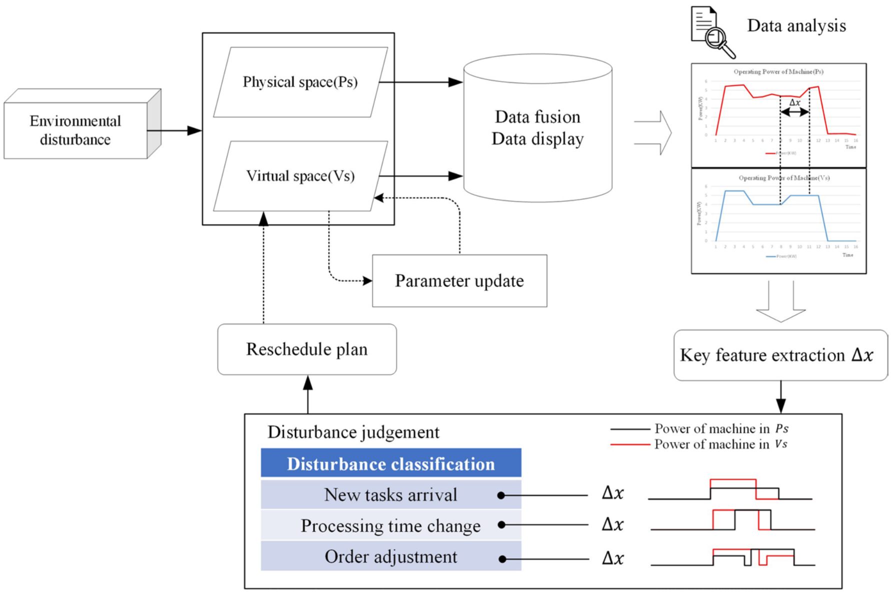


**Fig. 3** DT-based environmental disturbance detection

### 4.4 Scheduling performance evaluation

When a new scheduling plan is generated by the digital twin assembly plant due to a machine failure or environmental disturbance, the assembly plant generates a new scheduling plan. To evaluate the performance of the rescheduling strategy, the DT-based assembly shop then provides a specific performance evaluation module. As shown in Fig. 3, depending on the status of the human and machine, the new scheduling plan establishes a relationship between task–human–machine.

The virtual assembly cell will simulate the new scheduling plan learned by the reinforcement learning module and obtain the maximum completion time, total task delay, operator utilization, machine utilization, and total machine energy consumption for the assembly process. As shown in Fig. 4, when the maximum completion time and the maximum task delay after rescheduling are less than their respective set threshold, the operator utilization rate is greater than the average utilization rate, the machine utilization rate is greater than the average utilization rate, and the total energy consumption of the machine is less than the set maximum energy consumption, the DT will set the rescheduling plan as a feasible solution. Based on this, the system will mark these five types of results with different weights and make a weighted sum to obtain the score of the rescheduling plan. The rescheduling score is a reward value for the current scheduling policy, which is effective in guiding the scheduling system to find the optimal scheduling policy.

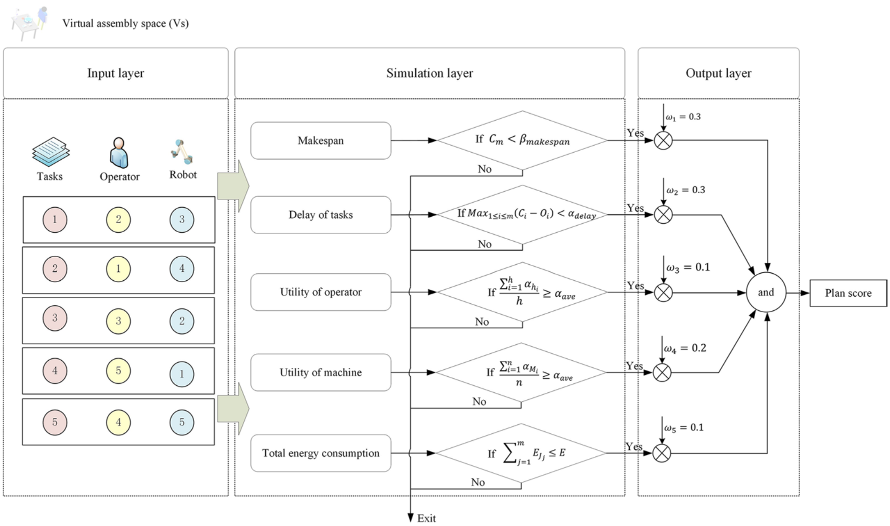


**Fig. 4** DT-based scheduling performance evaluation

# 5 Hierarchical RL-based flexible scheduling

## 5.1 Hierarchical RL framework

A two-stage reinforcement learning agent is constructed for a dynamic hierarchical deployment process on the shop floor for different assembly tasks, as shown in Fig. 5. In reinforcement learning, the most important role is the logic of the framework. The specific actor network and critical network are composed of a multilayer perceptron (MLP). The actor network takes the state information as input, and outputs an action *a1* according to a certain probability. The critic network, on the other hand, takes the {action, new state, reward} obtained after the interaction of the action with the environment as input to the network and obtains the value evaluation *q1* of the action. The specific process is as follows.

The inter-community scheduling agent selects a more suitable production line based on the line configuration information of the previous task and the actual production results, combined with the product requirements of the new task. At the same time, due to the different tasks, the original production lines cannot correspond to the existing product assembly requirements, so the inter-community agent needs to coordinate the relationship between production lines and deploy the corresponding number of production lines for the new task based on the actor_1 network. Among them, it will involve some cross-scheduling of production lines. Specifically, the agent within the community is the decision-maker, and through the actor_2 network, it collects information about the demand for new assembly tasks and the resource allocation of other production lines, makes corresponding actions, i.e., produces the machine execution orders of the current production line and the machine call demands of other production lines, and evaluates the results of the current selection in the critic_2 network.

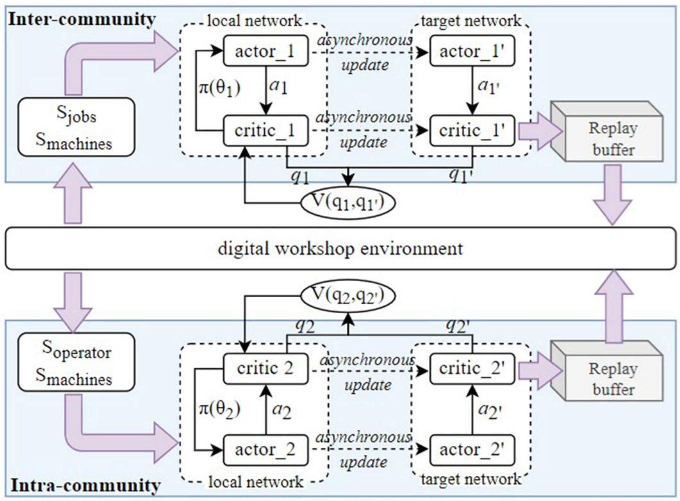


**Fig. 5** Hierarchical RL framework

The purpose of the target network is to perform a two-layer validation for the decision-making ability of the local network, i.e., to perform the next step of decision formation in the same framework based on the decision results of the local network and the state information updated after the decision execution, in order to evaluate the decision effect of the actor_1′ network using critical_1′. The network parameters of both actor_1′ and critic_1′ are copied from the local network using asynchronous updates to keep the action mechanism of the target network consistent with that of the local network, which can be regarded as a secondary evaluation of the performance of the local network. Finally, the critic_1 network and the critic_1′ network output the evaluation results q1 and q1′, respectively, calculate the error values between them, and feed the calculation results back to the critic_1 network to promote the stabilization of the decision–evaluation performance of the network. The mechanism of actor_2 and critic_2 is also similar.

## 5.2 State

In a dynamically task-oriented rescheduling system, an agent that schedules between communities and an agent that schedules within communities have a completely different set of malfunctions to consider when making scheduling decisions. The inter-community resource scheduling agent needs to make dynamic adjustments to different job task situations, while the intra-community agent mainly responds to machine failures and operator adjustments that require a secondary deployment of workstations. The job's state $S_{jobs}$, machine state $S_{macchines}$, and operator state $S_{operators}$ are shown in the following equations:

$$S_{jobs} = \{n, A\_S_{DT}, D_{all}, S_{O_{ik}}{}', E_{J_j}\} \tag{18}$$

$$S_{machines} = \{m, M', W_{M_i}, t, \alpha_{M_i}\} \tag{19}$$

$$S_{operators} = \{h, H', \alpha_{H_i}\} \tag{20}$$

## 5.3 Action and reward

In the scheduling system, the action of the agent is mainly expressed in the selection and matching of the workstation serial number with the machine serial number and operator serial number. The reward function is mainly affected by the combination of the maximum completion time, total task delay, operator utilization, machine utilization, and total machine energy consumption for the assembly process. Therefore, the reward function designed for agent scheduling decision evaluation is shown in Eq. (21).

$$R = \omega_1 Max\left(\sum_{i=1}^{n} C_i\right) + \omega_2(C_i - O_i) + \omega_3\frac{\sum_{i=1}^{h}\alpha_{Hi}}{h} + \omega_4\frac{\sum_{i=1}^{m}\alpha_{Mi}}{m} + \omega_5\sum_{j=1}^{n}E_{J_j} \tag{21}$$

## 6 Case study

The ventilator is characterized by high complexity, high precision, and multiple assembly constraints. After the outbreak of the epidemic, the targeted demand for medical machines surged, in line with the demand for personalized manufacturing in the context of Industry 4.0. All existing manufacturing and assembly lines need to be adjusted to dynamic customer demand in the short term, to complete the conversion most efficiently in order to adapt to customer demand as soon as possible.

In the assembly process of ventilators, unmanned production is not possible due to the complex structure and difficult arrangement of hoses and cables. However, in a manned manufacturing system, the chaotic distribution of tasks leads to inefficient dynamic adjustment of the production line. Therefore, hierarchical RL-based flexible scheduling is introduced, and the feasibility of the proposed method is verified by experimental analysis.

### 6.1 Human–machine collaborative workshop

In this case, to verify the effectiveness of the proposed method, the dynamic adjustment reconfiguration process from the simplified generator assembly plant (containing five production lines) to the respirator production plant is simulated. Similarly, to avoid the degree of experimental redundancy, the ventilator assembly process was categorized and organized so that the ventilator assembly line used for the study was also simplified. The original layout designed for the generator assembly line is shown in Fig. 6. The whole assembly process starts with M1 and ends with M11. M4 is the precondition for H1 execution, and the sequence of M5, M6, and M7 is also fixed. However, M5 and M4 are independent and can be executed separately at the same time. In addition to the uncertainty of the execution sequence, the operator can also be uncertain; for example, the work of M9 can be performed by a machine or by a human, and Fig. 6 only shows a form of the assembly sequence of the original production line. Before the production line was adjusted, each assembly line functioned with a total of five workers and 11 machines.

The traditional manual scheduling method is used as a comparative experiment. Its purpose is to prove that when facing new products, due to the lack of overall planning and information flow, the operation will be chaotic in the actual assembly process, resulting in extremely low efficiency. Human operators are often unable to find available machines in time. When the machine fails, the stagnation of progress and the lack of solutions will make workers anxious and reduce the safety of operators. In the hierarchical RL-based approach, all data is coordinated and planned at the DT

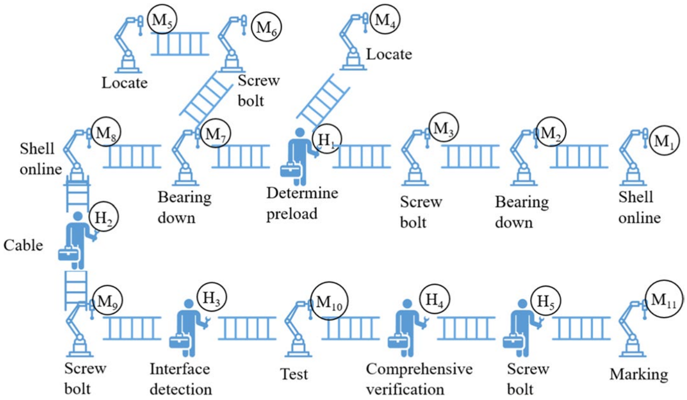


**Fig. 6** The layout of the original assembly line

side, dynamically adjusted based on reinforcement learning to ensure the efficiency of task execution, optimally allocate resources and progress, reduce unnecessary waiting time and energy expenses, and take into account the appropriate workload of operators while ensuring maximum efficiency to avoid excessive pressure.

## 6.2 Dynamic scheduling experiment

(1) Input and extract the state data of each production line in the existing generator assembly plant; the existing assembly cycle time is 500 s, and each production line contains 16 workstations. The total number of the existing workers and machines at the original production line is 16. And the station scheduling of the original task is shown in Fig. 7.

(2) In order to generate a dynamic scheduling task scheme, in the first step, initialization is performed, and initial data are loaded, such as priority relationships, minimum job times, task types and operational constraints, and layout information of each workstation. In the second step, the operational stability of each workstation is determined and the target evaluation function is developed; in the third step, the network parameters are initialized, the network training is started, and the update of the local network and the target network is performed every five steps. In the fourth step, based on the value of the error function, the decision behaviors with higher scores are stored in the replay buffer. Finally, the trained model is used in the scheduling planning process of the assembly line to guide the new assembly scheduling scheme.

(3) The comparative experimental results show that the assembly workshop is in the process of adjustment and upgrading in the face of dynamic tasks.

The comprehensive expenditure and result evaluation are shown in Table 5. The manual-based method belongs to the pre-planning type, and in the whole scheduling process, humans need to find malfunctions on their initiative and propose a scheduling scheme according to the machine status. The efficiency of the manual-based method is the lowest, while the delay during the development of a new scheduling

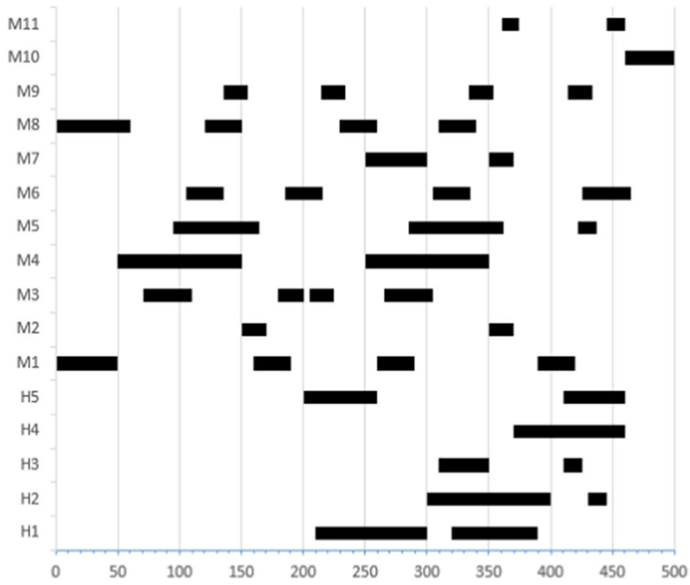


**Fig. 7** The station scheduling of the original task

**Table 5** Comparison of experimental results

| Method | Type | Time occupied | Total task delay | Operator utilization (%) | Machine utilization (%) |
|---|---|---|---|---|---|
| Manual-based | Pre-planning | 1760 | 530 | 200 | 180 |
| DT-based | Real-time planning | 1610 | 370 | 200 | 164 |
| Proposed method | Real-time planning | 1380 | 190 | 160 | 146 |

strategy is also the longest, at 530 s, due to the lack of real-time status monitoring. In addition, since human operations are more flexible and can adapt to more operational scenarios, the overall scheduling results are dominated by human involvement in order to ensure the executability of decisions when scheduling choices are made based on the human experience. And compared with the DT-based approach, the time to obtain the machine status is greatly reduced, and although the generated scheduling policy is still dominated by human involvement in the work, which reaches 200%, the timely detection of machine malfunctions effectively reduces the input of failed machines and lowers the machine occupancy ratio. The proposed method adds intelligent decision-making algorithms to the DT-based approach, which not only optimizes the proportion of human work, but also improves its overall operational efficiency. The overall comparison can be summarized in Table 6.

The comparison of task allocation is shown in Fig. 8. The flexible requirements of the three methods and the adaptive adjustment ability in dynamic scenarios are compared. It includes the comparison of two aspects: the scheduling planning for

**Table 6** Performance analysis of different methods

| Contrast items | Manual-based | Proposed method |
|---|---|---|
| Status monitoring | Check the equipment one by one, find the fault, check the fault information, and determine the replacement | Direct feedback of fault equipment number and historical service information in virtual space |
| Scheduling generation | Based on expert experience, the stability of decision-making effect is poor | Based on reinforcement learning algorithm, the decision is highly executable |
| Experience replay | Different experts have different experience bases and poor reusability | The experience of data processing is highly summarized and reusable |

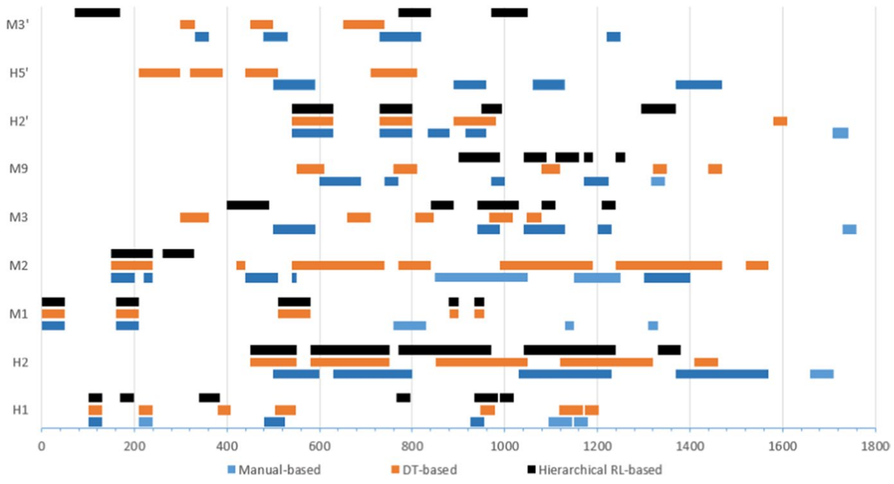


**Fig. 8** Comparison of task allocation

the conversion of the generator production line to the ventilator production line, and the anti-interference of different methods in the production line. Therefore, the M3 shutdown is set for 100 s at $t = 500$, indicating that the M3 suddenly fails, which is used as disturbance information in the manufacturing process.

Due to the complex structure of ventilator products and large assembly process data, only the scheduling status comparison of some stations is given in Fig. 8, where H2′, H5′, and M3′ respectively represent the scheduling for other production lines. The comparison revealed that the manual-based approach is delayed when facing the requirements of new assembly tasks, and the accuracy of the decisions was lower, with instances of interruption. For example, the task assigned to M2 at 440 s cannot be completed at once, forcing a downtime wait state to occur because the task of H1, which constrains the relevant behavior, has not yet been completed and cannot be followed up. The DT-based approach, however, can be used in downtime scenarios to discover alternative devices more efficiently and reduce the waiting time. Based on the method proposed in the article, better global planning can be performed, especially in M3′ and H5′, and more tasks are assigned to M3′, which effectively reduces the task load of H5′.

## 7 Conclusions

In this paper, the flexible scheduling technology of the human–machine collaborative assembly workshop is studied, and the digital twin environment model for scheduling is analyzed. Aiming at the dynamic adjustment efficiency of the assembly workshop system, a hierarchical reinforcement learning algorithm is constructed to adjust the allocation from the workshop level and production line level. The integration of the respective advantages of man and machine makes it possible to deal with more complex tasks. At the same time, the flexible scheduling method will further improve the adjustability of the manufacturing system.

In future work, we will further refine other influencing factors in the scheduling process, such as the different ease of migration of different machines, and more human factors engineering should be taken into account, such as the hazard level of the operation, the duration, and the energy input during the operation. In addition, it will be interesting to construct process knowledge graphs for assembly tasks and link the workstation attributes of operational tasks.

**Acknowledgements** This work is financially supported by National Key Research and Development Plan of China (Grant 2019YFB1706300), in part by the Fundamental Research Funds for the Central Universities and Graduate Student Innovation Fund of Donghua University (CUSF-DH-D-2022072).

## Declarations

**Conflict of interest** No conflict of interest exits in the submission of this manuscript.

**Ethical approval** This manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

## References

Çil ZA, Li Z, Mete S, Özceylan E (2020) Mathematical model and bee algorithms for mixed-model assembly line balancing problem with physical human-robot collaboration. Appl Soft Comput 93:106394

Dou J, Li J, Xia D, Zhao X (2021) A multi-objective particle swarm optimisation for integrated configuration design and scheduling in reconfigurable manufacturing system. Int J Prod Res 59(13):3975–3995

El-Shamouty M, Wu X, Yang S, Albus M, Huber MF (2020) Towards safe human-robot collaboration using deep reinforcement learning. In: 2020 IEEE international conference on robotics and automation (ICRA), pp 4899–4905

Ghadirzadeh A, Chen X, Yin W, Yi Z, Björkman M, Kragic D (2020) Human-centered collaborative robots with deep reinforcement learning. IEEE Robot Autom Lett 6(2):566–571

Ham A, Park MJ (2021) Human-robot task allocation and scheduling: Boeing 777 case study. IEEE Robot Autom Lettrs 6(2):1256–1263

He L, Chiong R, Li W, Dhakal S, Cao Y, Zhang Y (2021) Multiobjective optimization of energy-efficient job-shop scheduling with dynamic reference point-based fuzzy relative entropy. IEEE Trans Industr Inf 18(1):600–610

Hoseinpour Z, Kheirkhah AS, Fattahi P, Taghipour M (2020) The problem solving of bi-objective hybrid production with the possibility of production outsourcing through meta-heuristic algorithms. Management 4(2):1–17

Hoseinpour Z, Taghipour M, Beigi JH, Mahboobi M (2021) The problem solving of bi-objective hybrid production with the possibility of production outsourcing through imperialist algorithm, NGSA-II, GAPSO hybrid algorithms. Turk J Comput Math Educ (TURCOMAT) 12(13):8090–8111

Hu L, Liu Z, Hu W, Wang Y, Tan J, Wu F (2020) Petri-net-based dynamic scheduling of flexible manufacturing system via deep reinforcement learning with graph convolutional network. J Manuf Syst 55:1–14

Hu X, Guo J, Zhang Y (2019) Optimal strategies for the yard truck scheduling in container terminal with the consideration of container clusters. Comput Ind Eng 137:106083

Kim YG, Lee S, Son J, Bae H, Do Chung B (2020) Multi-agent system and reinforcement learning approach for distributed intelligence in a flexible smart manufacturing system. J Manuf Syst 57:440–450
Lin CH, Wang KJ, Tadesse AA, Woldegiorgis BH (2022) Human-robot collaboration empowered by hidden semi-Markov model for operator behaviour prediction in a smart assembly system. J Manuf Syst 62:317–333
Liu S, Bao J, Zheng P (2023) A review of digital twin-driven machining: From digitization to intellectualization. J Manuf Syst 67:361–378
Luo S, Zhang L, Fan Y (2021) Dynamic multi-objective scheduling for flexible job shop by deep reinforcement learning. Comput Ind Eng 159:107489
Pupa A, Van Dijk W, Secchi C (2021) A human-centered dynamic scheduling architecture for collaborative application. IEEE Robot Autom Lett 6(3):4736–4743
Rahman SM, Wang Y (2018) Mutual trust-based subtask allocation for human-robot collaboration in flexible lightweight assembly in manufacturing. Mechatronics 54:94–109
Rodríguez I, Nottensteiner K, Leidner D, Durner M, Stulp F, Albu-Schäffer A (2020) Pattern recognition for knowledge transfer in robotic assembly sequence planning. IEEE Robot Autom Lett 5(2):3666–3673
Wang H, Yan Q, Zhang S (2021) Integrated scheduling and flexible maintenance in deteriorating multi-state single machine system using a reinforcement learning approach. Adv Eng Inform 49:101339
Wikarek J, Sitek P, Nielsen P (2019) Model of decision support for the configuration of manufacturing system. IFAC-PapersOnLine 52(13):826–831
Xu W, Tang Q, Liu J, Liu Z, Zhou Z, Pham DT (2020) Disassembly sequence planning using discrete bees algorithm for human-robot collaboration in remanufacturing. Robot Computer-Integr Manuf 62:101860
Yu T, Huang J, Chang Q (2020) Mastering the working sequence in human-robot collaborative assembly based on reinforcement learning. IEEE Access 8:163868–163877. https://doi.org/10.1109/ACCESS.2020.3021904
Zhang C, Xu W, Liu J, Liu Z, Zhou Z, Pham DT (2019) A reconfigurable modeling approach for digital twin-based manufacturing system. Proc Cirp 83:118–125
Zhang Y, Tang D, Zhu H, Li S, Nie Q (2021) A flexible configuration method of distributed manufacturing resources in the context of social manufacturing. Comput Ind 132:103511
Zhang Z, Wu L, Zhang W, Peng T, Zheng J (2021) Energy-efficient path planning for a single-load automated guided vehicle in a manufacturing workshop. Comput Ind Eng 158:107397
Zhou T, Tang D, Zhu H, Zhang Z (2021) Multi-agent reinforcement learning for online scheduling in smart factories. Robot Computer-Integr Manuf 72:102202

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Rong Zhang** received the B.S. degree in North University of China, Taiyuan, in 2016 and the M.S. degree in Guangxi University of Science and Technology, Liuzhou, in 2019. He is currently pursuing the Ph.D. degree in College of Mechanical Engineering, Donghua University, Shanghai. His research interests are mainly reinforcement learning, computer vision and strategy research related to human-robot collaboration.

**Jianhao Lv** received the B.S. degree in Lanzhou University of Technology, Lanzhou, in 2018 and the M.S. degree in Lanzhou University of Technology, Lanzhou, in 2021. He is currently pursuing the Ph.D. degree in College of Mechanical Engineering, Donghua University, Shanghai. His research interests are mainly on computer vision, digital twin and modeling research related to human-robot collaboration.

**Jinsong Bao** received the Ph.D. degree in Shanghai Jiao Tong University, in 2002. He is currently working as the director of the Intelligent Manufacturing Research Institute at Donghua University. His research range from CPS/CPPS. Digital Thread/Digital Twin, Industrial CV/ NLP, Industrial Knowledge Graph, Big Data, Human Facto, Co-Robot, and Cognitive Computing to VR/AR/MR.

**Yu Zheng** received the Ph.D. degree in Shanghai Jiao Tong University, in 2015. She is currently an Associate Professor with the Institute of Intelligent Manufacturing and Information Engineering, Shanghai Jiao Tong University, Shanghai, China. Her research interests include intelligent manufacturing system, digital twin, data mining, and knowledge-based engineering.

## Authors and Affiliations

**Rong Zhang[1] · Jianhao Lv[1] · Jinsong Bao[1] · Yu Zheng[2]**

✉ Jinsong Bao
bao@dhu.edu.cn

Rong Zhang
1199074@mail.dhu.edu.cn

Jianhao Lv
c953749@163.com

Yu Zheng
yuzheng@sjtu.edu.cn

[1] College of Mechanical Engineering, Donghua University, Shanghai 201620, China

[2] Institute of Intelligent Manufacturing and Information Engineering, Shanghai Jiao Tong University, Shanghai 200240, China